

Data Science e Machine Learning

Dai Dati alla Conoscenza

Immagini del libro a colori

MICHELE DI NUZZO

Copyright © 2021 Michele di Nuzzo

Tutti i diritti riservati.

Codice ISBN: 979-8-4636-4718-4

Indice

1.	Data Science e Machine Learning.....	1
2.	R per la Data Science.....	6
3.	Matematica e Algebra.....	18
4.	Statistica Descrittiva.....	20
5.	Teoria della Probabilità.....	33
6.	Principali Distribuzioni.....	36
7.	Statistica Inferenziale.....	44
8.	Il Campionamento.....	51
9.	La preparazione dei dati.....	54
10.	Feature Engineering.....	58
11.	Machine Learning.....	61
12.	La Validazione.....	66
13.	Gli algoritmi di Regressione.....	72
14.	Algoritmi di Classificazione.....	77
15.	Algoritmi di Clustering.....	84
16.	Riduzione della Dimensionalità.....	97
17.	Algoritmi Semi Supervisionati.....	99
18.	Lazy Learning.....	106
19.	Metodi Ensemble.....	109
20.	Algoritmi di associazione.....	110
21.	Serie Temporal.....	113
22.	Reti Neurali.....	117
23.	Deep Learning.....	128
24.	Automated Machine Learning.....	145
25.	Visualizzare i dati.....	146
26.	Comunicare i dati.....	150

1. DATA SCIENCE E MACHINE LEARNING

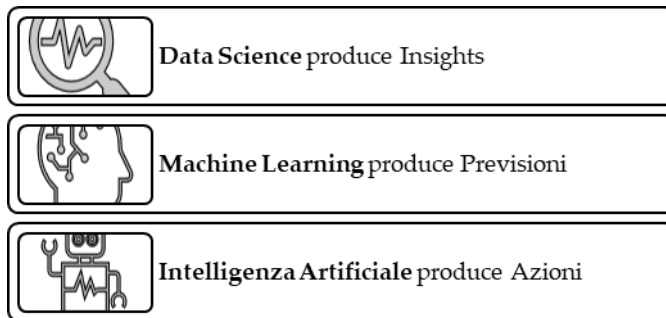


Figura 1.1 - Data Science, Machine Learning e AI

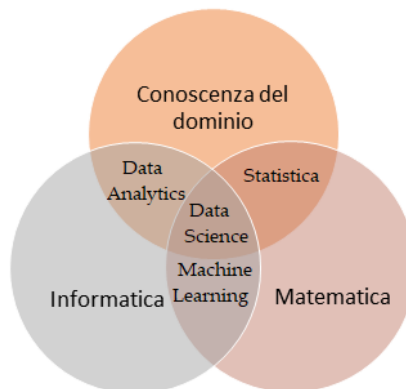


Figura 1.2 - Le skills del Data Scientist

Definire gli obiettivi	<ul style="list-style-type: none"> • Definire gli obiettivi • Creare un piano di progetto
Recuperare i dati	<ul style="list-style-type: none"> • Dati interni • Dati esterni
Preparare i dati	<ul style="list-style-type: none"> • Pulizia dei dati • Trasformazione dei dati • Fusione dei dati
Esplorare i dati	<ul style="list-style-type: none"> • Analisi univariata • Analisi multivariata • Correlazioni
Creare il modello	<ul style="list-style-type: none"> • Feature Engineering • Modeling • Validazione
Presentare i risultati	<ul style="list-style-type: none"> • Presentare i dati • Automatizzare le analisi

Figura 1.3 - Il processo di un progetto di data science

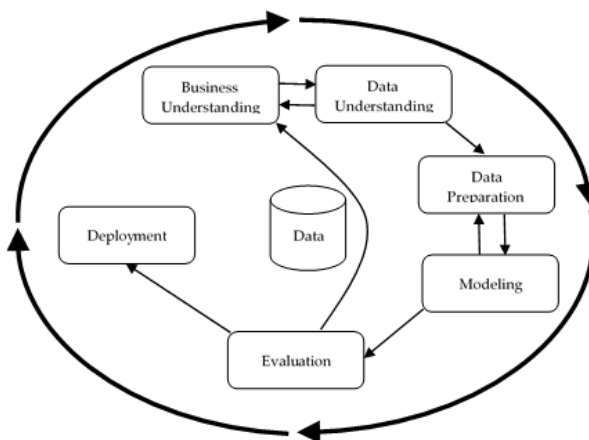


Figura 1.4 - Le fasi del modello CRISP-DM

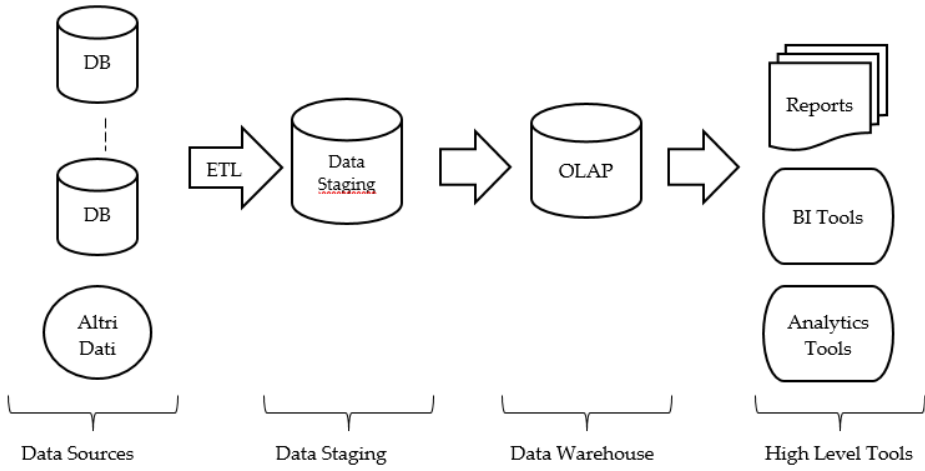


Figura 1.5 - Le aree di un sistema di Business Intelligence

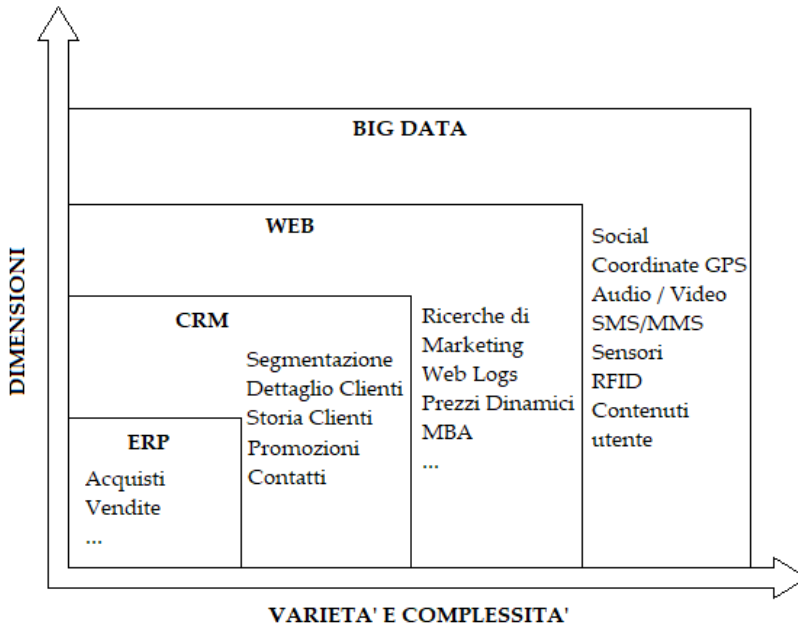


Figura 1.6 - Come si sono evoluti i Big Data

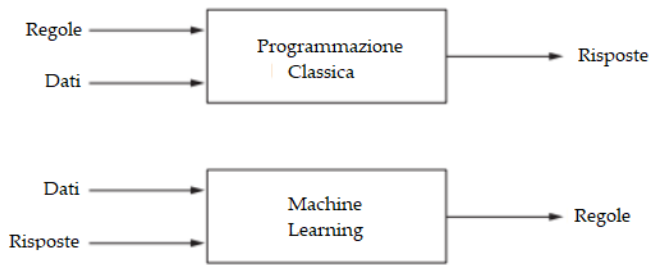


Figura 1.7 - Programmazione classica vs Machine Learning

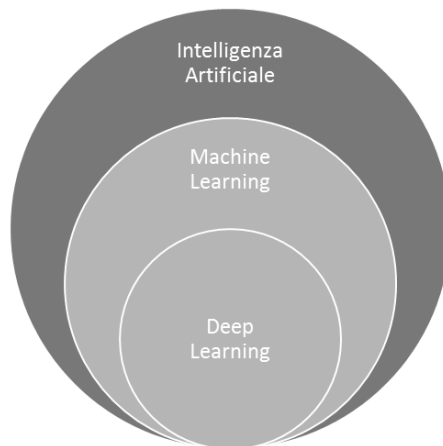


Figura 1.8 - Intelligenza Artificiale, Machine Learning e Deep Learning

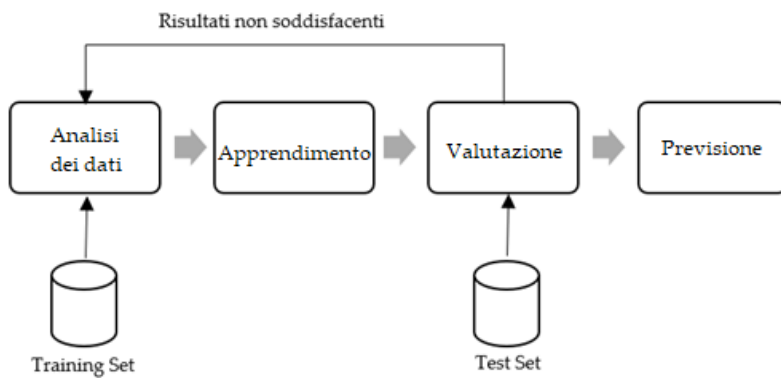


Figura 1.9 - Le fasi di un progetto di Machine Learning

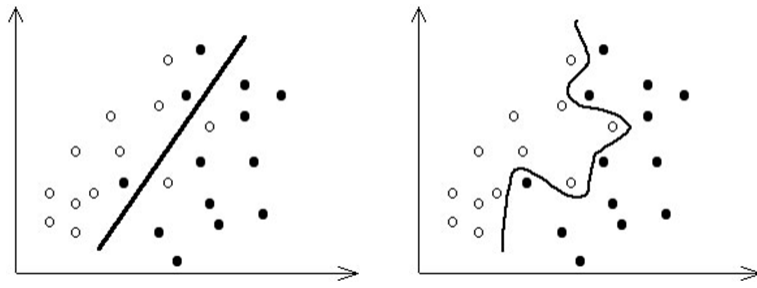


Figura 1.10 - Classificatori con underfitting e con overfitting

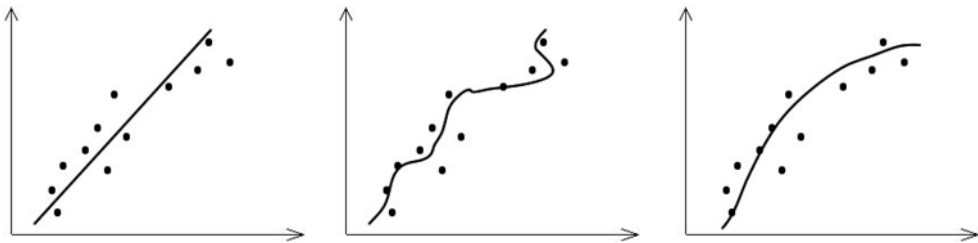


Figura 1.11 - Regressore con underfitting, con overfitting e bilanciato

2. R PER LA DATA SCIENCE

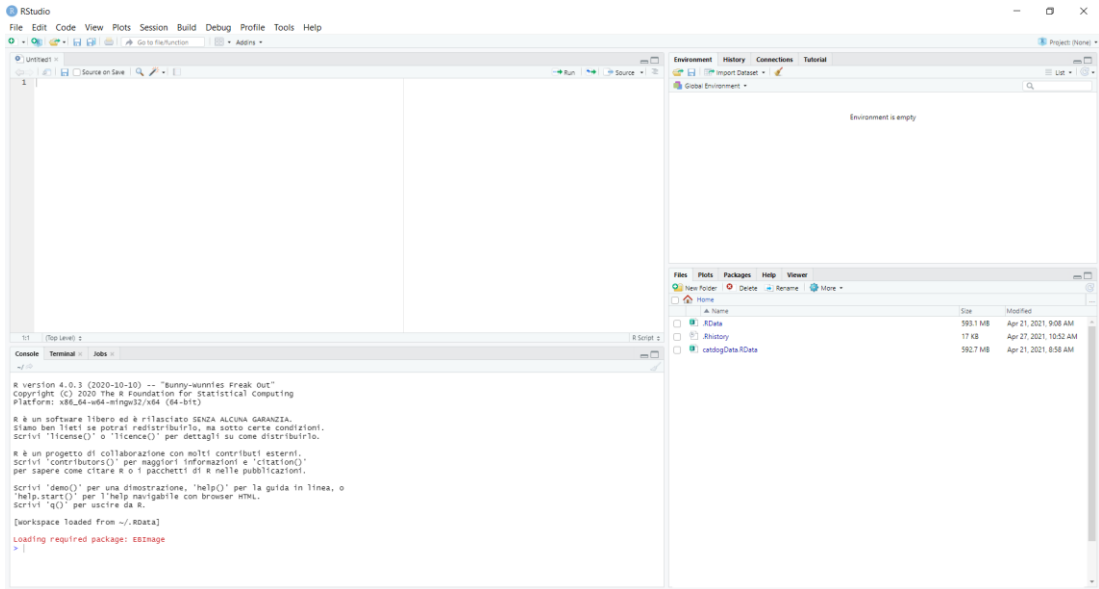


Figura 2.1 - L'IDE R Studio

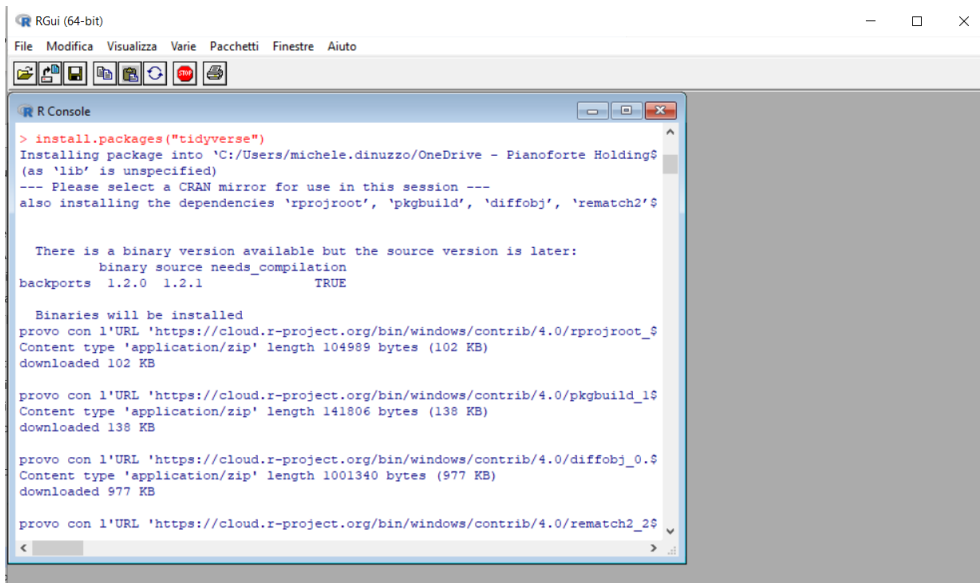


Figura 2.2 - Installazione del pacchetto tidyverse

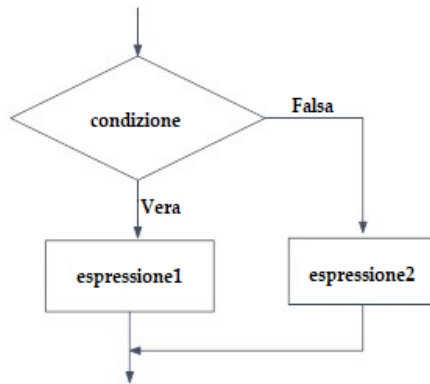


Figura 2.3 - Diagramma del flusso if-else

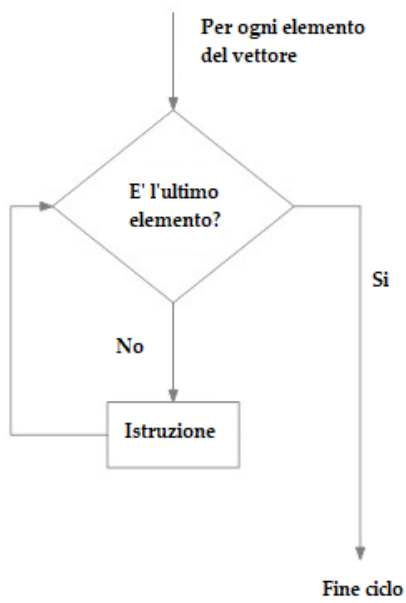


Figura 2.4 – Il ciclo for

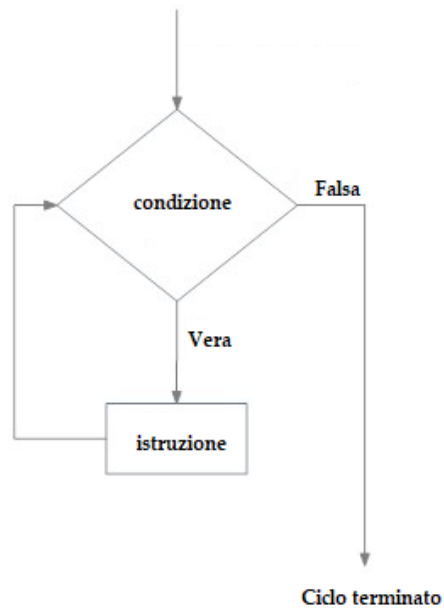


Figura 2.5 - Il ciclo while

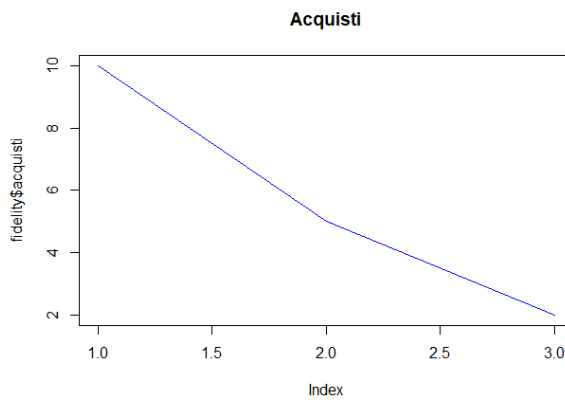


Figura 2.6 - Un grafico prodotto con la funzione plot

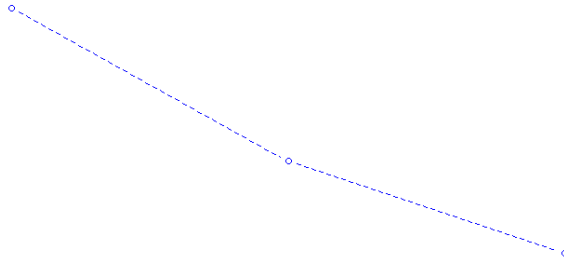


Figura 2.7 - Il nostro nuovo grafico

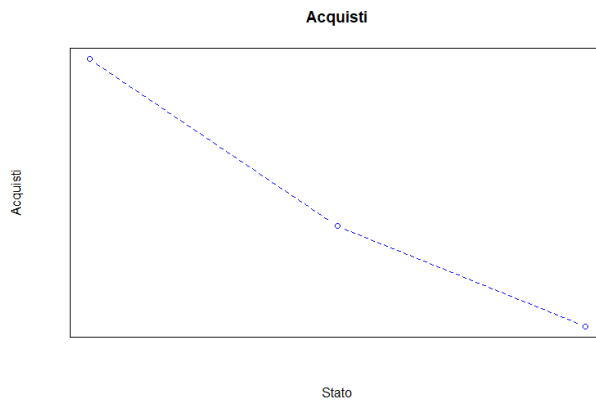


Figura 2.8 - Il grafico con l'aggiunta dei titoli e del box

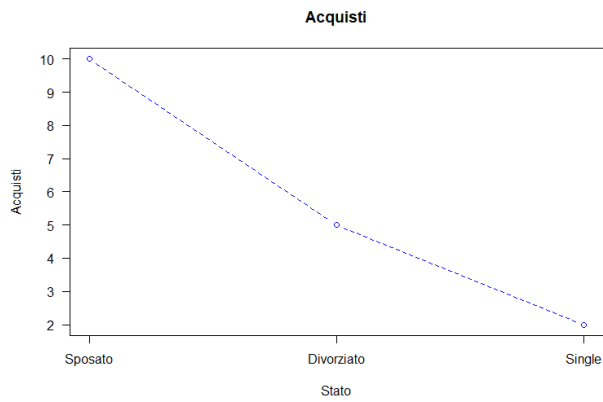


Figura 2.9 - Il nostro grafico finale

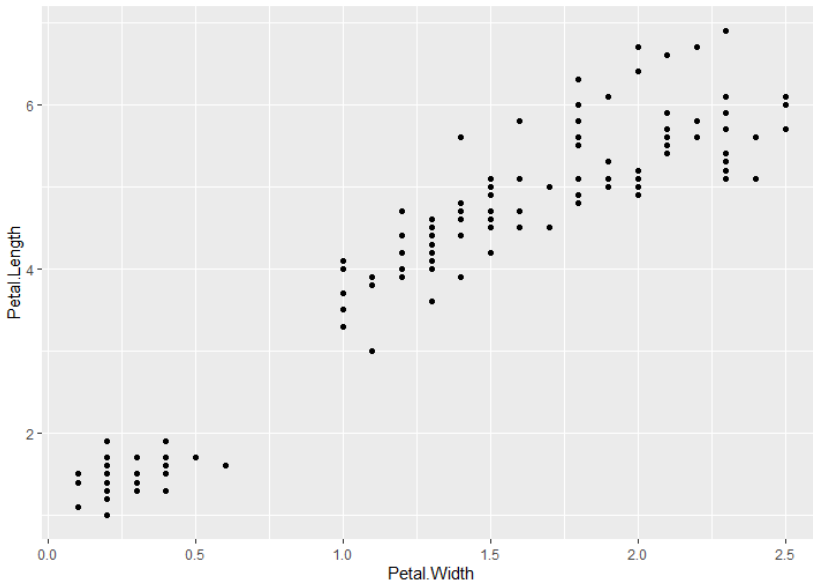


Figura 2.10 - Plot dei dati con la funzione `qplot()`

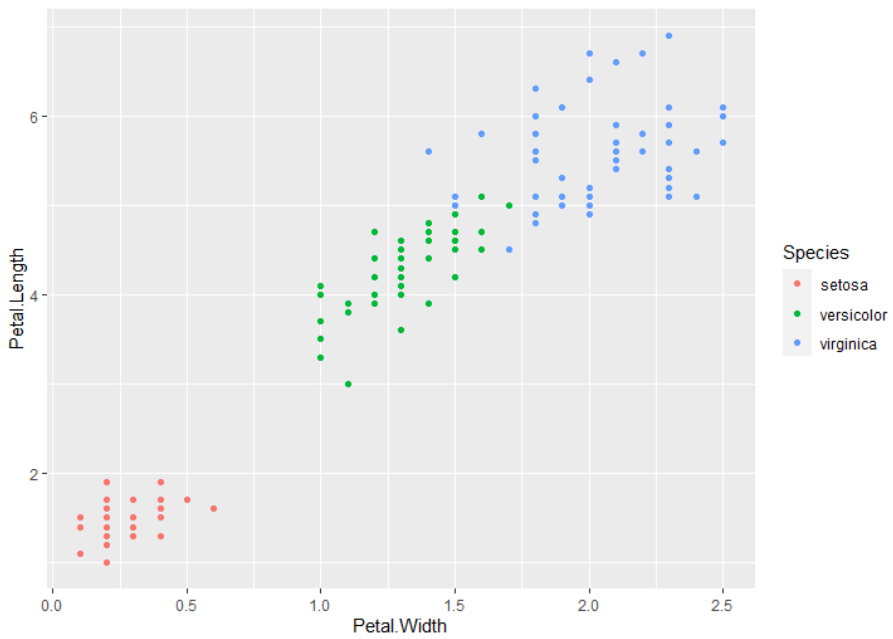


Figura 2.11 - Dati raggruppati per colore

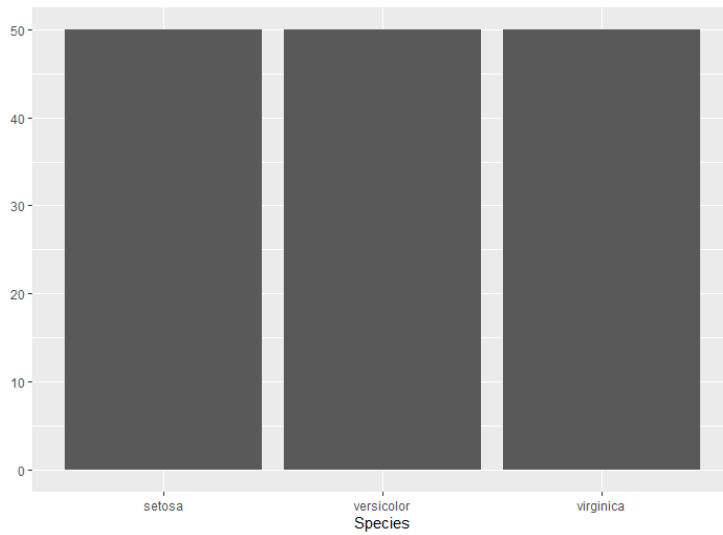


Figura 2.12 - *qplot* con una variabile

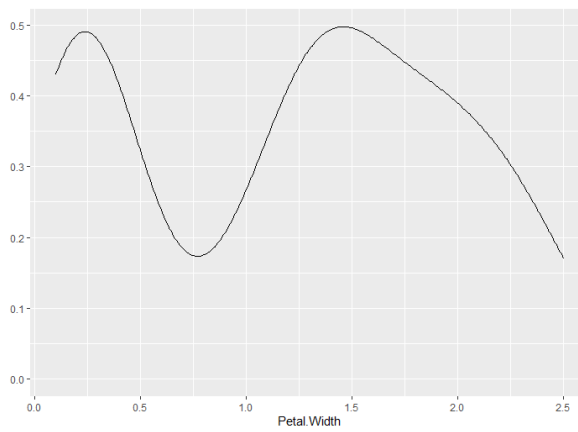


Figura 2.13 - Grafico di una variabile con *geom = density*

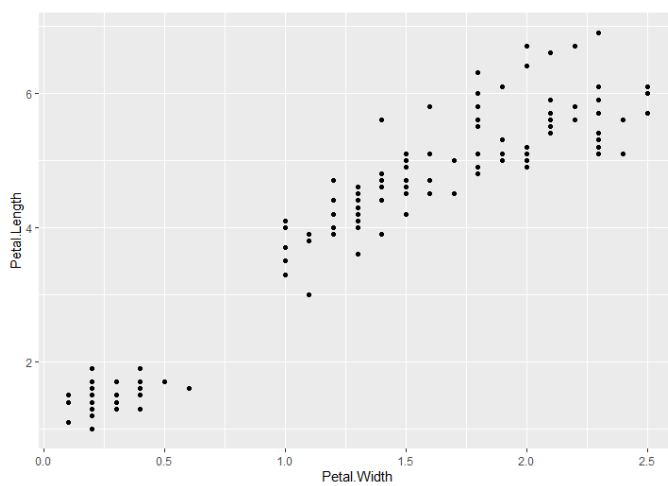


Figura 2.14 - Scatterplot di due variabili

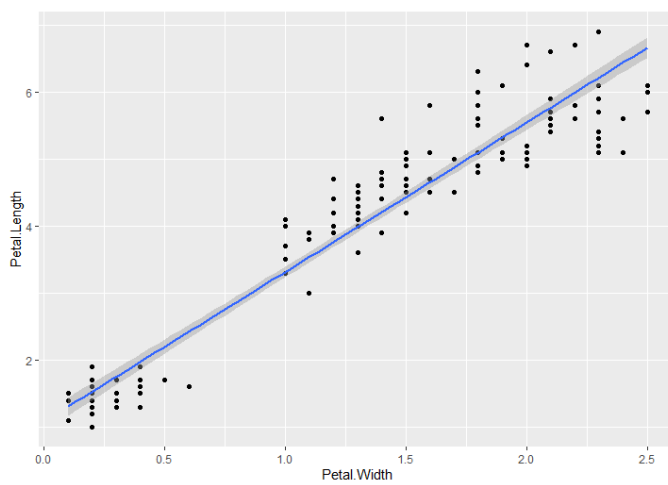


Figura 2.15 - Aggiunta di una retta di regressione

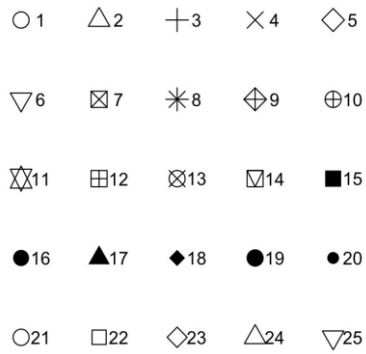


Figura 2.16 - Le possibili forme dei punti in R

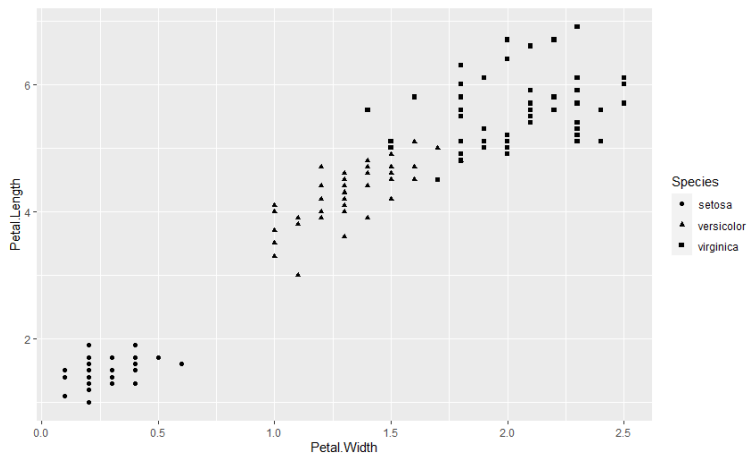


Figura 2.17 - Forme del punto per classe

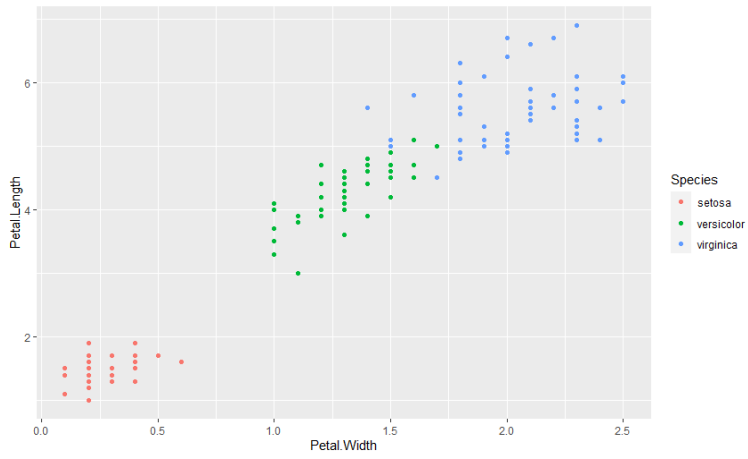


Figura 2.18 - Colore per classe

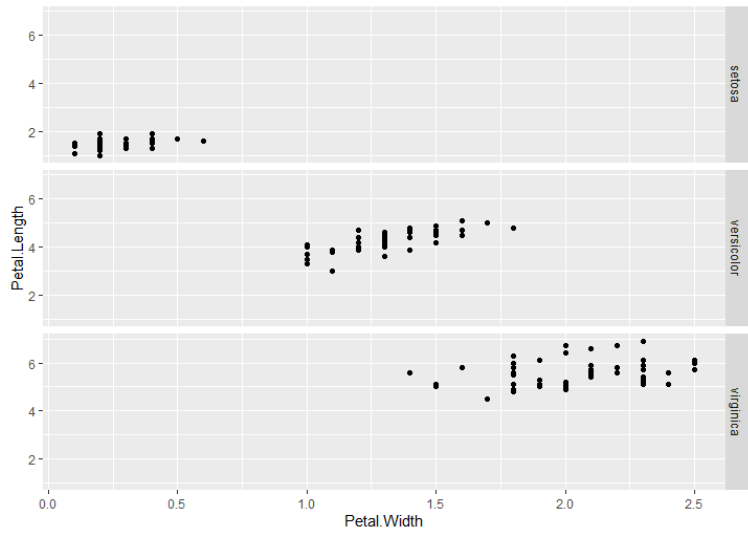


Figura 2.19 - Grafico suddiviso in righe

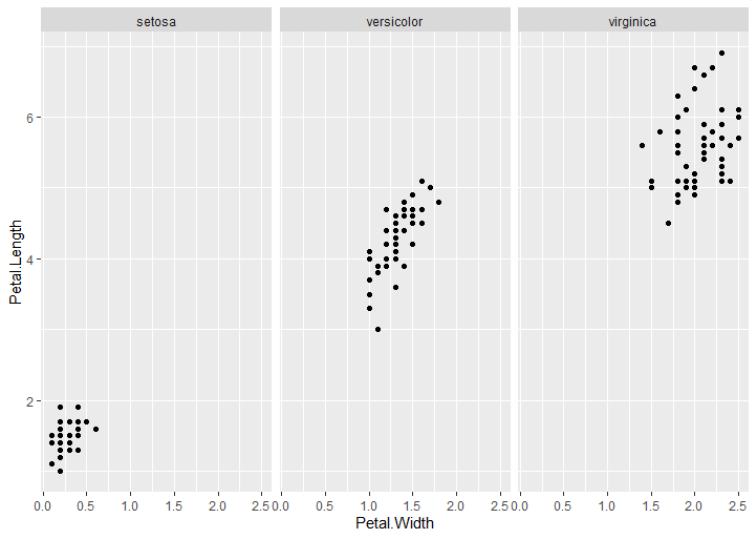


Figura 2.20 - Grafico suddiviso in colonne

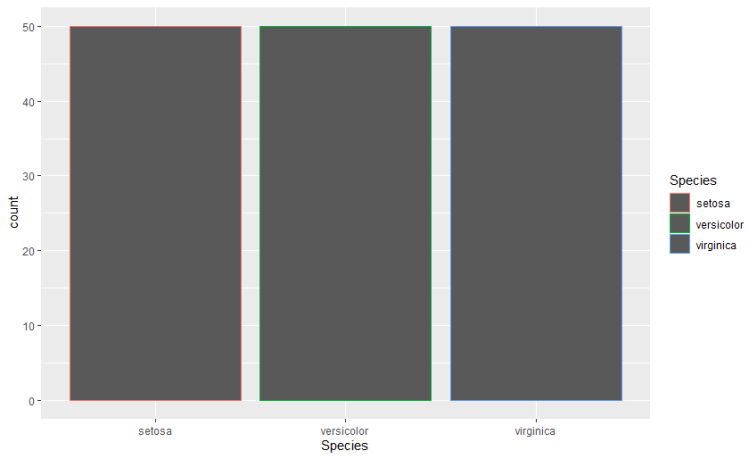


Figura 2.21 - Il parametro color

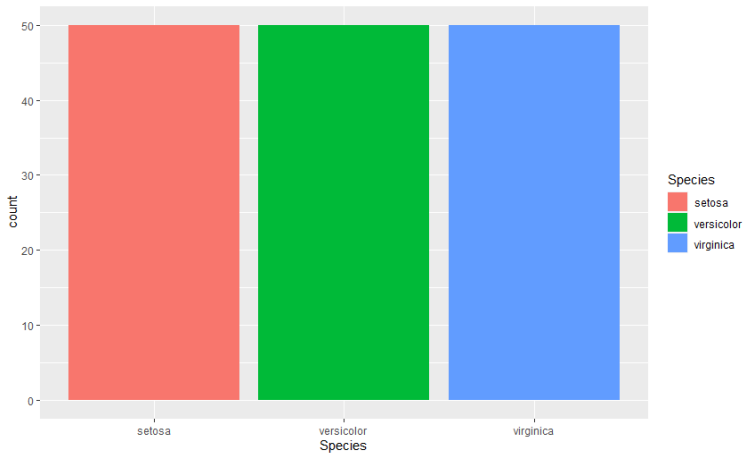


Figura 2.22 - Il parametro fill

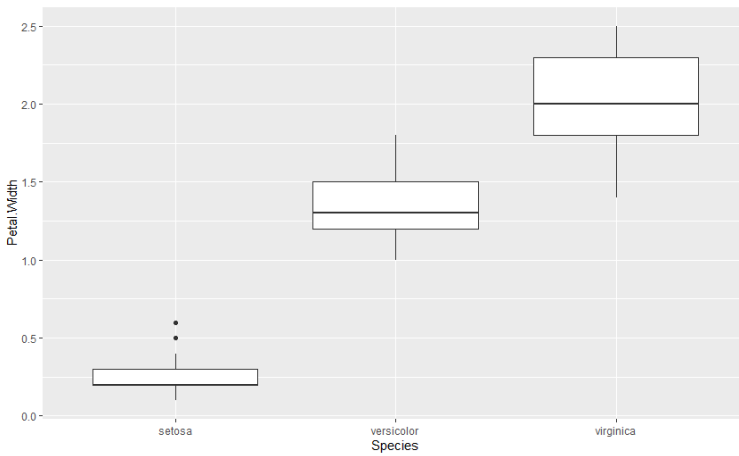


Figura 2.23 - Boxplot di Petal.Width

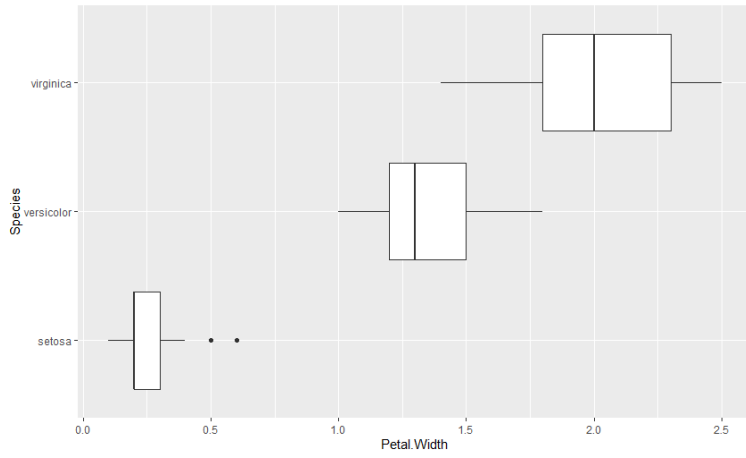


Figura 2.24 - Boxplot di Petal.Width con assi invertiti

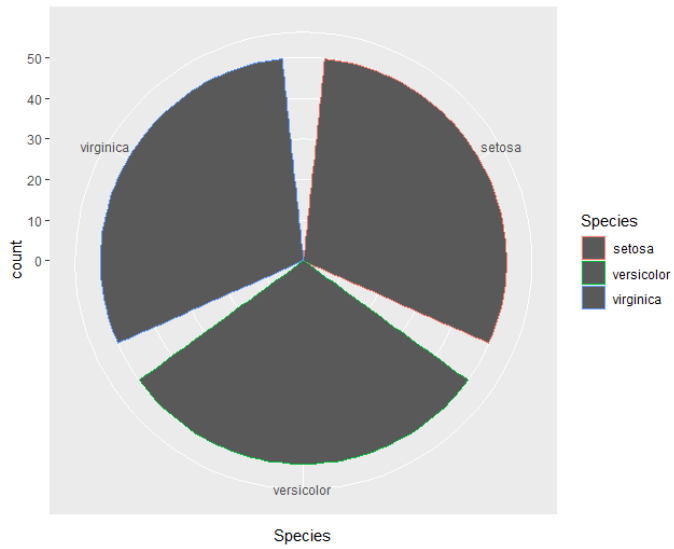


Figura 2.25 - Coordinate polari

3. MATEMATICA E ALGEBRA

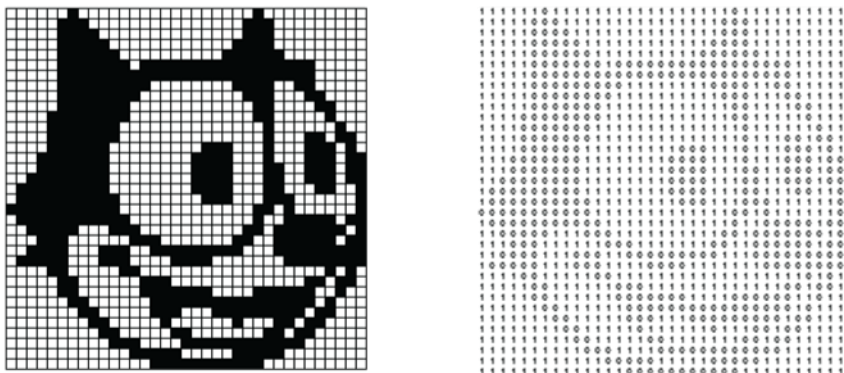


Figura 3.1 - Rappresentazione matriciale dell'immagine

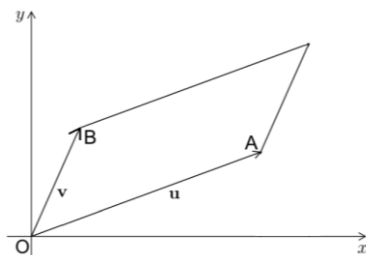


Figura 3.2 - Rappresentazione vettoriale

$$\begin{array}{|c|} \hline A \\ \hline m * n \\ \hline \end{array} = \begin{array}{|c|} \hline U \\ \hline m * p \\ \hline \end{array} \begin{array}{|c|} \hline D \\ \hline p * p \\ \hline \end{array} \begin{array}{|c|} \hline V^T \\ \hline p * n \\ \hline \end{array}$$

Figura 3.3 - SVD

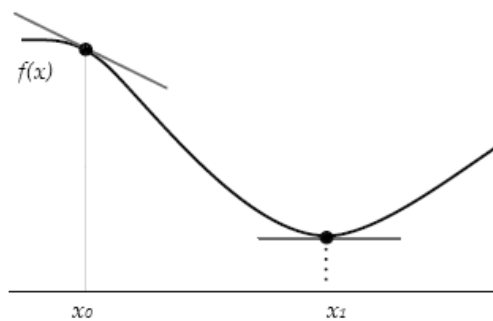


Figura 3.4 - Una generica funzione

4. STATISTICA DESCRITTIVA

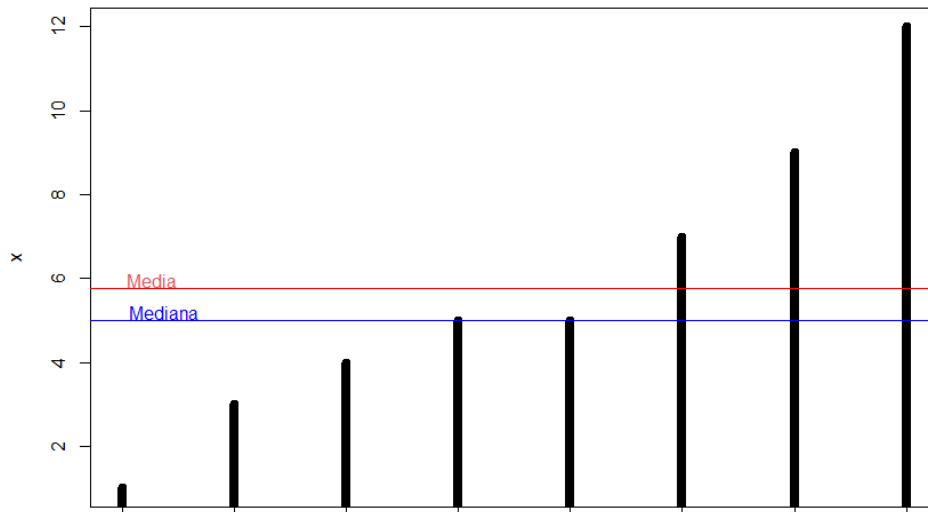


Figura 4.1 - Media e mediana

Famiglia	Reddito
Unità 1	32.000
Unità 2	43.000
Unità 3	28.000
Unità 4	29.500
Unità 5	34.000
Unità 6	42.000
Unità 7	47.000
Unità 8	-
Media	36.500

Figura 4.2 - Esempio di media

Famiglia	Reddito
Unità 1	32.000
Unità 2	43.000
Unità 3	28.000
Unità 4	29.500
Unità 5	34.000
Unità 6	42.000
Unità 7	47.000
Unità 8	1.300.000
Media	194.438

Figura 4.3 - Esempio di media

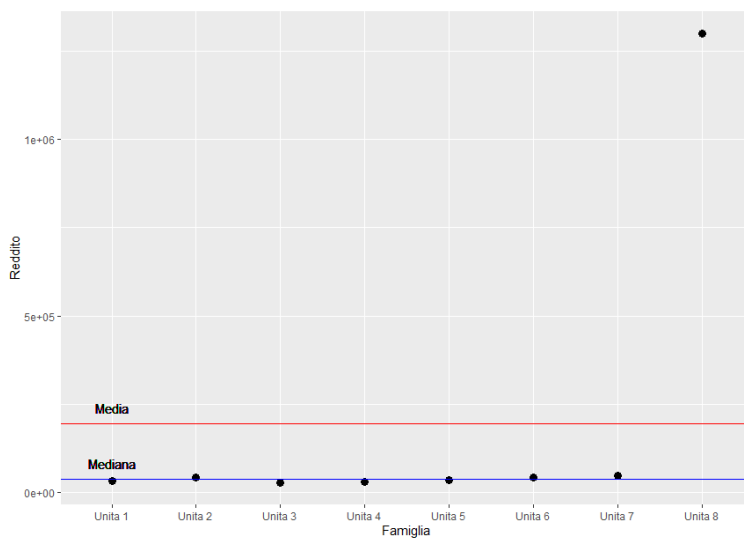


Figura 4.4 - Grafico dei valori con media e mediana

Osservazione	Figli Campione A	Figli Campione B	Figli Campione C
1	2	1	1
2	2	2	5
3	2	3	2
4	2	3	2
5	2	2	0
6	2	1	1
7	2	2	3
Media	2	2	2
Dev. Standard	0	0,76	1,51
Varianza	0	0,57	2,29

Figura 4.5 - Tre campioni con la stessa media

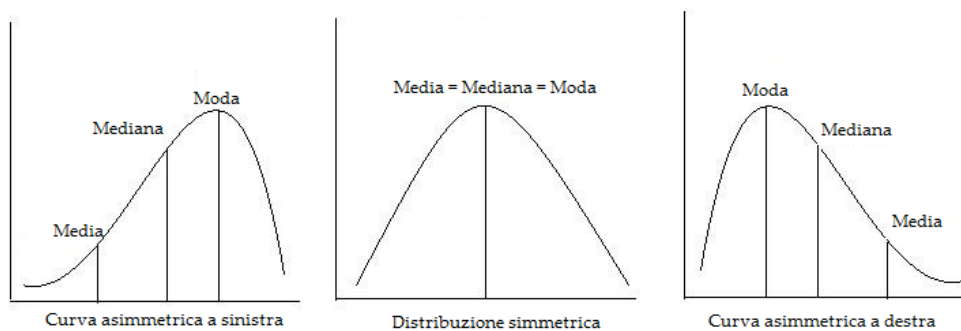


Figura 4.6 - Tipi di asimmetria

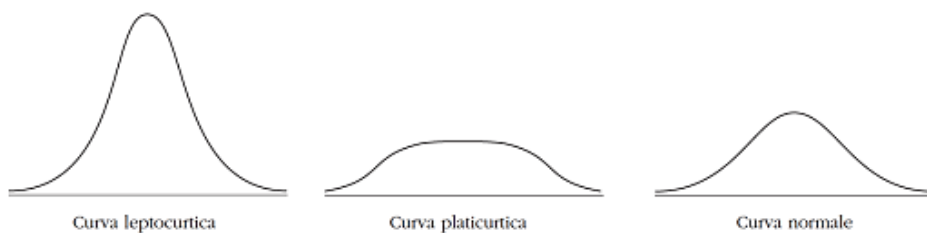


Figura 4.7 - Curtosi

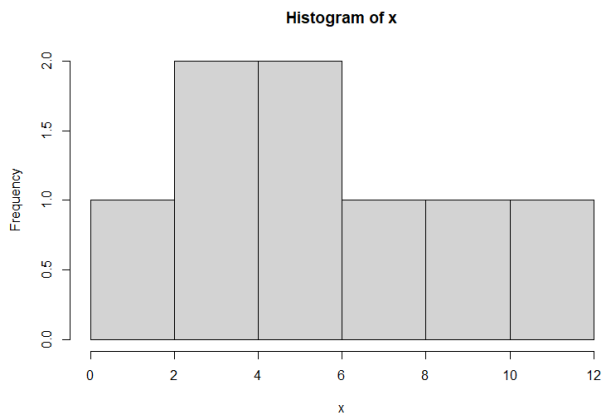


Figura 4.8 - Istogramma della nostra variabile

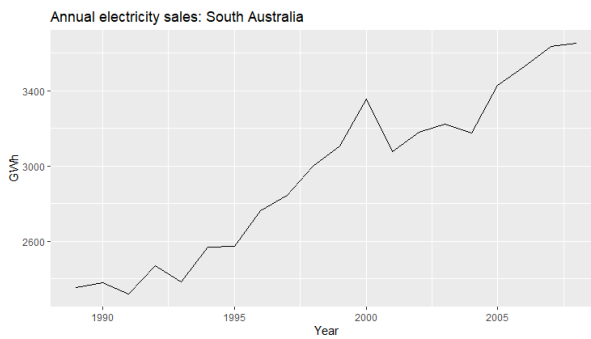


Figura 4.9 - Il dataset elecsales

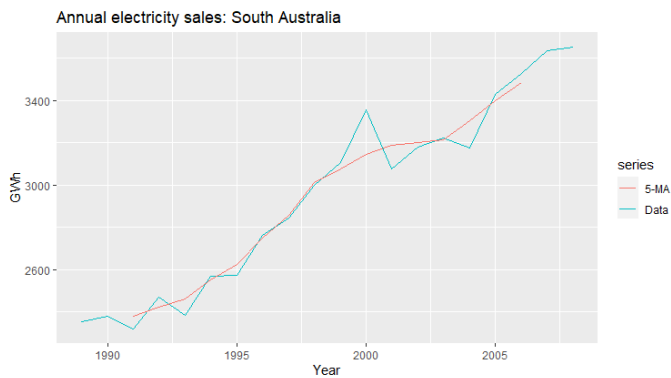


Figura 4.10 - Plot della media 5-MA

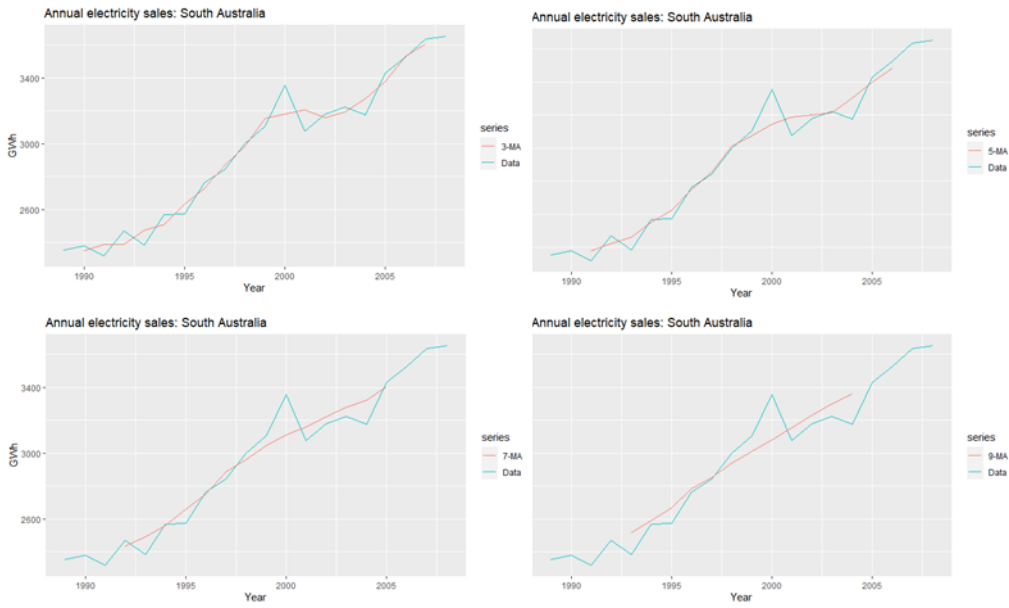


Figura 4.11 - Media mobile per diversi valori di m

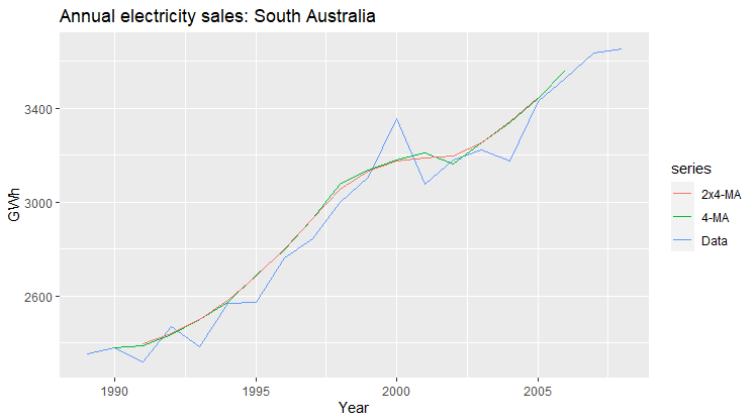


Figura 4.12 - 2x4MA

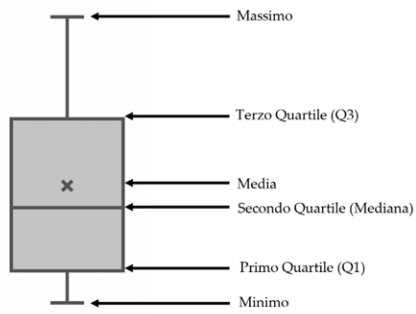


Figura 4.13 - Diagramma Boxplot

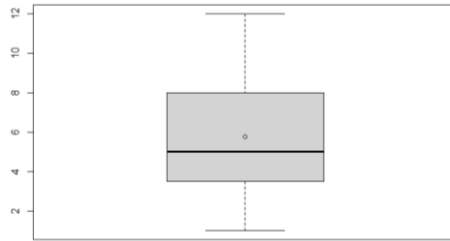


Figura 4.14 - Il boxplot con la media

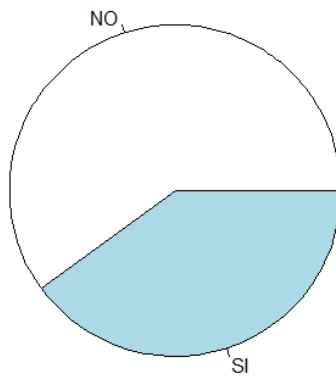


Figura 4.15 - Diagramma a torta di una variabile discreta

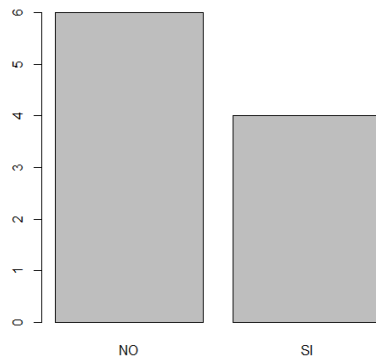


Figura 4.16 - Diagramma a barre

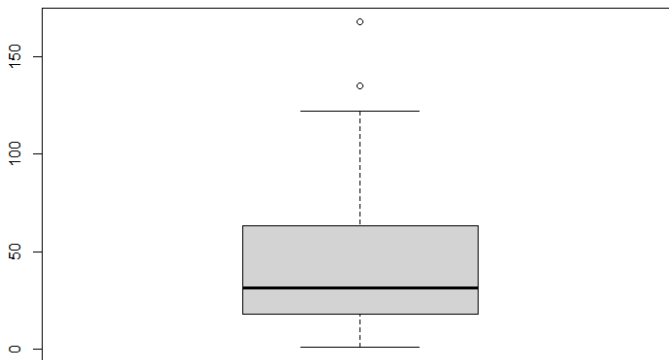


Figura 4.17 - Il boxplot della variabile Ozone

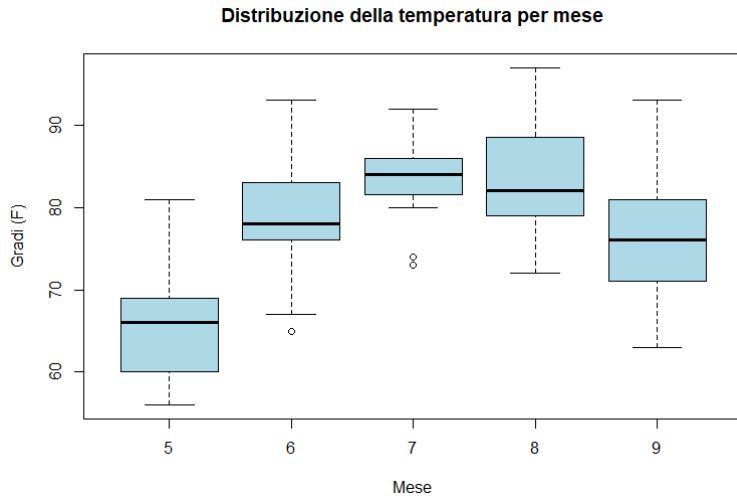


Figura 4.18 - Boxplot della variabile Temp per Mese

Fidelizzato	Uomini	Donne	TOTALE
SI	$f_{si,u}$	$f_{si,d}$	f_{si}
NO	$f_{no,u}$	$f_{no,d}$	f_{no}
TOTALE	f_u	f_d	f

Figura 4.19 - Tabella di contingenza

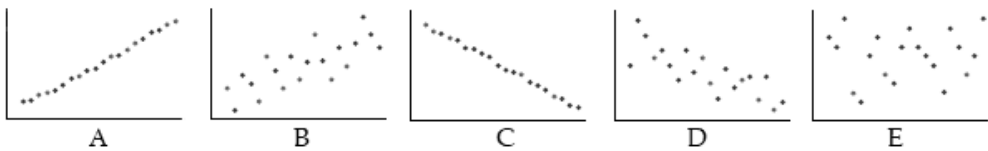


Figura 4.20 - Esempio di grafici a dispersione con diversi livelli di correlazione

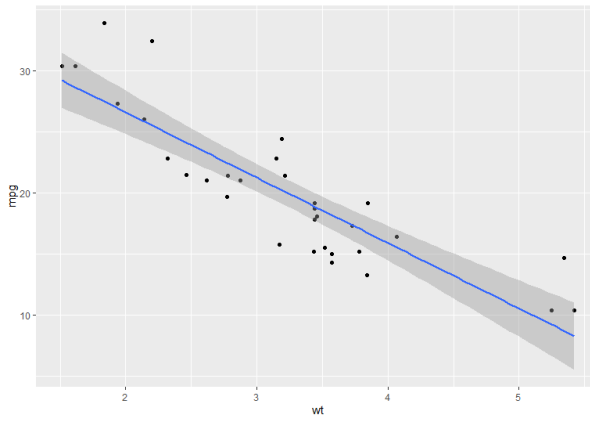


Figura 4.21 - Scatterplot delle variabili mpg e wt

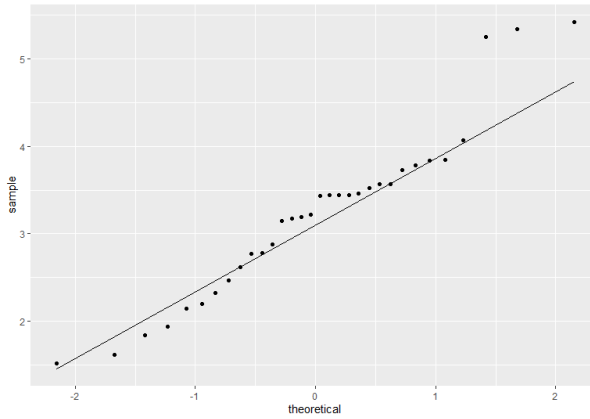


Figura 4.22 - Grafico Q-Q della variabile wt

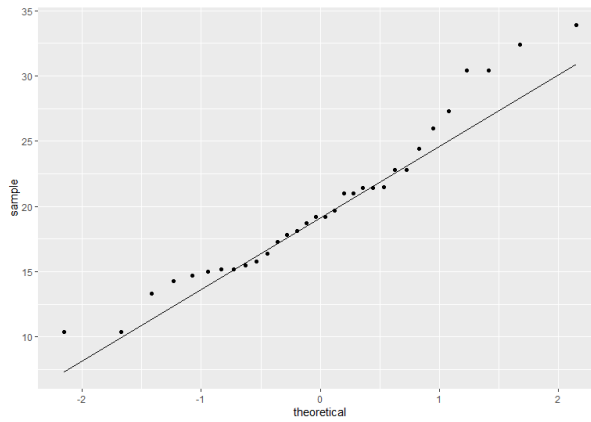


Figura 4.23 - Grafico Q-Q della variabile mpg

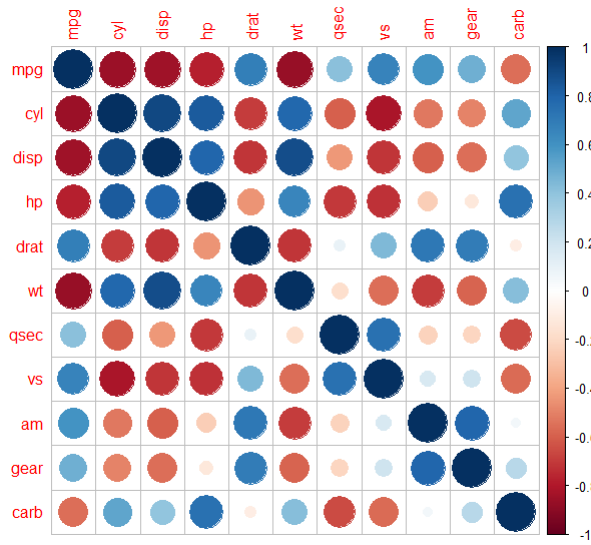


Figura 4.24 - Grafico delle correlazioni

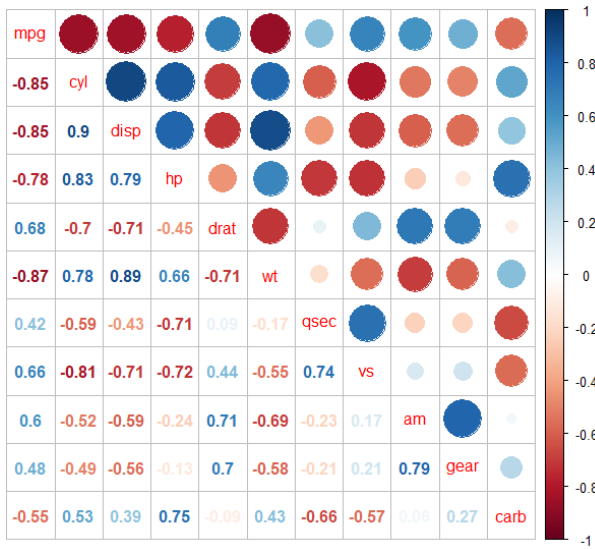


Figura 4.25 - Grafico e valore delle correlazioni

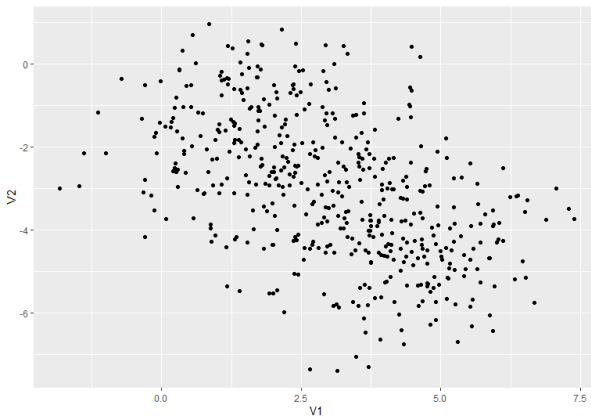


Figura 4.26 - I dati generati dalla simulazione

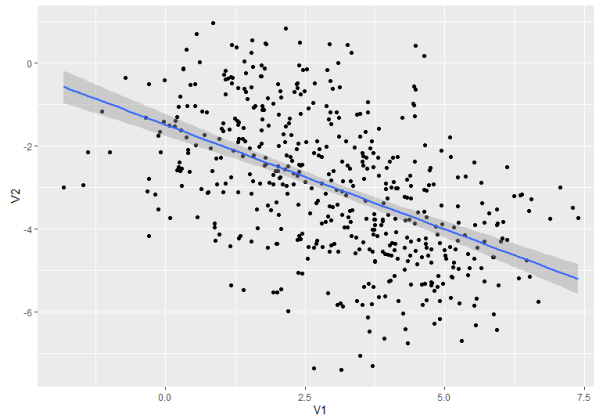


Figura 4.27 - La retta di regressione tra le variabili

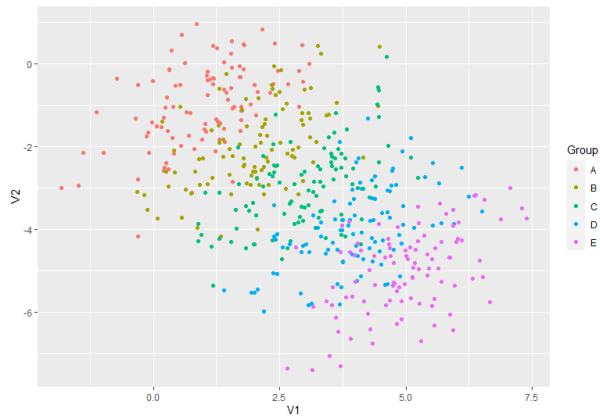


Figura 4.28 - Suddivisione per gruppo

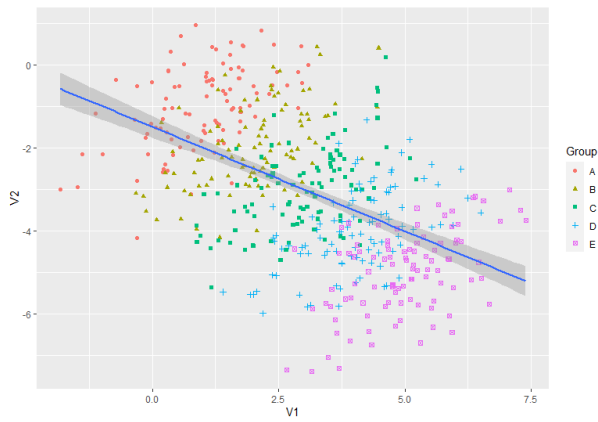


Figura 4.29 - La retta di regressione

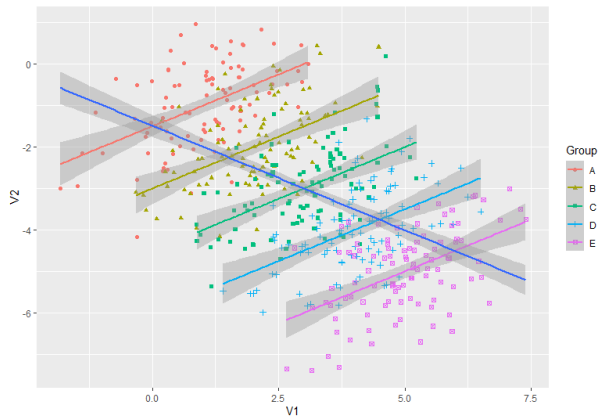


Figura 4.30 - Le correlazioni tra gruppi

5. TEORIA DELLA PROBABILITÀ

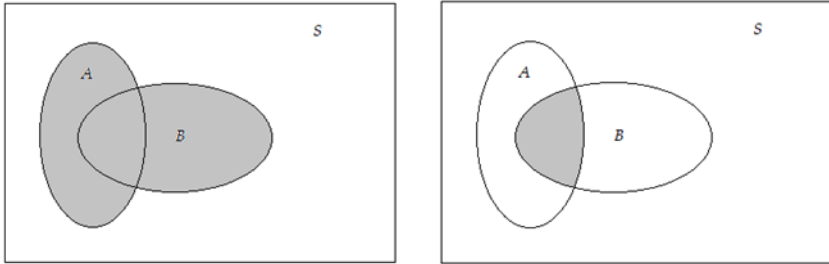


Figura 5.1 – Unione e intersezione di due eventi

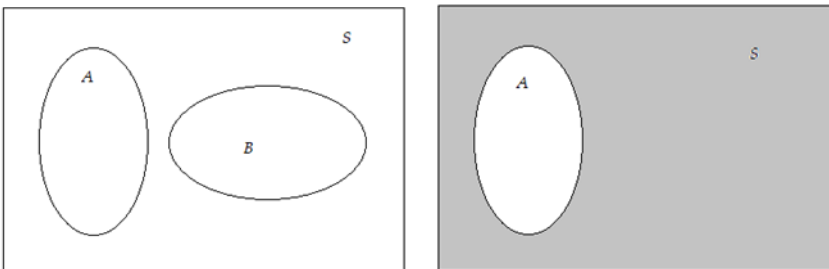


Figura 5.2 - Eventi disgiunti ed evento complementare

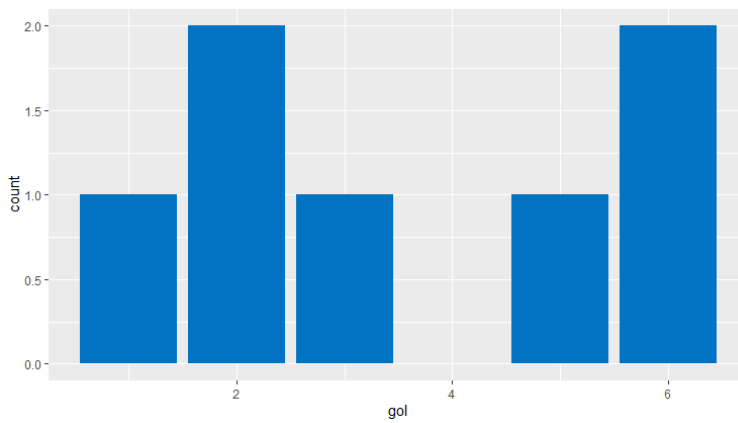


Figura 5.3 - Distribuzione della variabile gol

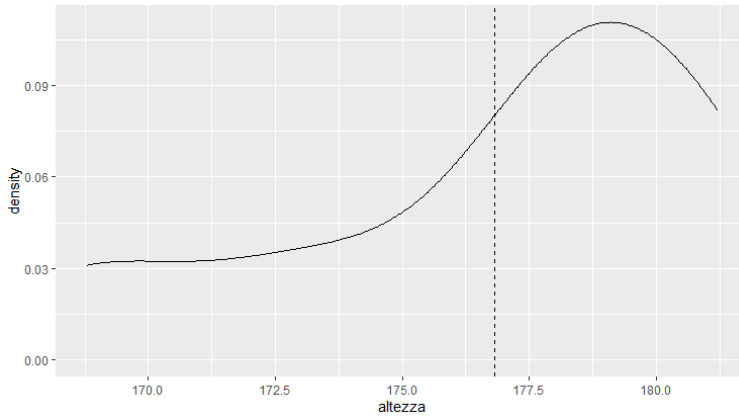


Figura 5.4 - Distribuzione della variabile altezza

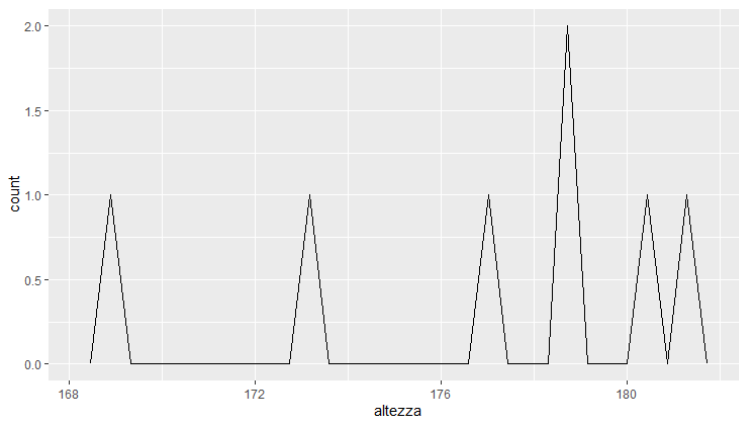


Figura 5.5 - Plot delle frequenze della variabile continua altezza

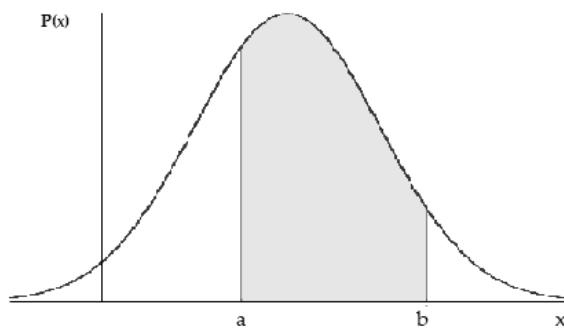


Figura 5.6 - Probability Density Function

X	P(X)	F(X)
1	1/6	1/6
2	1/6	2/6
3	1/6	3/6
4	1/6	4/6
5	1/6	5/6
6	1/6	6/6

Figura 5.7 - CDF in forma tabellare

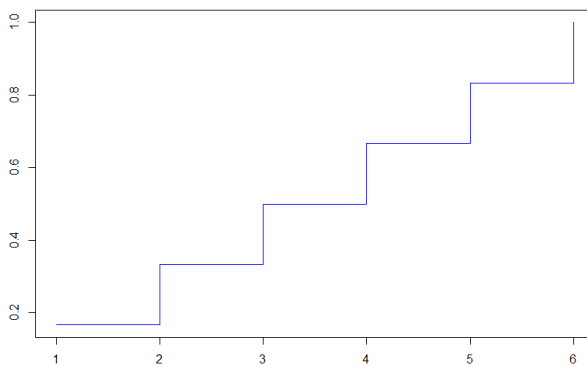


Figura 5.8 - Grafico della CDF di un lancio di dadi

6. PRINCIPALI DISTRIBUZIONI

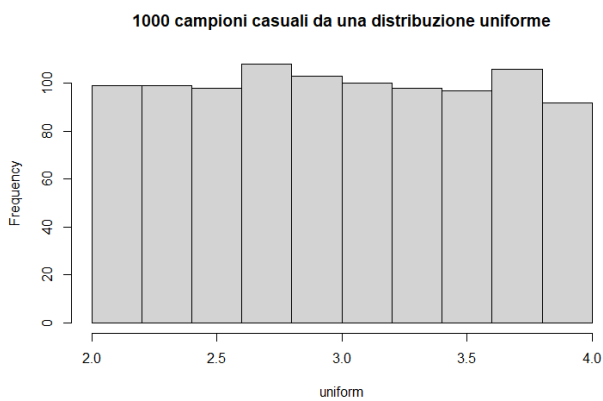


Figura 6.1 - Campioni da una distribuzione uniforme

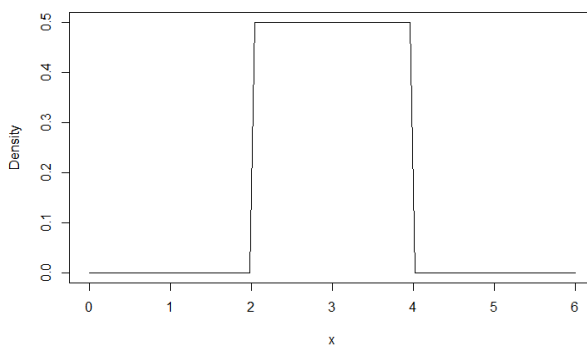


Figura 6.2 - Distribuzione uniforme

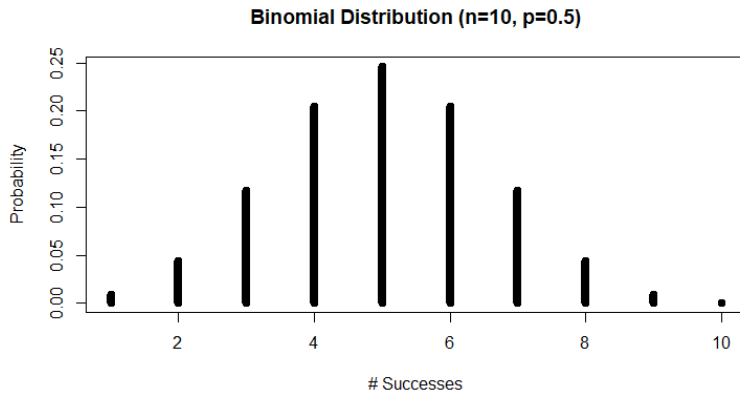


Figura 6.3- Distribuzione binomiale B(10,0.5)

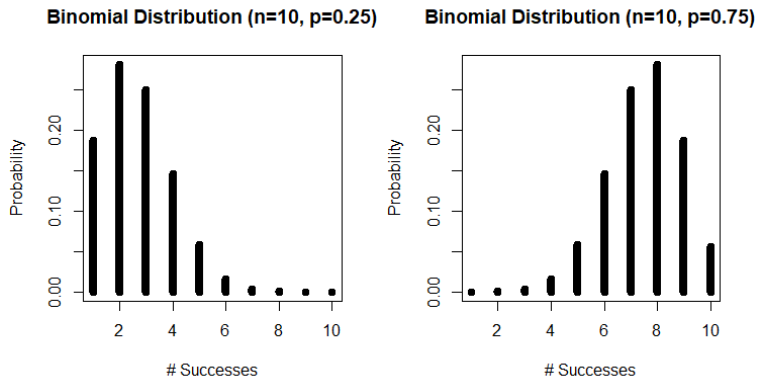


Figura 6.4 - Distribuzioni binomiali B(10,0.25) e B(10,0.75)

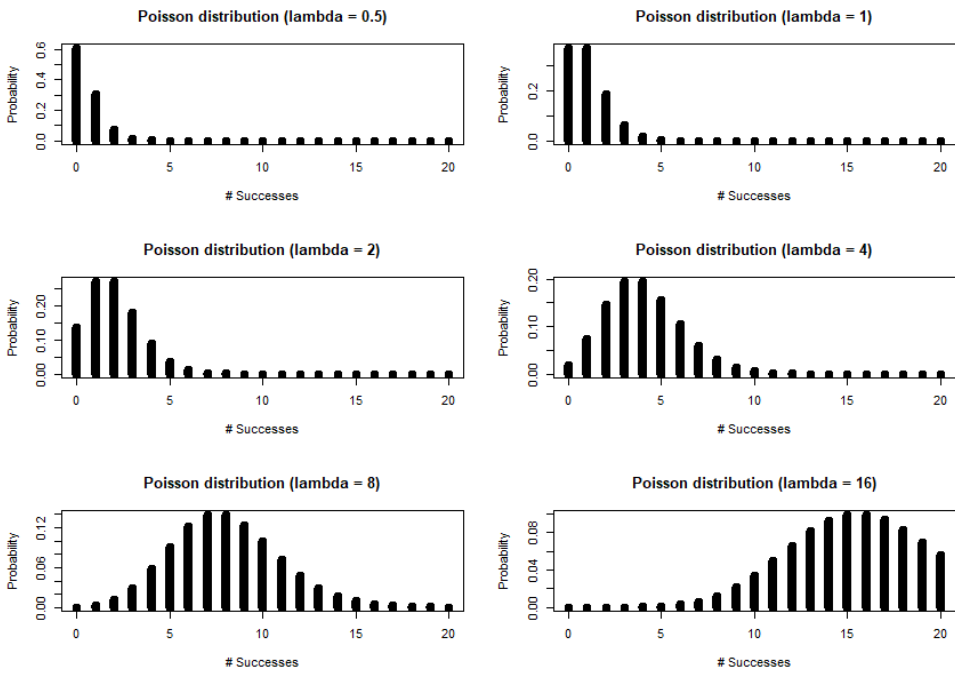


Figura 6.5 - Distribuzione di Poisson per diversi valori di lambda

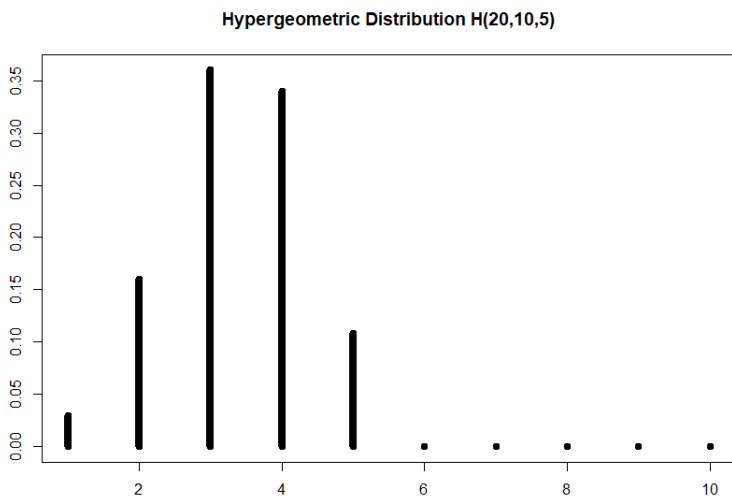


Figura 6.6 - Esempio di distribuzione Ipergeometrica $H(20,10,5)$

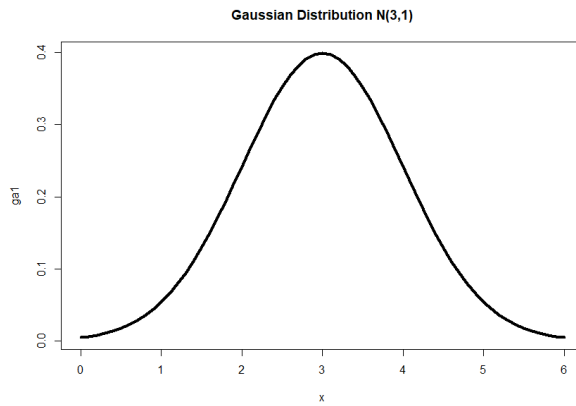


Figura 6.7 - Distribuzione normale con media $\mu = 3$ e deviazione standard $\sigma = 1$

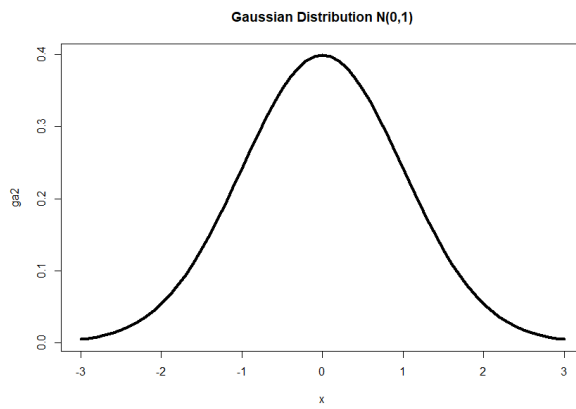


Figura 6.8 - Distribuzione normale con media $\mu=0$ e deviazione standard $\sigma = 1$

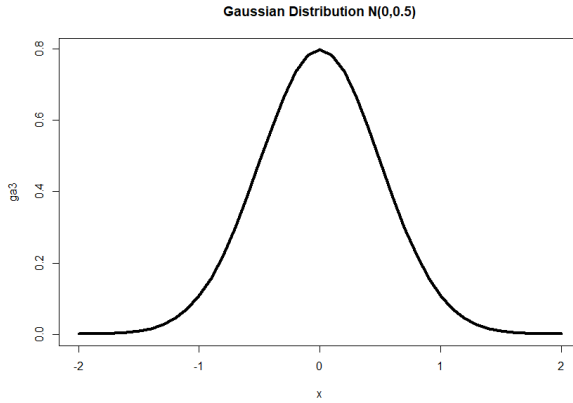


Figura 6.9 – Distribuzione normale con media $\mu=0$ e deviazione standard $\sigma = 0,5$

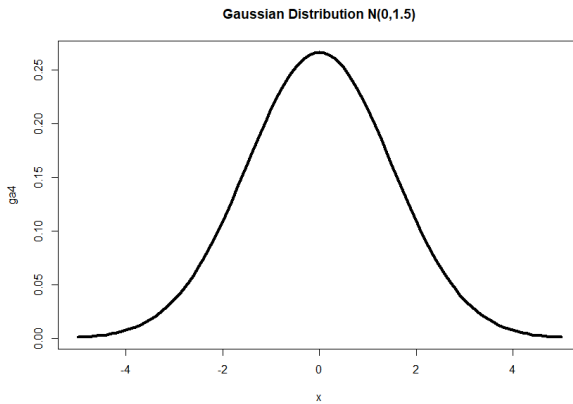


Figura 6.10 – Distribuzione normale con media $\mu=0$ e deviazione standard $\sigma = 1,5$

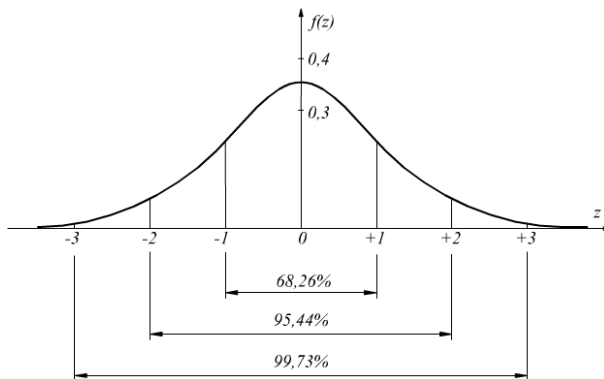


Figura 6.11 - Regola 68-95-99

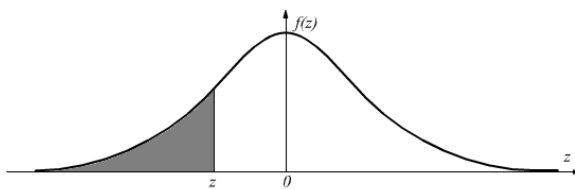


Figura 6.12 - Curva normale standard per valori negativi di z

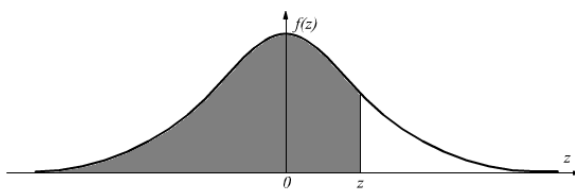


Figura 6.13 - Curva normale standard per valori positivi di z

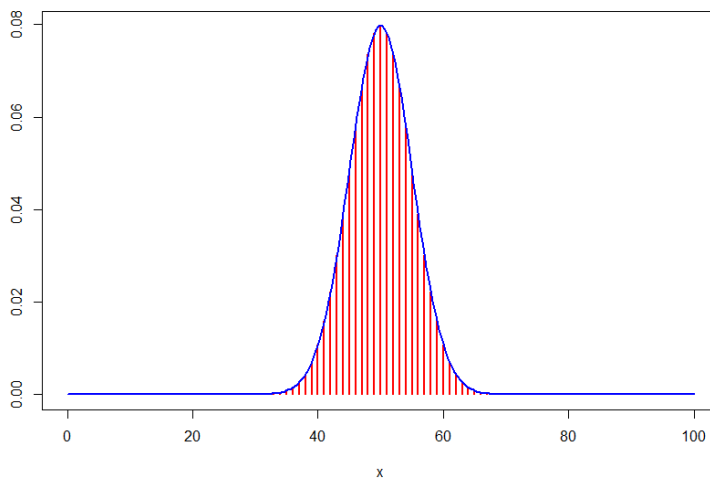


Figura 6.14 - Approssimazione della distribuzione binomiale alla normale

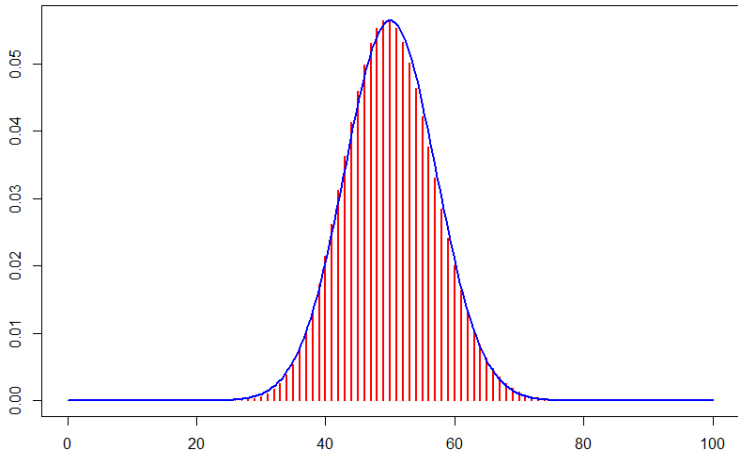


Figura 6.15 - Approssimazione di una distribuzione di Poisson con la normale

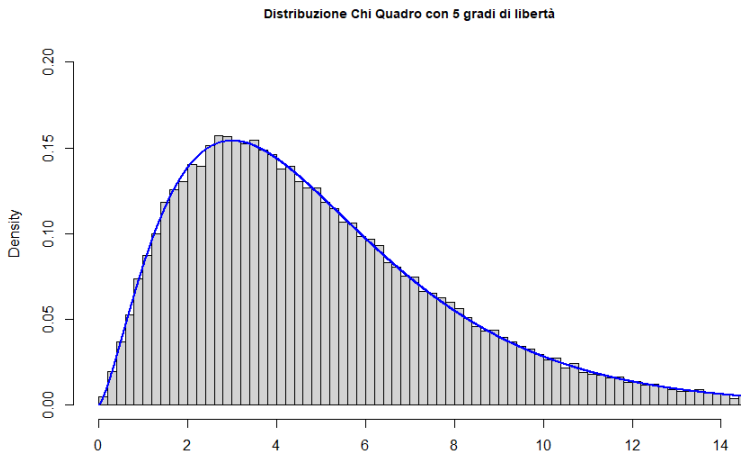


Figura 6.16 - Distribuzione Chi Quadro con 5 gradi di libertà

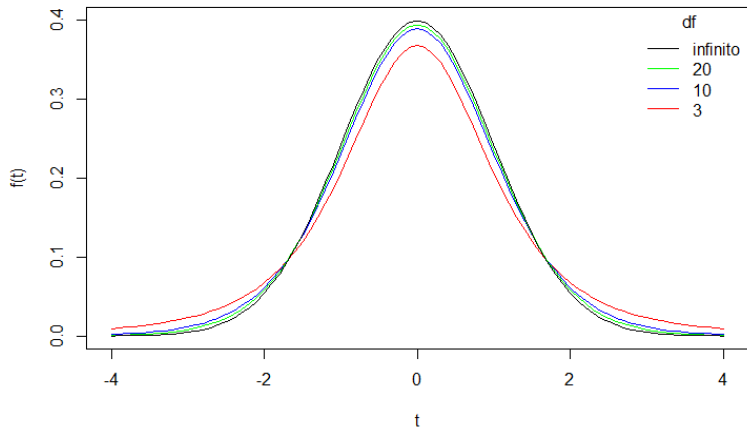


Figura 6.17 - Distribuzione t di student con diversi gradi di libertà

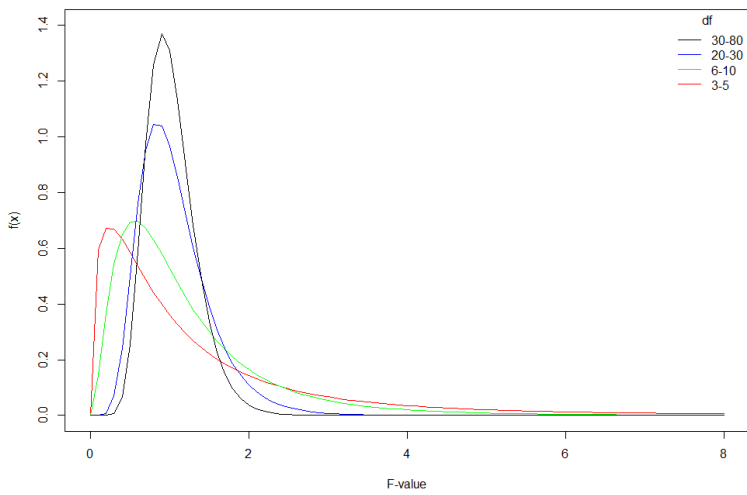


Figura 6.18 - Distribuzione F per diverse coppie di gradi di libertà

7. STATISTICA INFERENZIALE

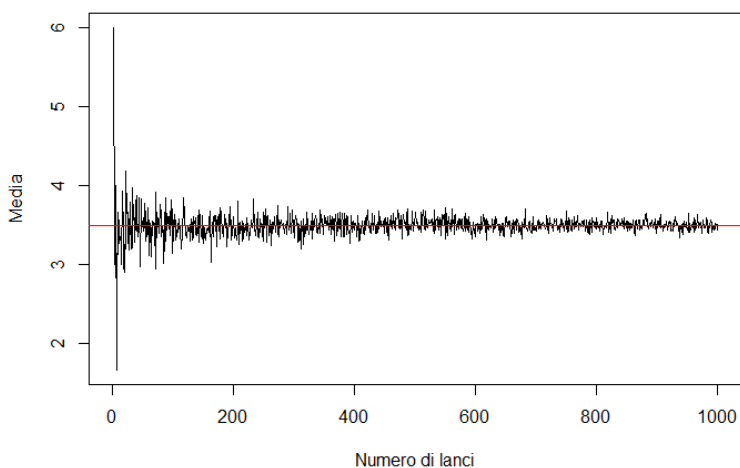


Figura 7.1 - Aumentando il numeri di campioni la media si avvicina al valore atteso

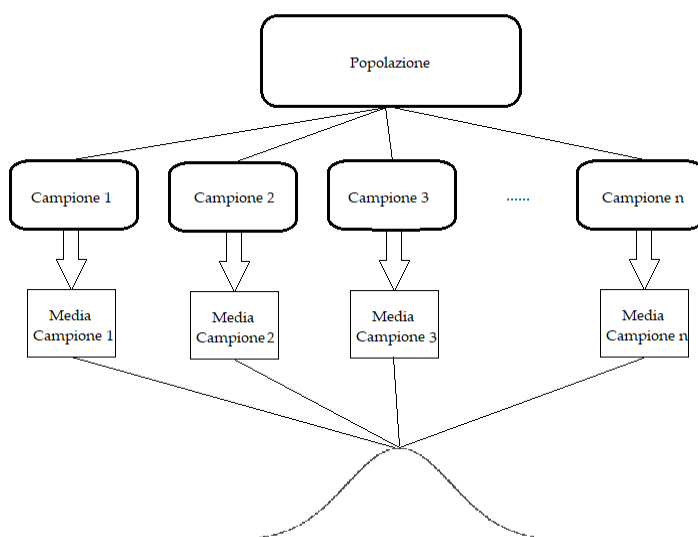


Figura 7.2 - Teorema del limite centrale

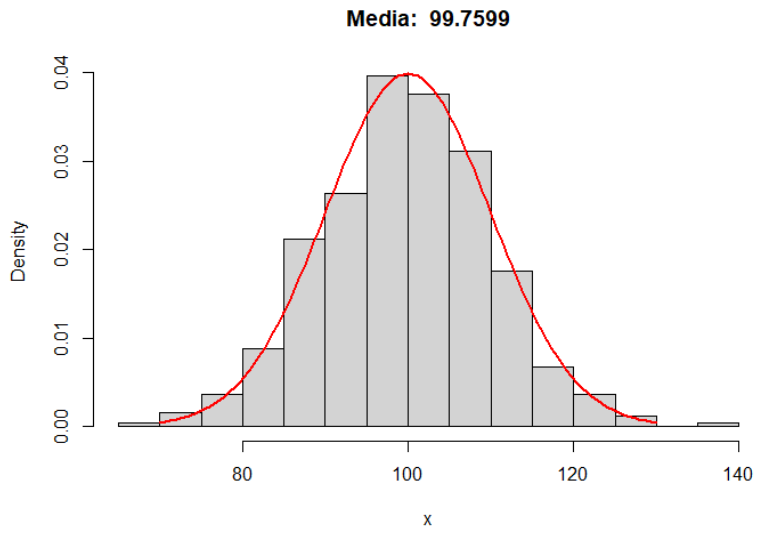


Figura 7.3 - Distribuzione dei 500 elementi estratti

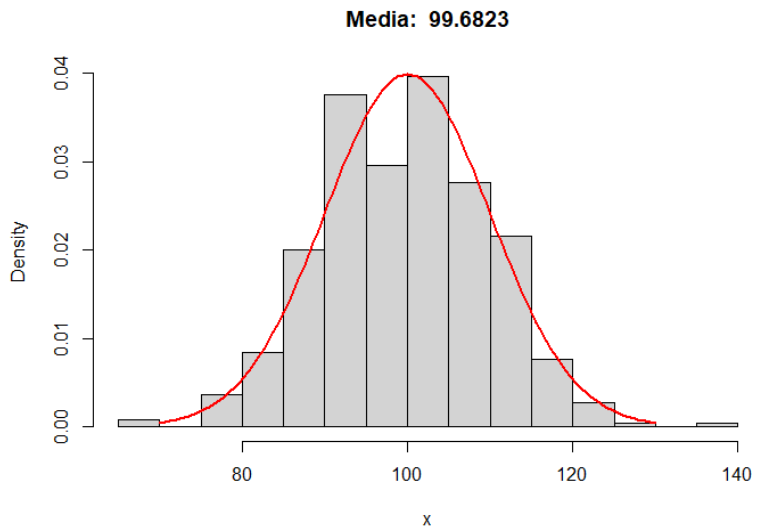


Figura 7.4 - Ulteriore estrazione di 500 elementi

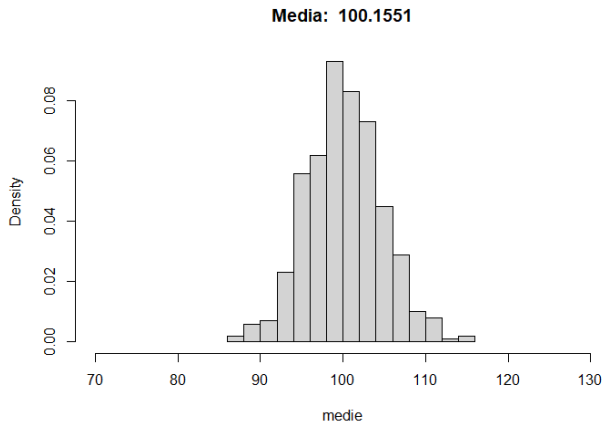


Figura 7.5 - Distribuzione delle media per $n = 5$

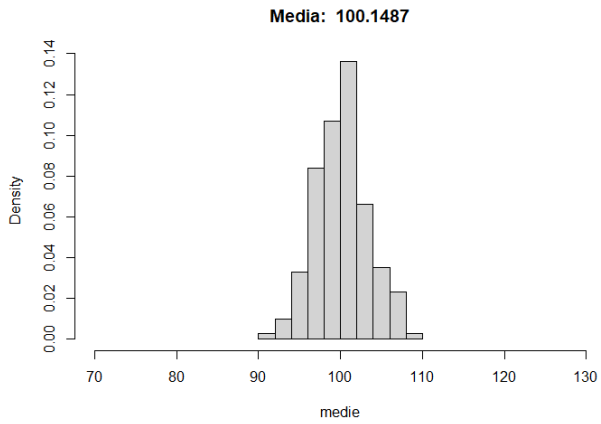


Figura 7.6 - Distribuzione delle medie per $n = 10$

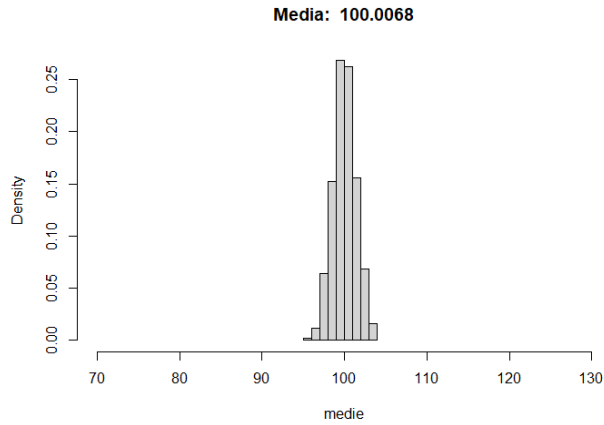


Figura 7.7 - Distribuzione delle medie per $n = 50$

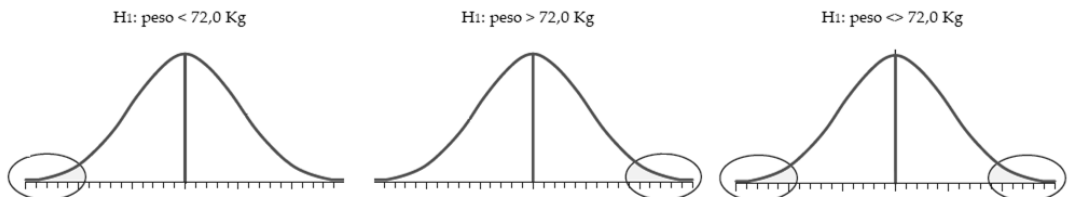


Figura 7.8 - Uso dell'ipotesi alternativa per determinare le code del test

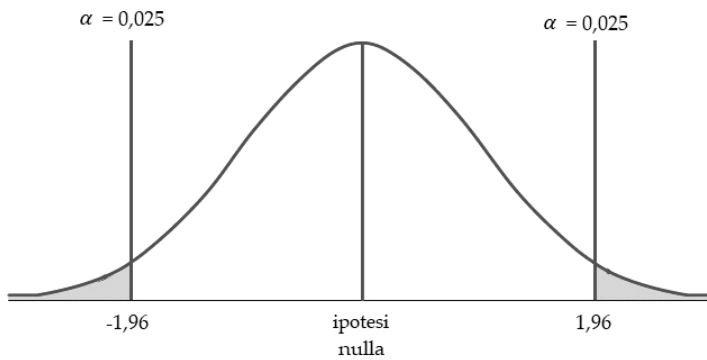


Figura 7.9 - Regione di accettazione e di rifiuto per il test a due code

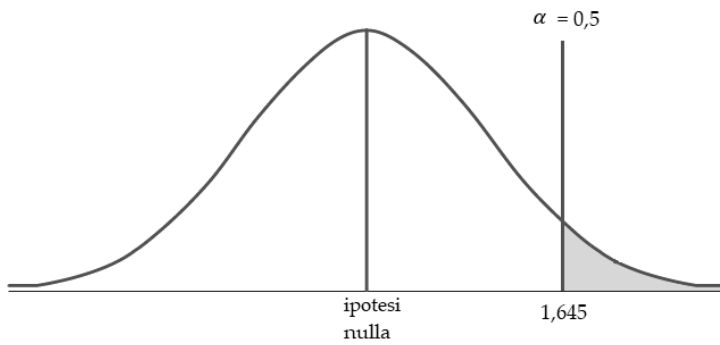


Figura 7.10 - Regione di accettazione e di rifiuto per test a una coda

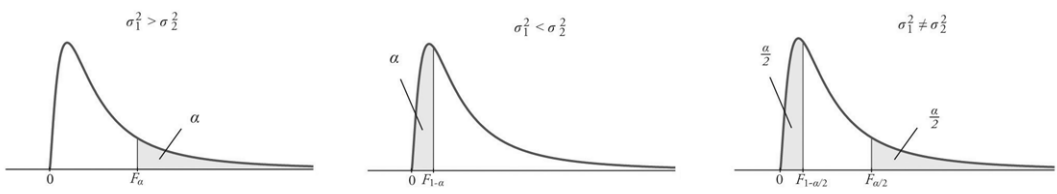


Figura 7.11 - Distribuzione F

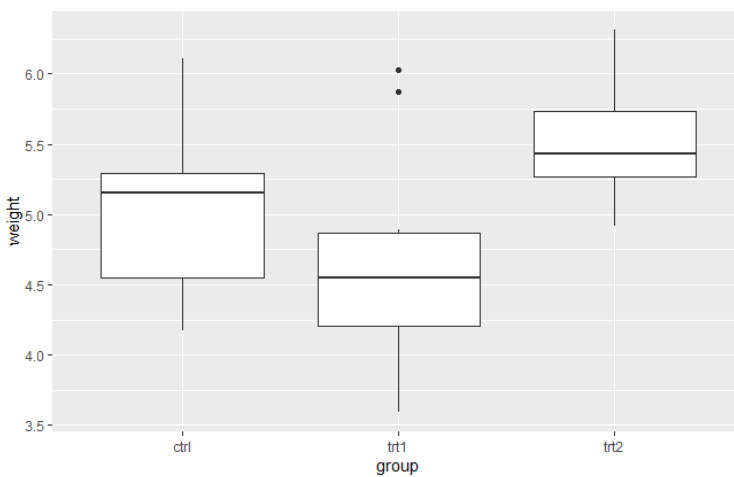


Figura 7.12 - Il boxplot della variabile weight

		SESSO		Totale
		Femmina	Maschio	
TIPO	Non Fidelizzato	105	15	120
	Fidelizzato	80	300	380
Totale		185	315	500

Figura 7.13 - Esempio di due variabili discrete

		SESSO		Totale
		Femmina	Maschio	
TIPO	Non Fidelizzato	$\frac{120 \cdot 185}{500}$	$\frac{120 \cdot 315}{500}$	120
	Fidelizzato	$\frac{380 \cdot 185}{500}$	$\frac{380 \cdot 315}{500}$	380
Totale		185	315	500

Figura 7.14 - Costruzione della tabella delle frequenze attese

		SESSO		Totale
		Femmina	Maschio	
TIPO	Non Fidelizzato	44,4	75,6	120
	Fidelizzato	140,6	239,4	380
Totale		185	315	500

Figura 7.15 - Tabella delle frequenze attese

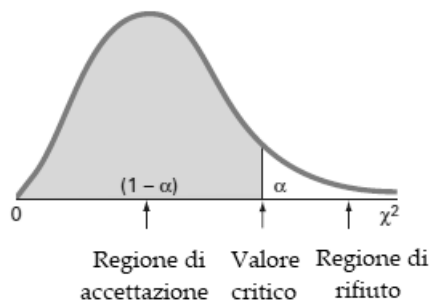


Figura 7.16 - Distribuzione Chi Quadro

<i>gl/alfa</i>	0.20	0.10	0.05	0.025	0.01	0.005	0.001	0.0005	0.0001
1	1.6424	2.7055	3.8415	5.0239	6.6349	7.8794	10.8274	12.1153	15.1343
2	3.2189	4.6052	5.9915	7.3778	9.2104	10.5965	13.8150	15.2014	18.4247
3	4.6416	6.2514	7.8147	9.3484	11.3449	12.8381	16.2660	17.7311	21.1040
4	5.9886	7.7794	9.4877	11.1433	13.2767	14.8602	18.4662	19.9977	23.5064
5	7.2893	9.2363	11.0705	12.8325	15.0863	16.7496	20.5147	22.1057	25.7507
6	8.5581	10.6446	12.5916	14.4494	16.8119	18.5475	22.4575	24.1016	27.8527
7	9.8032	12.0170	14.0671	16.0128	18.4753	20.2777	24.3213	26.0179	29.8814
8	11.0301	13.3616	15.5073	17.5345	20.0902	21.9549	26.1239	27.8674	31.8268
9	12.2421	14.6837	16.9190	19.0228	21.6660	23.5893	27.8767	29.6669	33.7247
10	13.4420	15.9872	18.3070	20.4832	23.2093	25.1881	29.5879	31.4195	35.5572
11	14.6314	17.2750	19.6752	21.9200	24.7250	26.7569	31.2635	33.1382	37.3647
12	15.8120	18.5493	21.0261	23.3367	26.2170	28.2997	32.9092	34.8211	39.1306
13	16.9848	19.8119	22.3620	24.7356	27.6882	29.8193	34.5274	36.4768	40.8735
14	18.1508	21.0641	23.6848	26.1189	29.1412	31.3194	36.1239	38.1085	42.5752
15	19.3107	22.3071	24.9958	27.4884	30.5780	32.8015	37.6978	39.7173	44.2596
16	20.4651	23.5418	26.2962	28.8453	31.9999	34.2671	39.2518	41.3077	45.9255
17	21.6146	24.7690	27.5871	30.1910	33.4087	35.7184	40.7911	42.8808	47.5591
18	22.7595	25.9894	28.8693	31.5264	34.8052	37.1564	42.3119	44.4337	49.1853
19	23.9004	27.2036	30.1435	32.8523	36.1908	38.5821	43.8194	45.9738	50.7873
20	25.0375	28.4120	31.4104	34.1696	37.5663	39.9969	45.3142	47.4977	52.3832

Figura 7.17 - Tabella dei valori attesi della distribuzione chi quadro

8. IL CAMPIONAMENTO

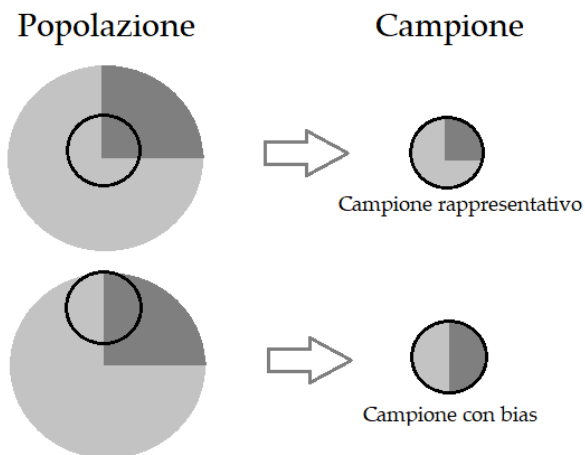


Figura 8.1 - Campionamento corretto e con bias

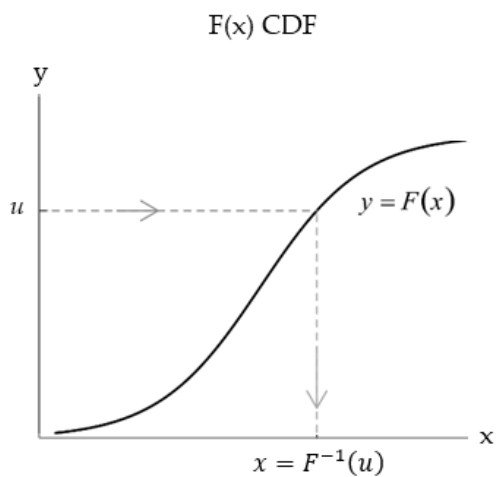


Figura 8.2 - Inverse Transform Sampling

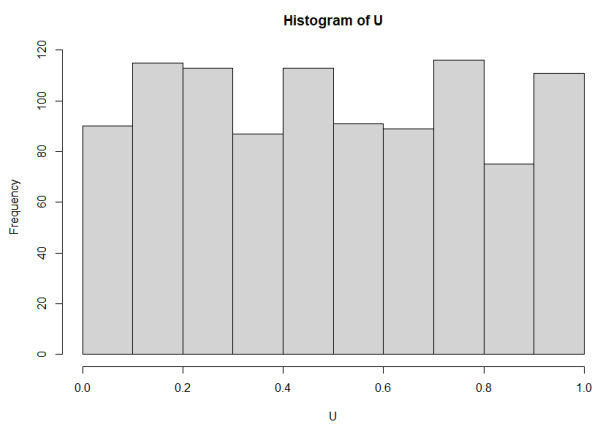


Figura 8.3 - Istogramma della distribuzione U

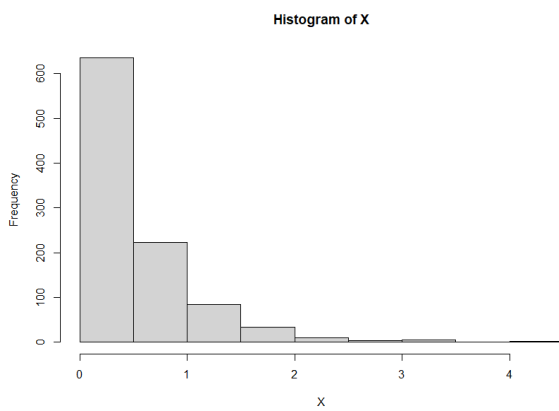


Figura 8.4 - La nuova distribuzione

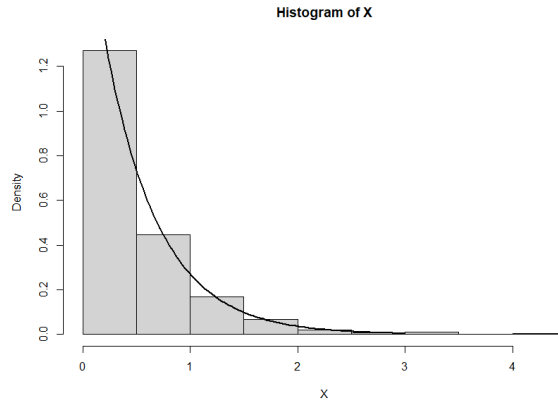


Figura 8.5 - Andamento esponenziale della distribuzione

9. LA PREPARAZIONE DEI DATI

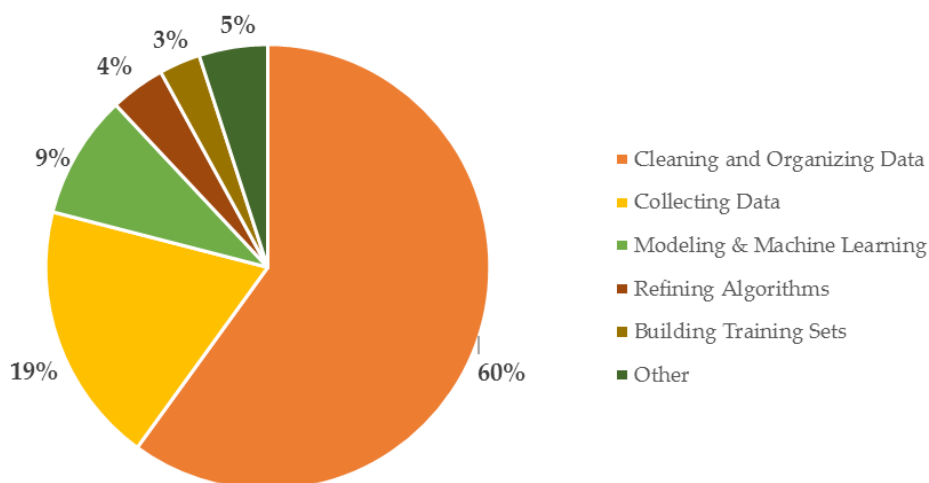


Figura 9.1 - Tempo medio speso in un progetto

Cliente	Sesso	Stato	Lavoro	Acquisti	Spesa	Fidelizzato
1	Donna	Sposato	Impiegato	10	1000	SI
2	Uomo	Divorziato	Libero Professionista	5	500	SI
3	Uomo	Single	Impiegato	2	200	NO
4	Donna	Single	Impiegato	1	100	NO
5	Donna	Single	Imprenditore	1	100	NO
6	Uomo	Divorziato	Impiegato	8	800	SI
7	Donna	Divorziato	Impiegato	0	0	NO
8	Uomo	Sposato	Impiegato	13	1300	SI
9	Uomo	Sposato	Imprenditore	2	200	NO
10	Uomo	Sposato	Libero Professionista	7	700	SI

Figura 9.2 . Esempio di dataset clienti

Cliente	Data	Prodotto	Quantità	Valore
1	06/01/2020	Prodotto A	1	100
1	08/02/2020	Prodotto B	1	100
1	15/03/2020	Prodotto A	1	100
1	16/03/2020	Prodotto C	1	100
1	03/05/2020	Prodotto C	1	100
1	04/05/2020	Prodotto B	1	100
1	06/06/2020	Prodotto A	1	100
1	31/07/2020	Prodotto A	1	100
1	24/09/2020	Prodotto C	1	100
1	15/12/2020	Prodotto B	1	100

Figura 9.3 - Esempio di transazioni di acquisto di un cliente

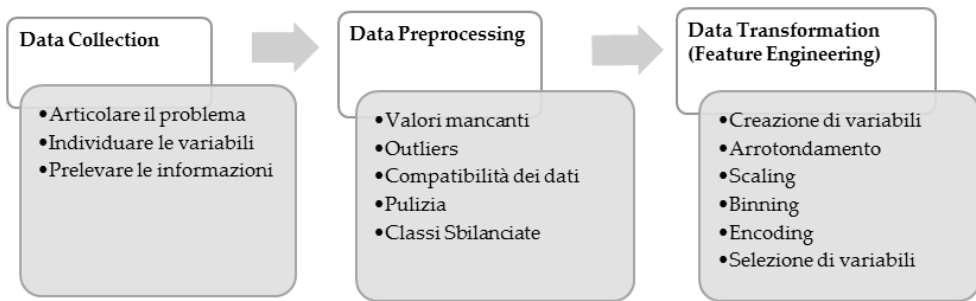


Figura 9.4 - Le fasi del processo di preparazione dei dati

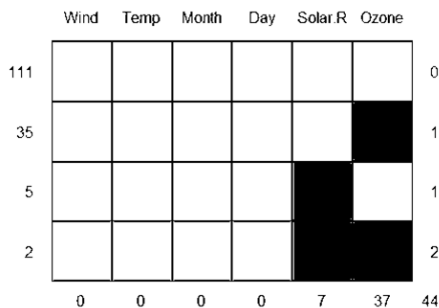


Figura 9.5 - Mappa dei valori mancanti

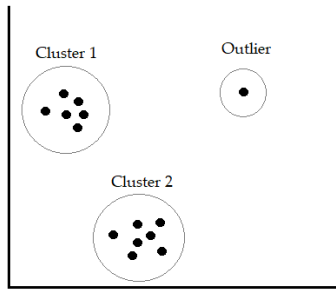


Figura 9.6 - Clustering per l'identificazione di outliers

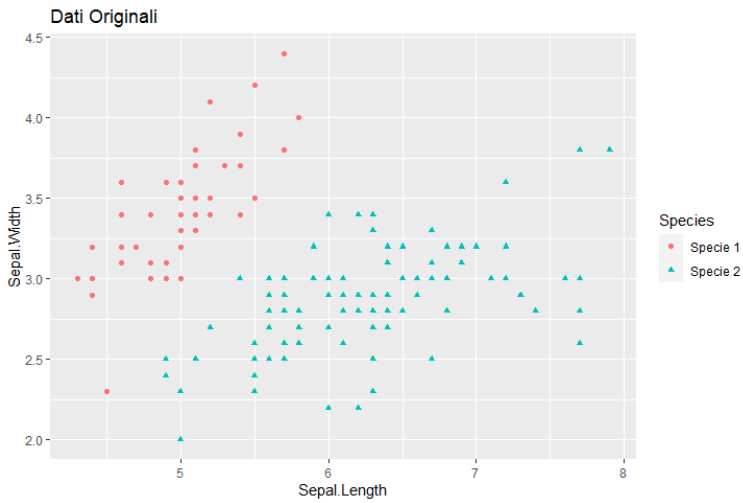


Figura 9.7 - Il dataset prima del bilanciamento

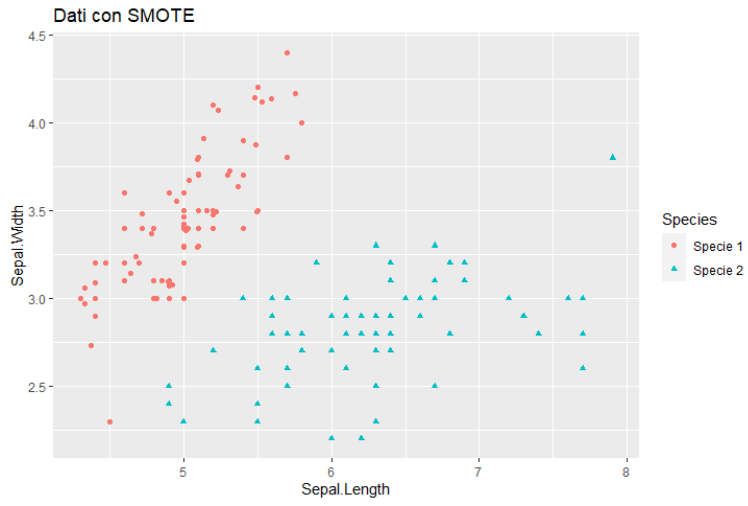


Figura 9.8 - Il dataset dopo il bilanciamento

10. FEATURE ENGINEERING

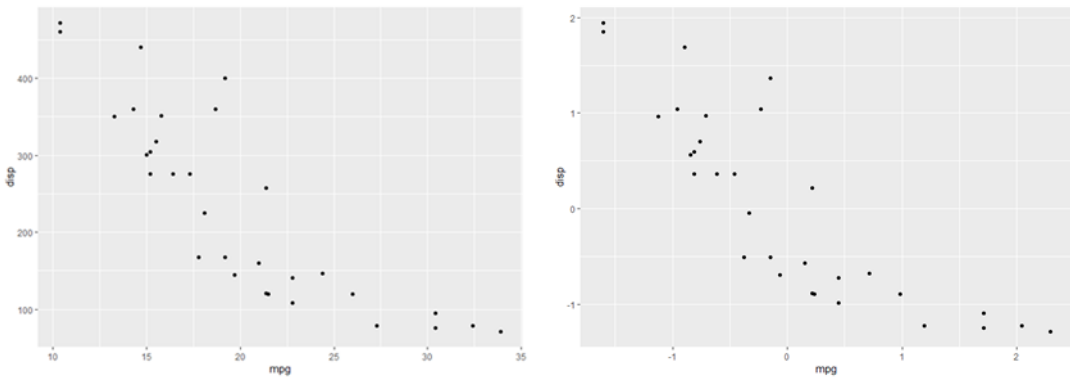


Figura 10.1 - Le variabili prima e dopo lo scaling

Eta	Bin
12	15
16	15
18	15
22	26
25	26
30	26
35	41
42	41
45	41
48	51
50	51
55	51

Figura 10.2 - Equi-depth binning

Eta	Bin
12	[12,22]
16	[12,22]
18	[12,22]
22	[12,22]
25	[23,33]
30	[23,33]
35	[34,44]
42	[34,44]
45	[45,55]
48	[45,55]
50	[45,55]
55	[45,55]

Figura 10.3 - Equi-width binning

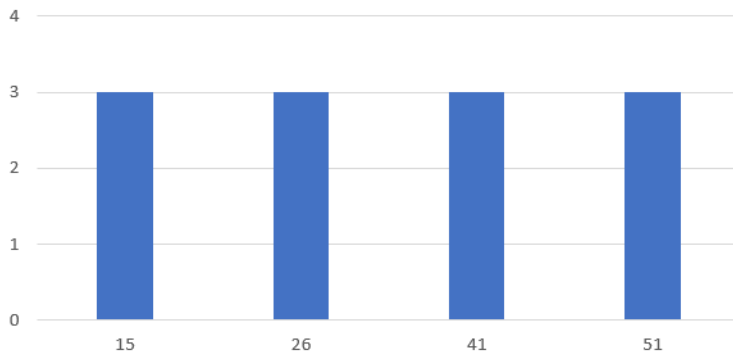


Figura 10.4 - Equi-depth binning

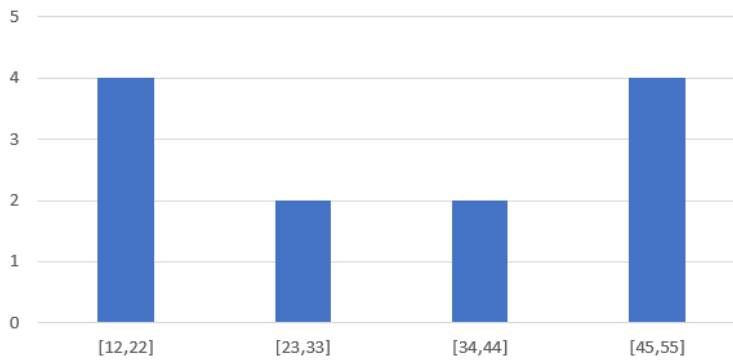


Figura 10.5 - Equi-width binning

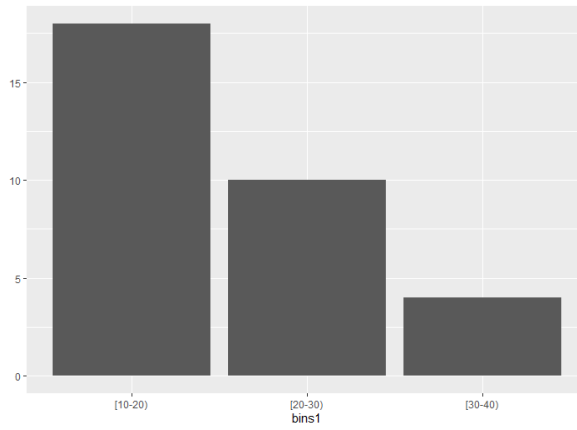


Figura 10.6 - Plot dei bins

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb	bins
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4	[20-30]
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4	[20-30]
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1	[20-30]
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1	[20-30]
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2	[10-20]
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1	[10-20]
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4	[10-20]
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2	[20-30]
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2	[20-30]
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4	[10-20]
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4	[10-20]

Figura 10.7 - Il dataset con la colonna dei bins

11. MACHINE LEARNING

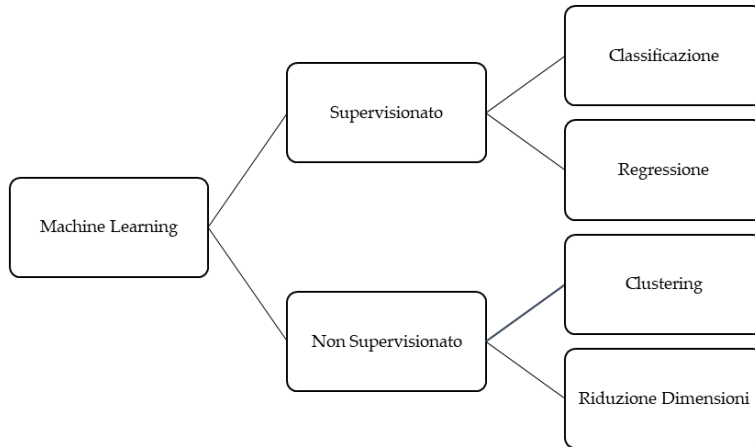


Figura 11.1 – Alcuni tipi di algoritmi

N.	Colore	Forma	Dimensioni	Risultato
1	Rosso	Rotonda	Piccola	Commestibile
2	Verde	Rotonda	Piccola	Commestibile
3	Giallo	Allungata	Piccola	Commestibile

Esempio Feature Classe

Figura 11.2 - Elementi necessari alla fase di addestramento

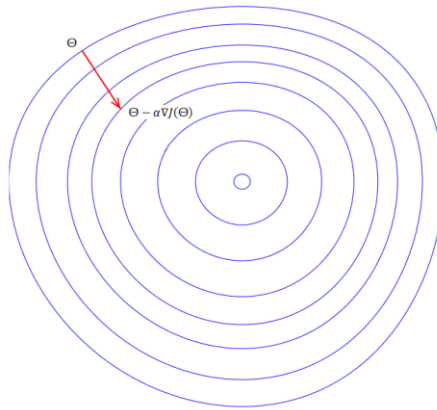


Figura 11.3 - Discesa del gradiente

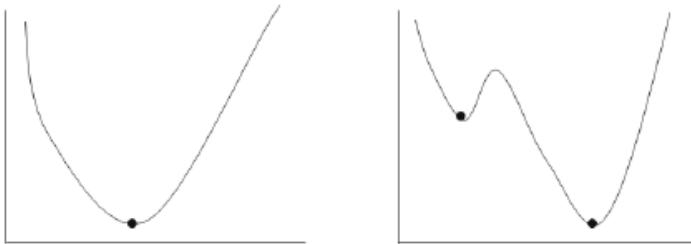


Figura 11.4 - Esempio di funzione convessa e funzione non convessa

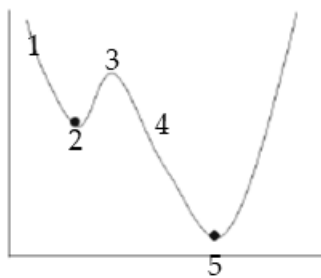


Figura 11.5 - Funzione non convessa

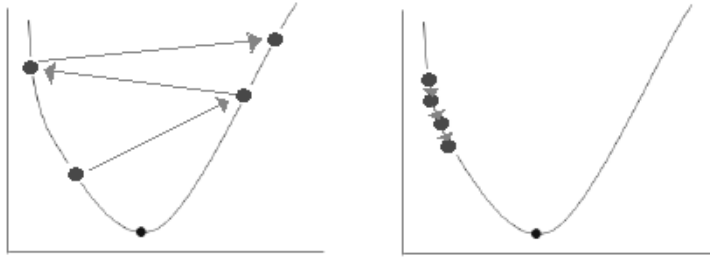


Figura 11.6 - Effetto del learning rate



Figura 11.7 - Processo di identificazione degli iperparametri

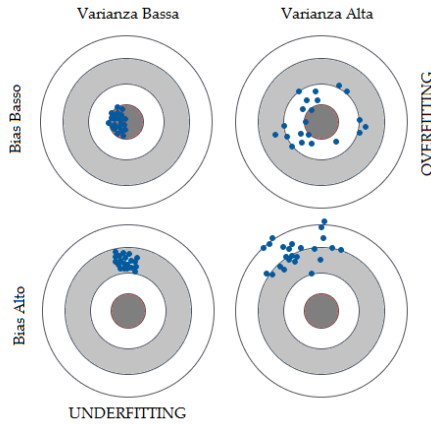


Figura 11.8 - Effetto della varianza e del bias sul fitting

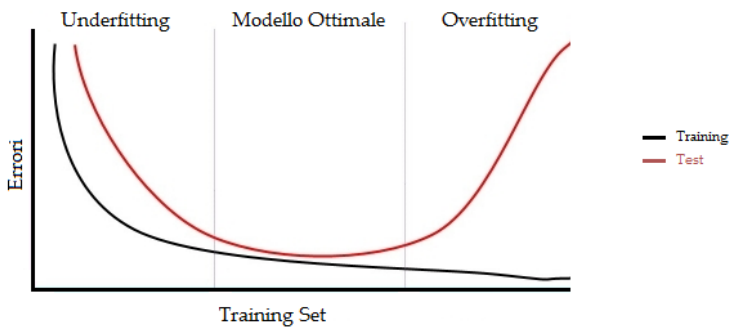


Figura 11.9 - Esempio di curva di apprendimento

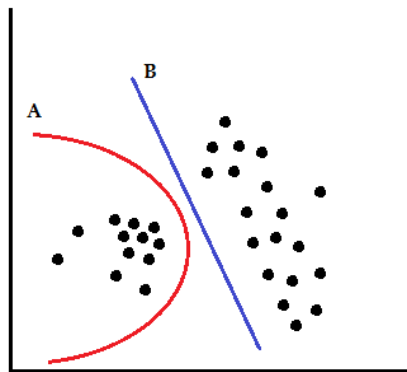


Figura 11.10 - Rasoio di Occam: la soluzione B è da preferire perché è la più semplice

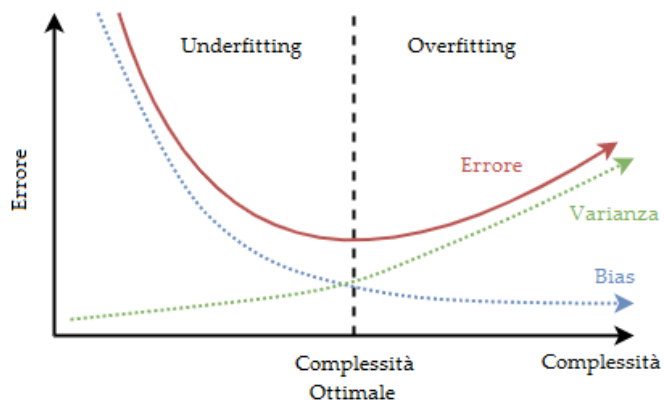


Figura 11.11 - Bias-Variance Trade-Off

12. LA VALIDAZIONE

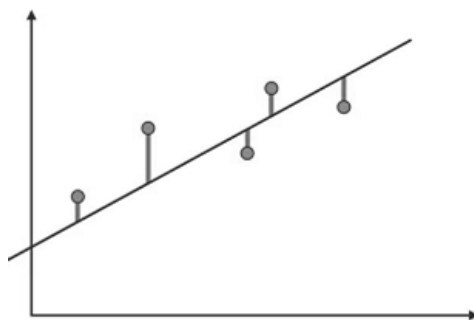


Figura 12.1 - Errore nella regressione

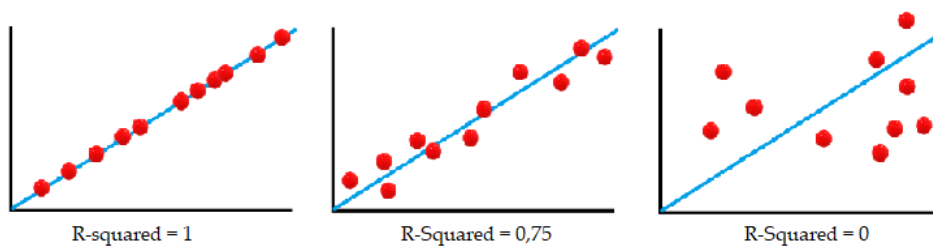


Figura 12.2 - Interpretazione di R-squared

		Previsione		
		SI	NO	Tot
Reale	SI	TP	FN	TP+FN
	NO	FP	TN	FP+TN
Tot		TP+FP	TN+FN	P+N

Figura 12.3 - Esempio di matrice di confusione

Decile (% campioni)	TPR	TPR	Lift
	Modello Base	Nostro Modello	
0	0	0	0,00
10	10	60	6,00
20	20	75	3,75
30	30	82	2,73
40	40	88	2,20
50	50	92	1,84
60	60	95	1,58
70	70	98	1,40
80	80	100	1,25
90	90	100	1,11
100	100	100	1,00

Figura 12.4 - Tabella per il Gain Chart

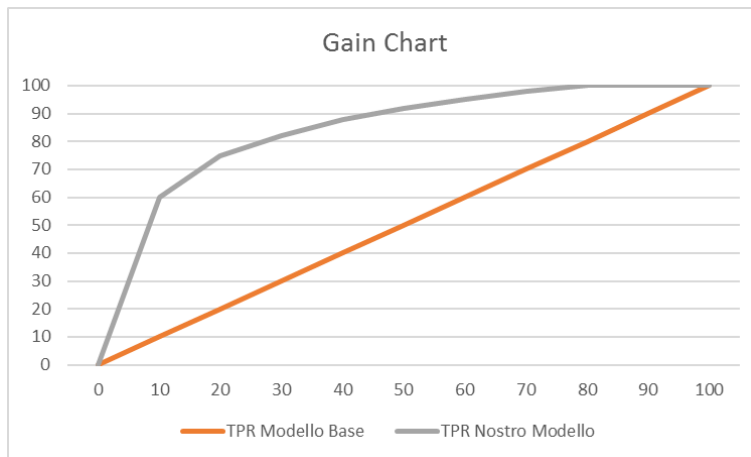


Figura 12.5 - Gain Chart

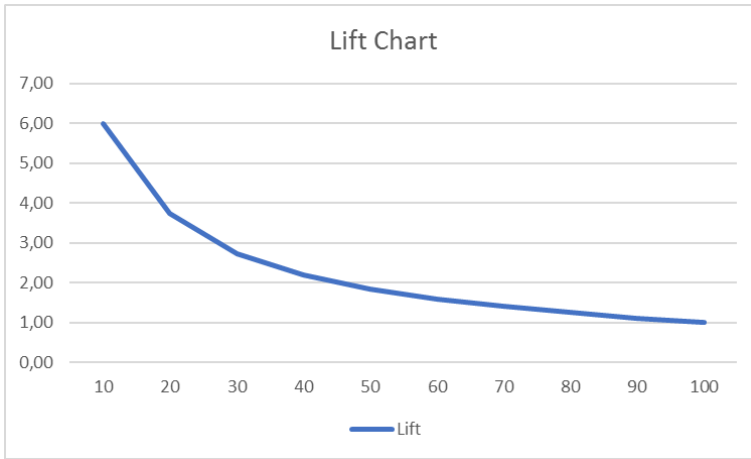


Figura 12.6 - Lift Chart

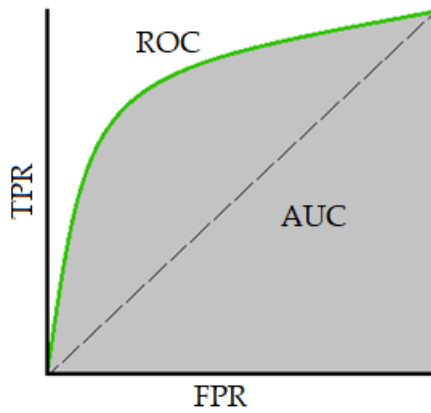


Figura 12.7 - Esempio di curva ROC

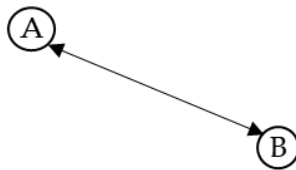


Figura 12.8 - Distanza Euclidea

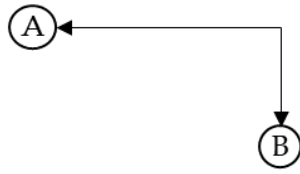


Figura 12.9 - Distanza Manhattan

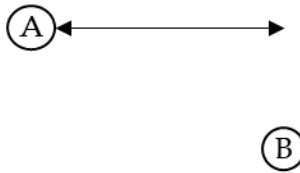


Figura 12.10 - Distanza di Lagrange

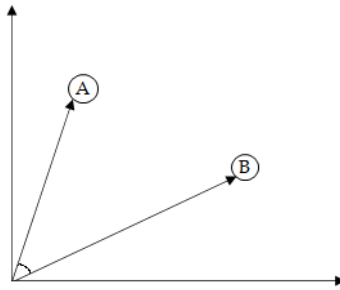


Figura 12.11 - La distanza coseno

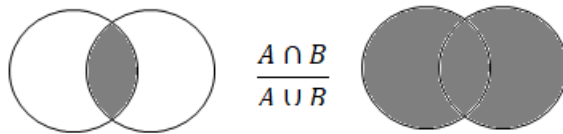


Figura 12.12 - Distanza di Jaccard

A	1	0	1	1	0	0
B	1	1	1	0	0	0

Figura 12.13 - Distanza di Hamming

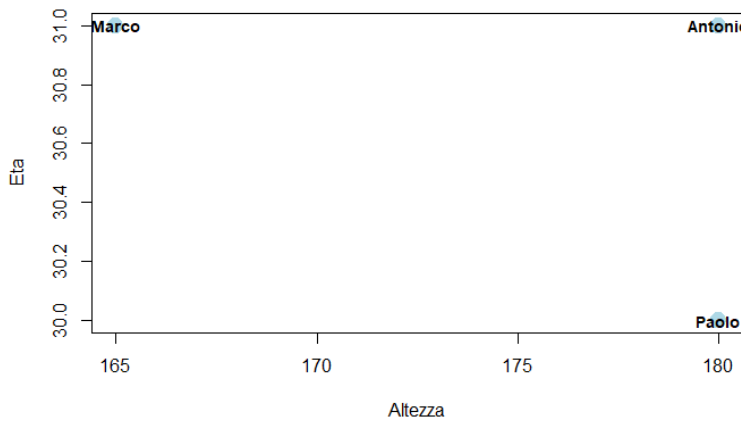


Figura 12.14 - Grafico della matrice x

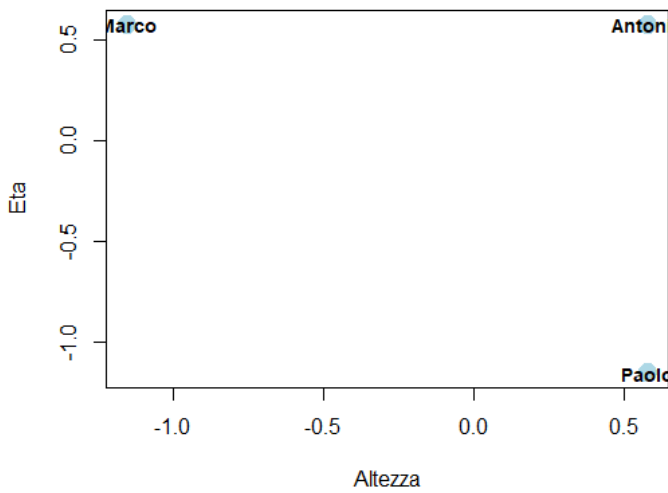


Figura 12.15 - Grafico delle variabili scalate



Figura 12.16 - Resubstitution Validation



Figura 12.17 - Hold-out Validation

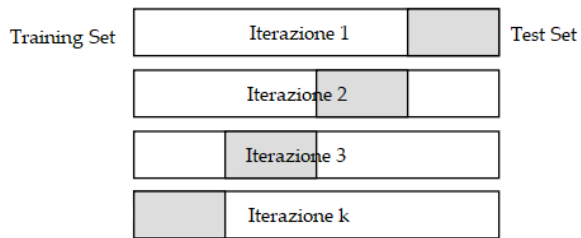


Figura 12.18 - K-fold Cross Validation

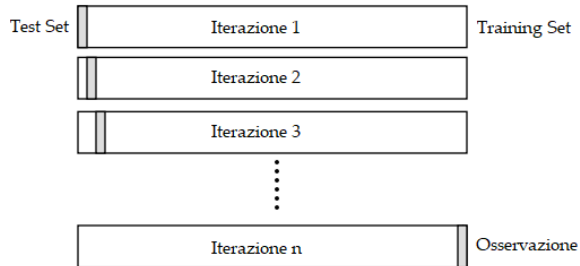


Figura 12.19 - Leave One Out Cross Validation

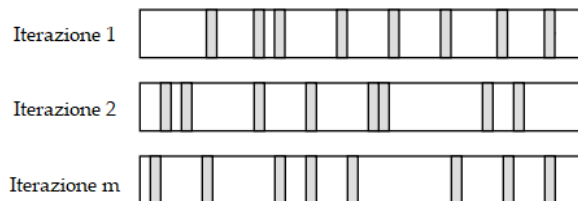


Figura 12.20 - Random Subsampling Validation

13. GLI ALGORITMI DI REGRESSIONE

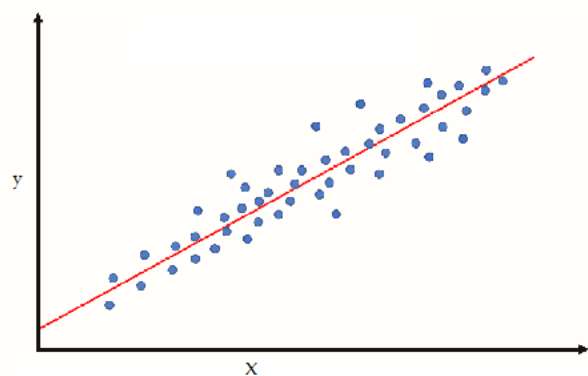


Figura 13.1 - Esempio di regressione lineare

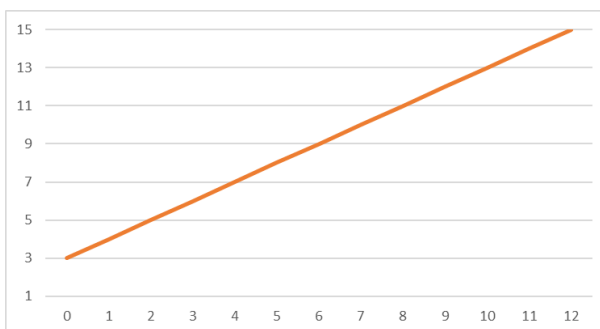


Figura 13.2 - Grafico del peso di un neonato per mese

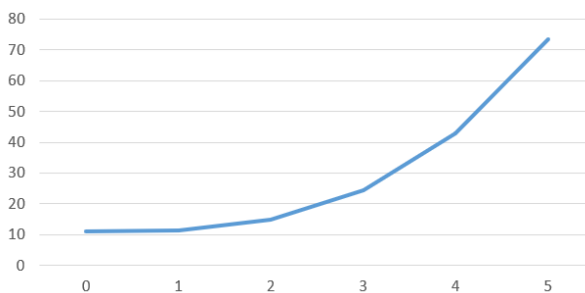


Figura 13.3 - Esempio di relazione non lineare

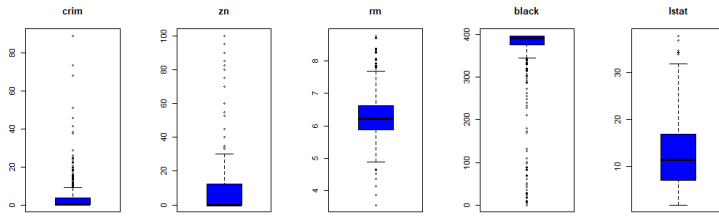


Figura 13.4 - Boxplot di alcune variabili

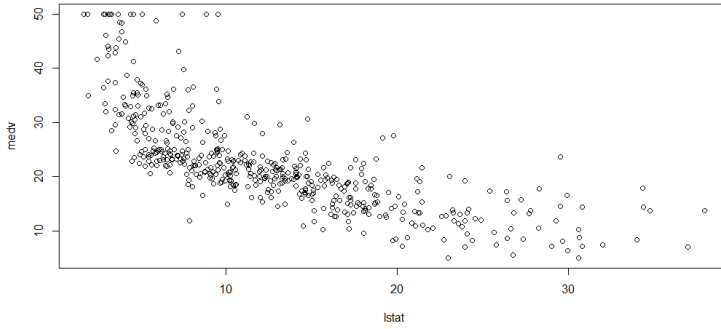


Figura 13.5 - Relazione tra la feature lstat e la variabile di output medv

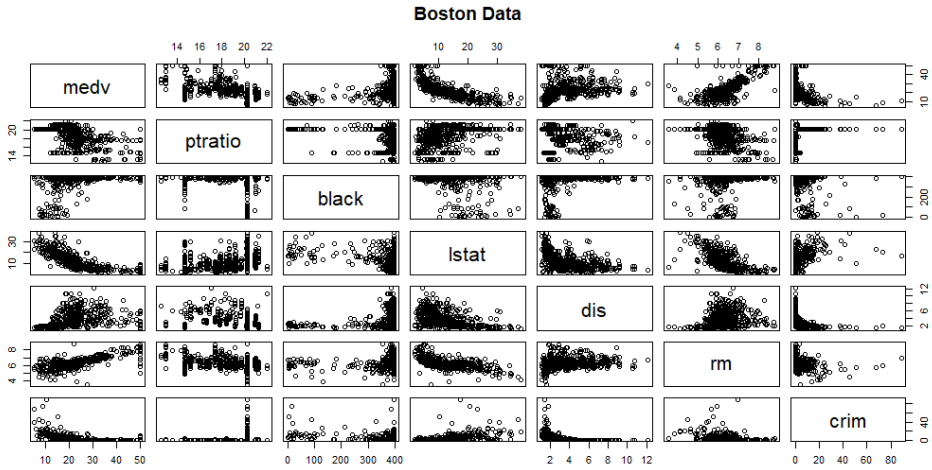


Figura 13.6 - Relazione tra tutte le variabili

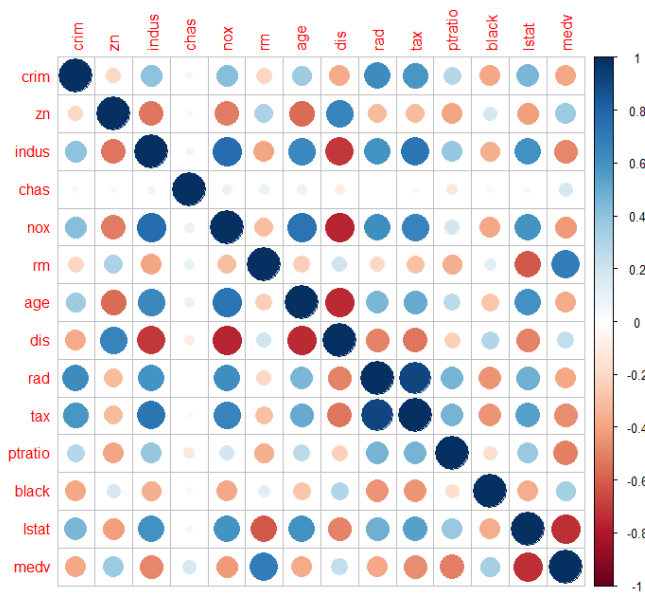


Figura 13.7 - Grafico delle correlazioni tra le variabili

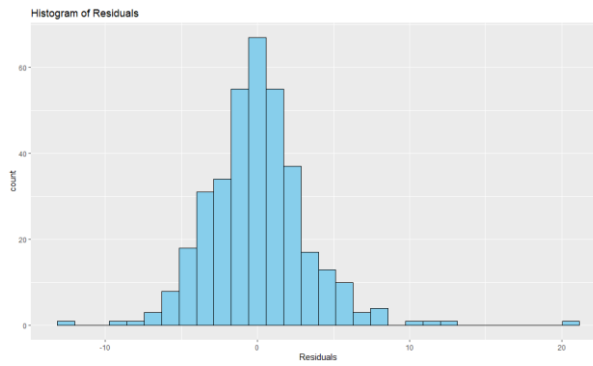


Figura 13.8 - Distribuzione dei residui

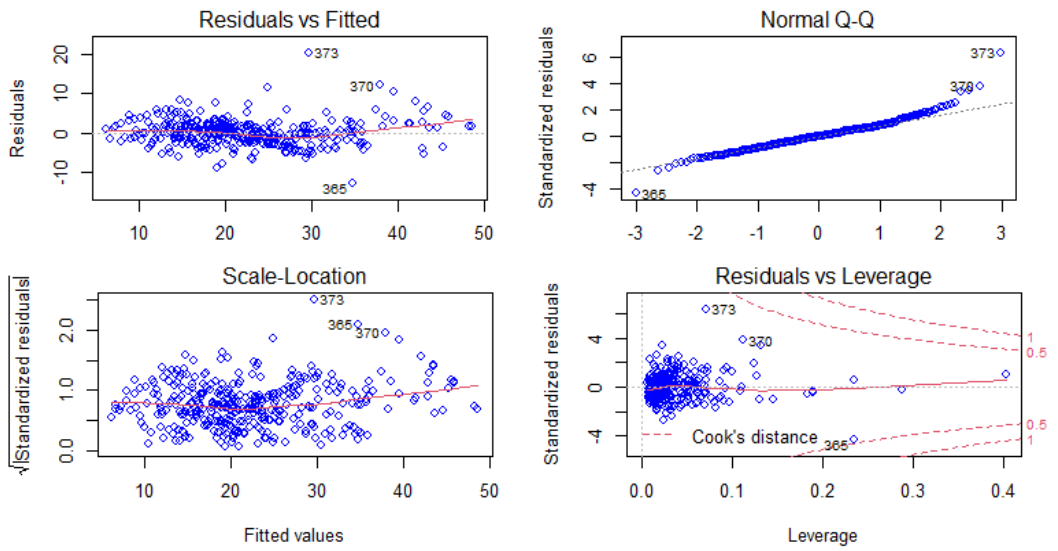


Figura 13.9 - Plot del modello

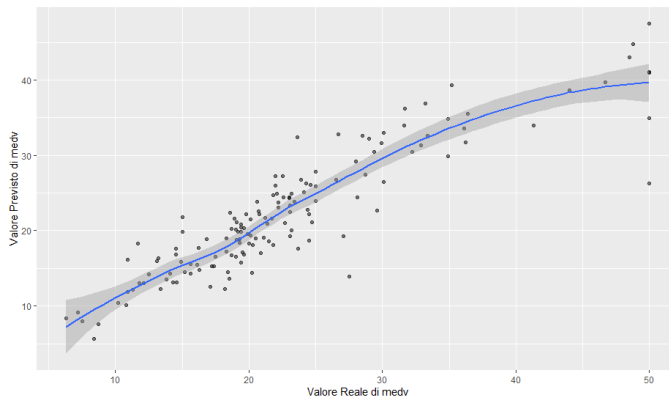


Figura 13.10 - Confronto tra la variabile predetta e quella reale

	AtBat	Hits	HmRun	Runs	RBI	Walks	Years	CAtBat	CHits	ChmRun	CRuns	CRBI	CWalks	League	Division	PutOuts	Assists	Errors	Salary	NewLeague
-Andy Allanson	293	66	1	30	29	14	1	293	66	1	30	29	14	A	E	446	33	20	NA	A
-Alan Ashby	315	81	7	24	38	39	14	3449	835	69	321	414	375	N	W	632	43	10	475,000	N
-Alvin Davis	479	130	18	66	72	76	3	1624	457	63	224	266	263	A	W	580	82	14	480,000	A
-Andre Dawson	496	141	20	65	78	37	11	5628	1575	225	828	838	354	N	E	200	11	3	500,000	N
-Andres Galarraga	321	87	10	39	42	30	2	396	101	12	48	46	33	N	E	805	40	4	91,500	N
-Alfredo Griffin	594	169	4	74	51	35	11	4408	1133	19	501	336	194	A	W	282	421	25	750,000	A
-Al Newman	185	37	1	23	8	21	2	214	42	1	30	9	24	N	E	76	127	7	70,000	A
-Argenis Salazar	298	73	0	24	24	7	3	509	108	0	41	37	12	A	W	121	283	9	100,000	A
-Andres Thomas	323	81	6	26	32	8	2	341	86	6	32	34	8	N	W	143	290	19	75,000	N
-Andre Thornton	401	92	17	49	66	65	13	5206	1332	253	784	890	866	A	E	0	0	0	1100,000	A
-Alan Trammell	574	159	21	107	75	59	10	4631	1300	90	702	504	488	A	E	238	445	22	517,148	A
-Alex Trevino	202	53	4	31	26	27	9	1876	467	15	192	186	161	N	W	304	45	11	512,500	N
-Andy VanSlyke	418	113	13	48	61	47	4	1512	392	41	205	204	203	N	E	211	11	7	550,000	N
-Alan Wiggins	239	60	0	30	11	22	6	1941	510	4	309	103	207	A	E	121	151	6	700,000	A
-Bill Almon	196	43	7	29	27	30	13	3231	825	36	376	290	238	N	E	80	45	8	240,000	N
-Billy Beane	183	39	3	20	15	11	3	201	42	3	20	16	11	A	W	116	0	0	NA	A
-Buddy Bell	568	158	20	89	75	73	15	8068	2273	177	1045	993	732	N	W	105	290	10	775,000	N

Figura 13.11 - Il dataset Hitters

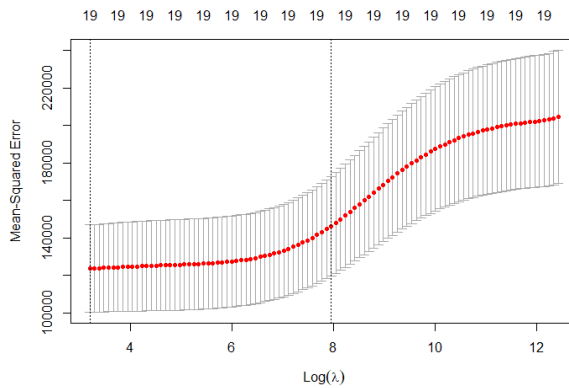


Figura 13.12 - L'MSE al variare di lambda

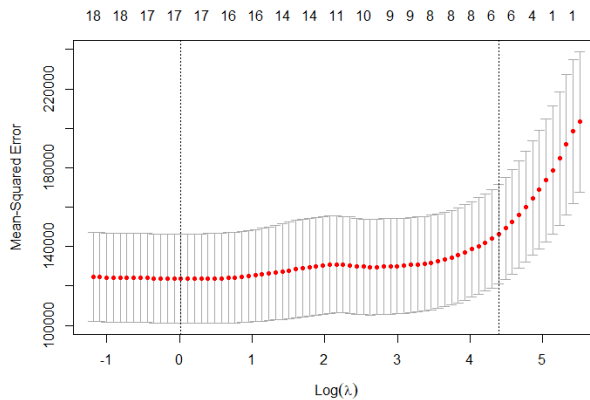


Figura 13.13 - MSE al variare di lambda per un modello LASSO

14. ALGORITMI DI CLASSIFICAZIONE

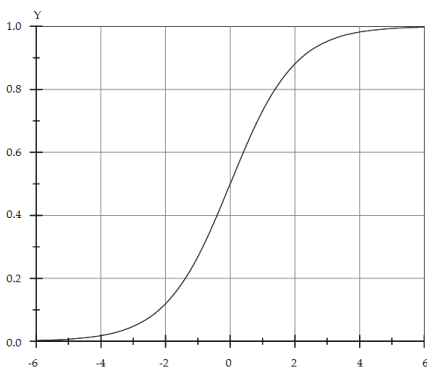


Figura 14.1 - Grafico della funzione sigmoide

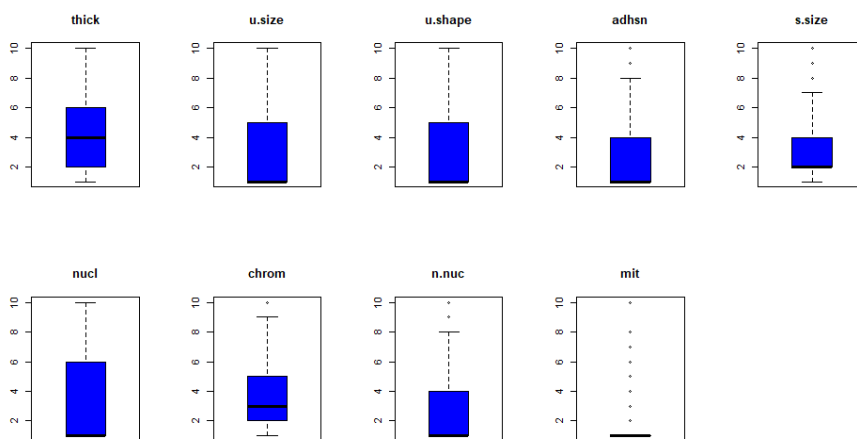


Figura 14.2 - I boxplot delle variabili

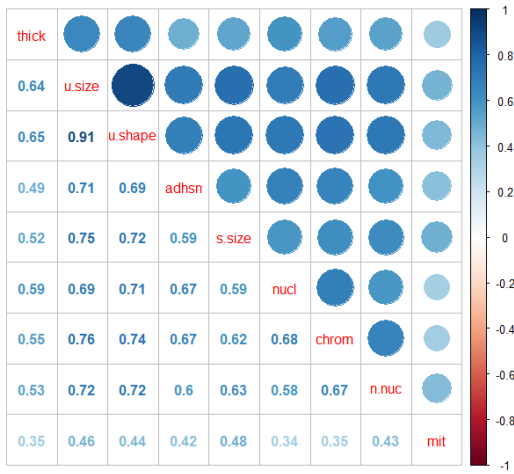


Figura 14.3 - I valori delle correlazioni

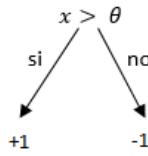


Figura 14.4 - Un nodo di un albero di classificazione

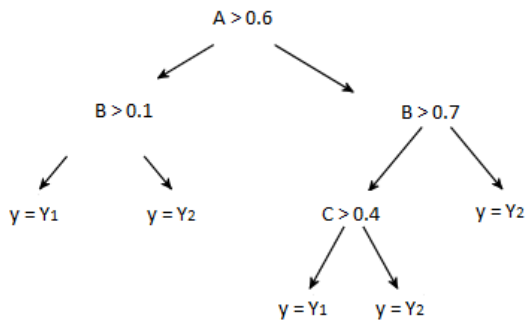


Figura 14.5 - Esempi di nodi binari

	X2	X30	X2.1	X1	X2.2	X2.3	X2.4	X2.5	X1.1	X2.6	X2.7	X2.8	X2.9	X2.10	X1.00	X85	X18	X4.0	X	X1.2	
1	2	50	1	1	2		2	2	2	1	2	2	2	2	2	0.9	135	42	3.5	NA	1
2	2	78	1	2	2		1	2	2	2	2	2	2	2	2	0.7	96	32	4.0	NA	1
3	2	31	1	NA	1	2	2	2	2	2	2	2	2	2	2	0.7	46	52	4.0	80	1
4	2	34	1	2	2	2	2	2	2	2	2	2	2	2	2	1.0	NA	200	4.0	NA	1
5	2	34	1	2	2	2	2	2	2	2	2	2	2	2	2	0.9	95	28	4.0	75	1
6	1	51	1	1	2	1	2	1	2	2	1	1	2	2	2	NA	NA	NA	NA	NA	1
7	2	23	1	2	2	2	2	2	2	2	2	2	2	2	2	1.0	NA	NA	NA	NA	1
8	2	39	1	2	2	1	2	2	2	1	2	2	2	2	2	0.7	NA	48	4.4	NA	1
9	2	30	1	2	2	2	2	2	2	2	2	2	2	2	2	1.0	NA	120	3.9	NA	1
10	2	39	1	1	1	2	2	2	1	1	2	2	2	2	2	1.3	78	30	4.4	85	1
11	2	32	1	2	1	1	2	2	2	1	2	1	2	2	2	1.0	59	249	3.7	54	1
12	2	41	1	2	1	1	2	2	2	1	2	2	2	2	2	0.9	81	60	3.9	52	1
13	2	30	1	2	2	1	2	2	2	1	2	2	2	2	2	2.2	57	144	4.9	78	1
14	2	17	1	1	1	2	2	2	2	2	2	2	2	2	2	NA	NA	NA	NA	NA	1

Figura 14.6 - Il dataset originale

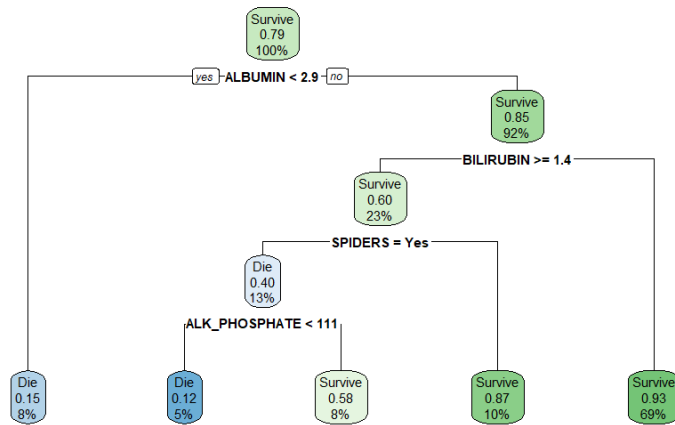


Figura 14.7 - L'albero del dataset hepatitis

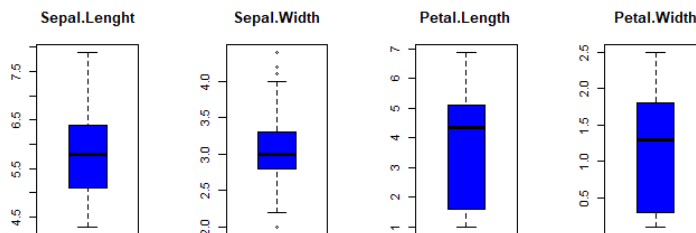


Figura 14.8 - Il boxplot delle features

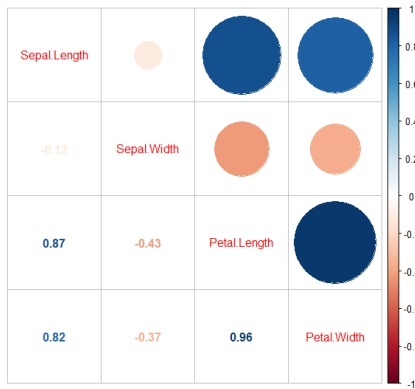


Figura 14.9 - Corplot delle features

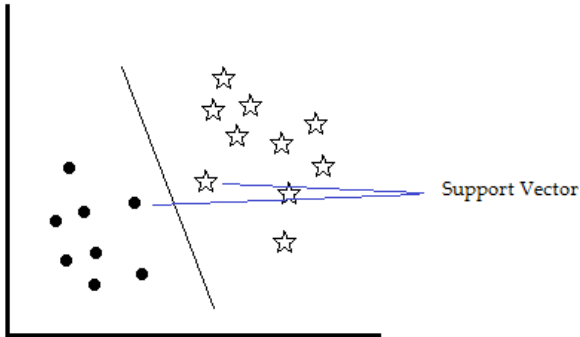


Figura 14.10 - Support Vectors

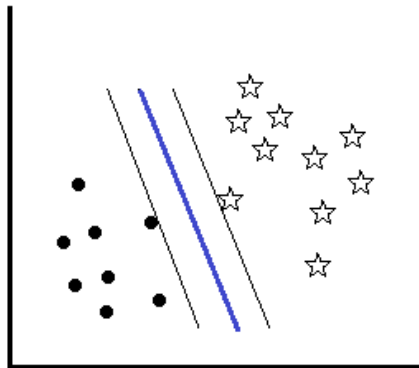


Figura 14.11 - Massimizzazione della distanza tra i vettori

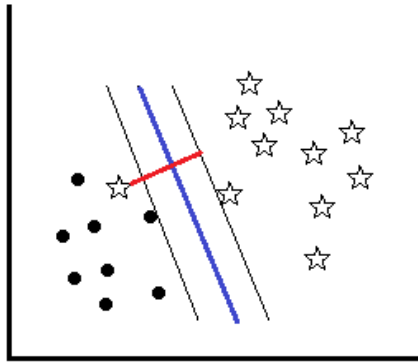


Figura 14.12 - Classi non linearmente separabili

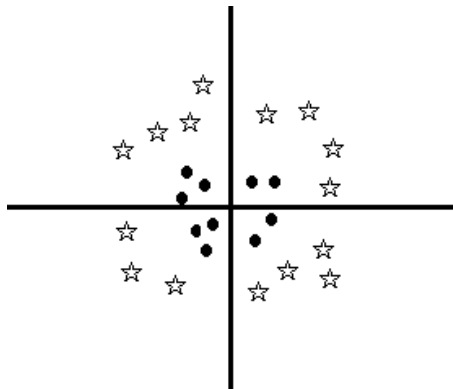


Figura 14.13 - Esempio di classi non separabili linearmente

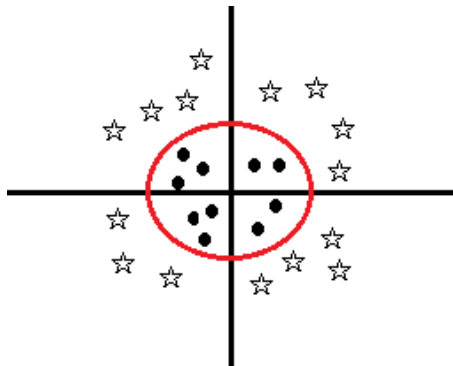


Figura 14.14 - Separazione con un modello di dimensioni superiori

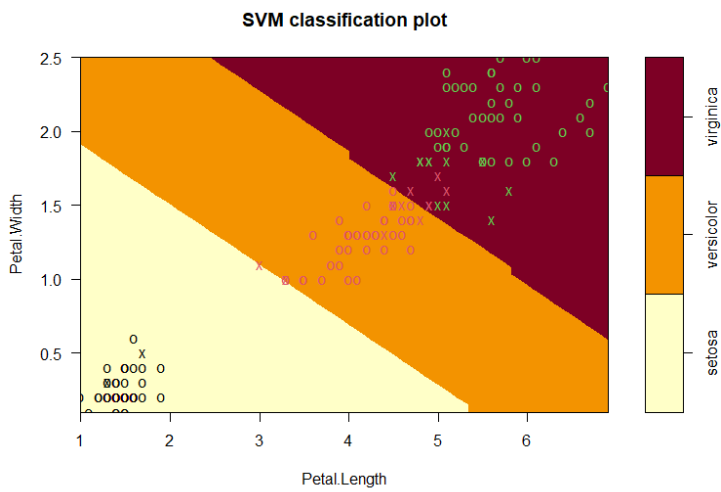


Figura 14.15 - Grafico del primo modello SVM

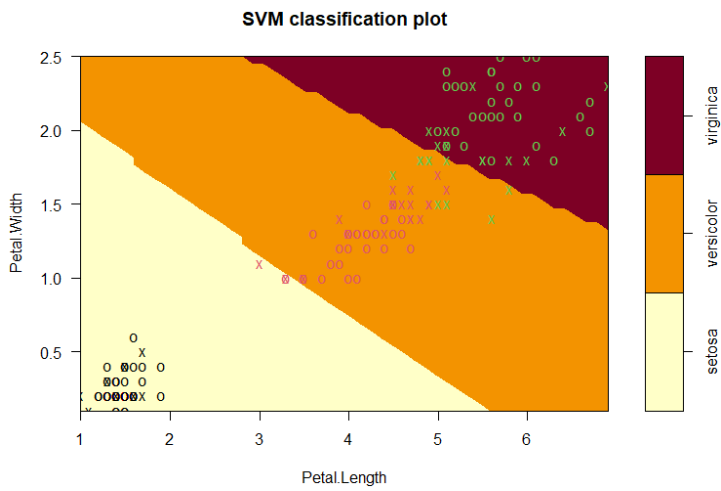


Figura 14.16 - Grafico del secondo modello SVM

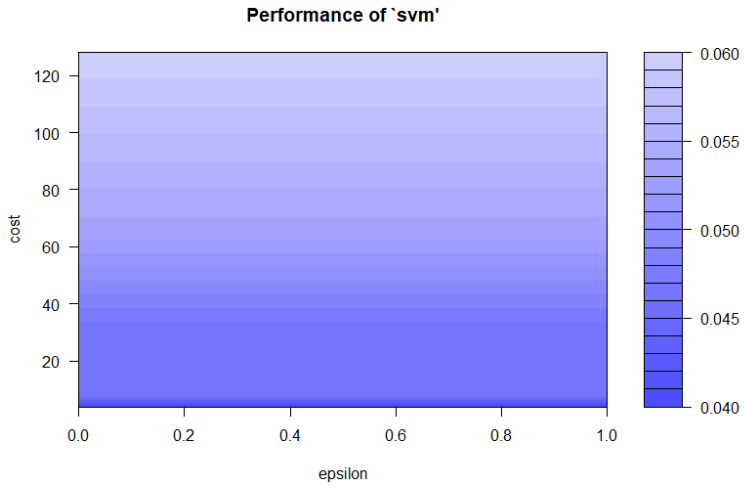


Figura 14.17 - Performance del SVM

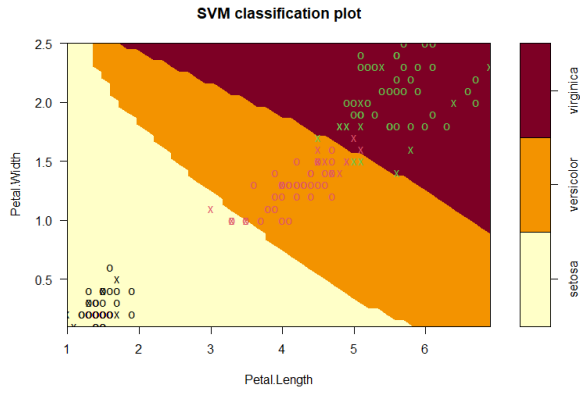


Figura 14.18 - Modello SVM migliore

15. ALGORITMI DI CLUSTERING



Figura 15.1 - Esempio di clustering

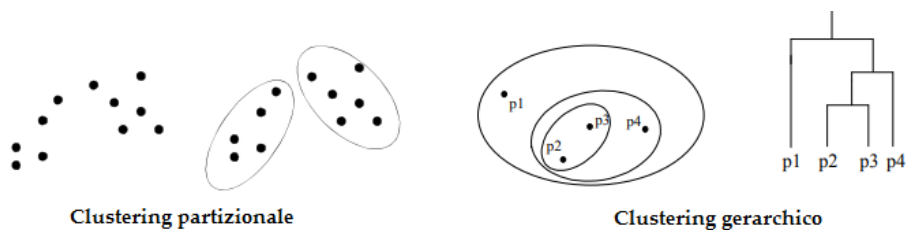


Figura 15.2 - Algoritmi di clustering

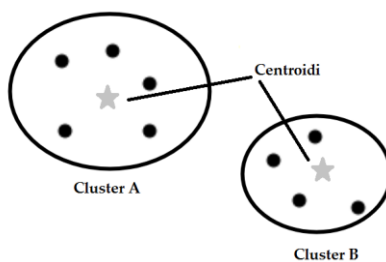


Figura 15.3 - Cluster e loro centroidi

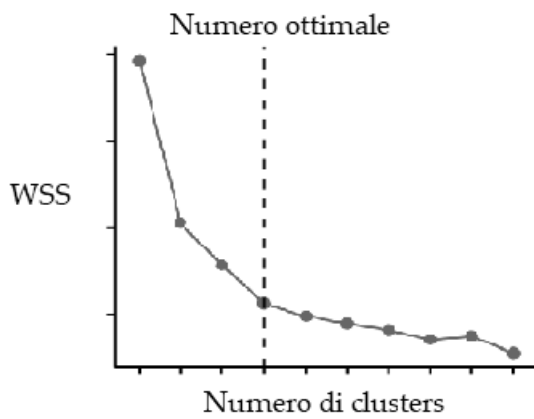


Figura 15.4 - Il metodo Elbow

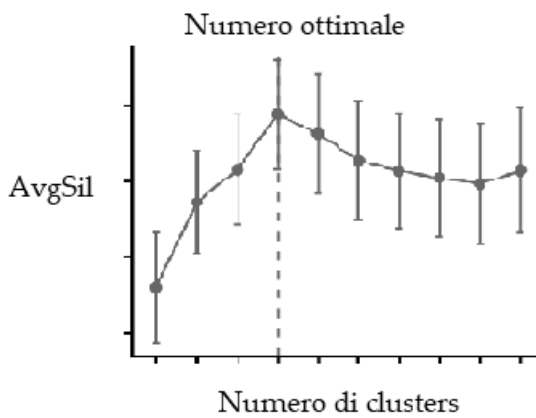


Figura 15.5 - Il metodo della silhouette media

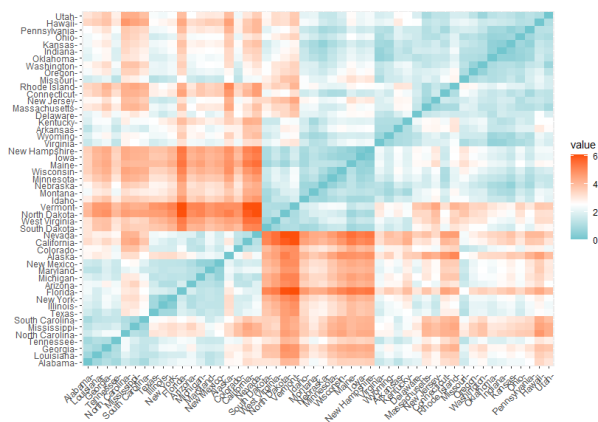


Figura 15.6 - La matrice delle distanze

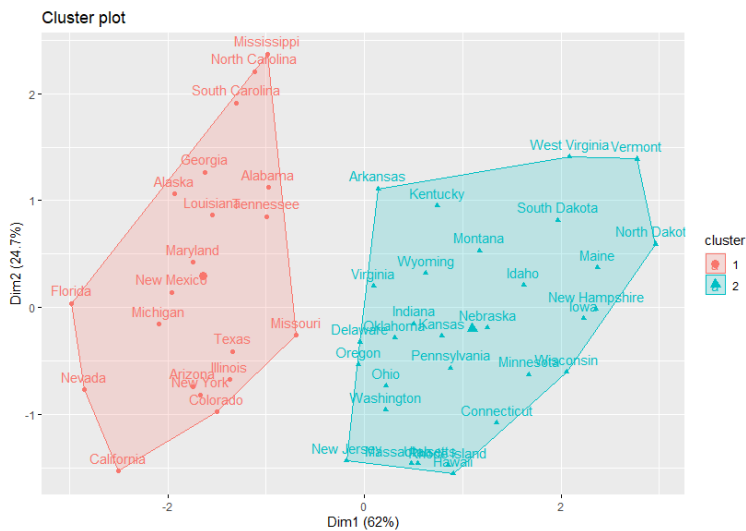


Figura 15.7 - Grafico con $k = 2$

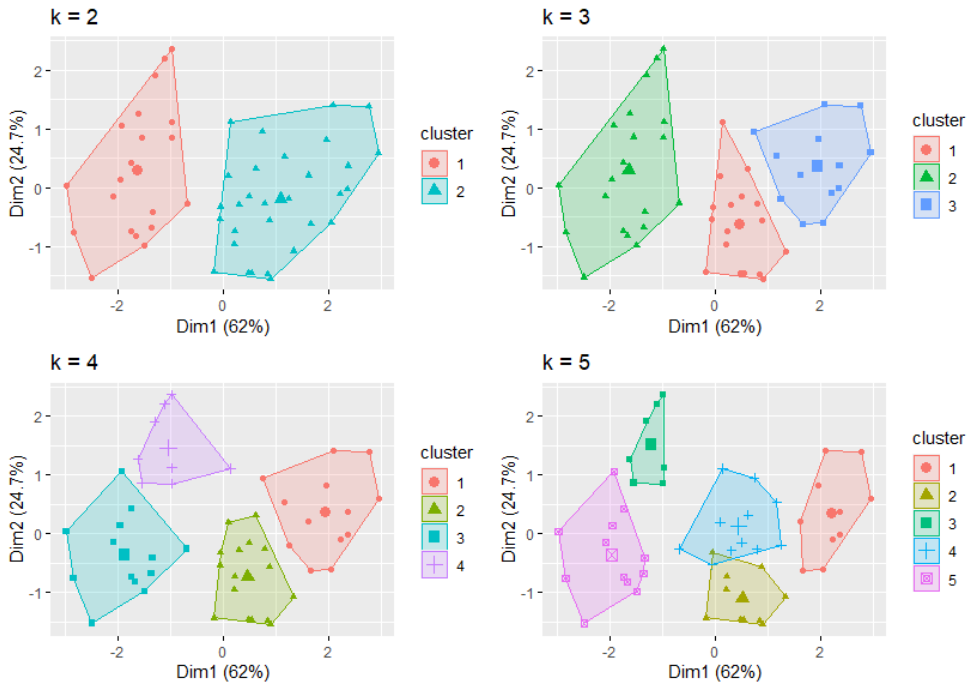


Figura 15.8 - K-means per diversi valori di k

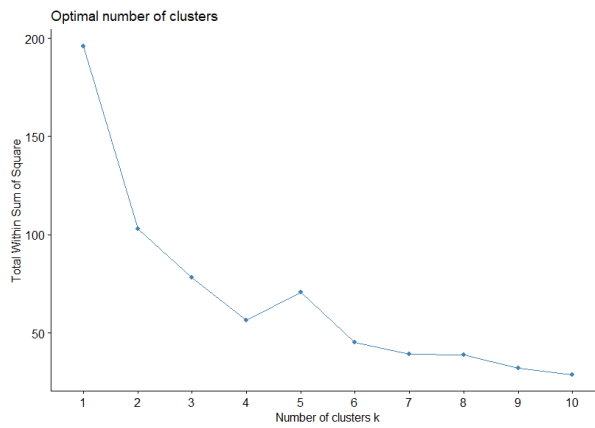


Figura 15.9 - Numero ottimale di cluster col metodo Elbow

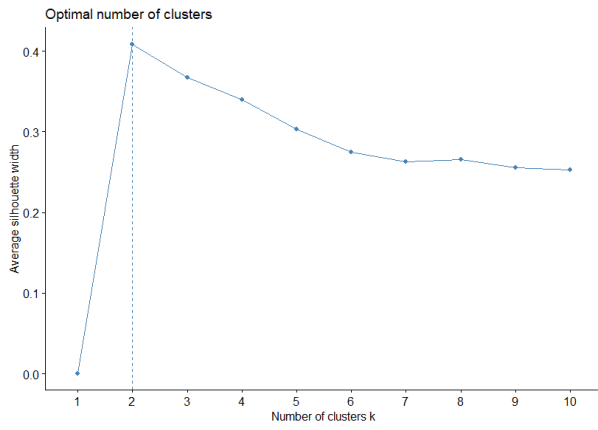


Figura 15.10 - Numero ottimale di cluster col metodo silhouette

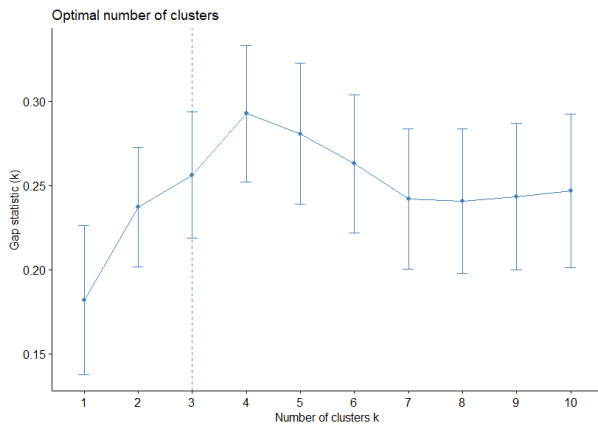


Figura 15.11 - Numero ottimale di cluster col metodo gap statistics

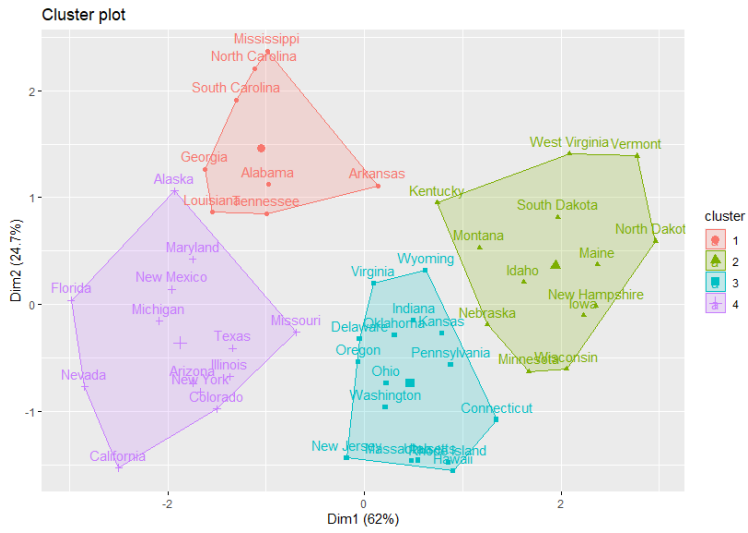


Figura 15.12 - Clustering con $k = 4$

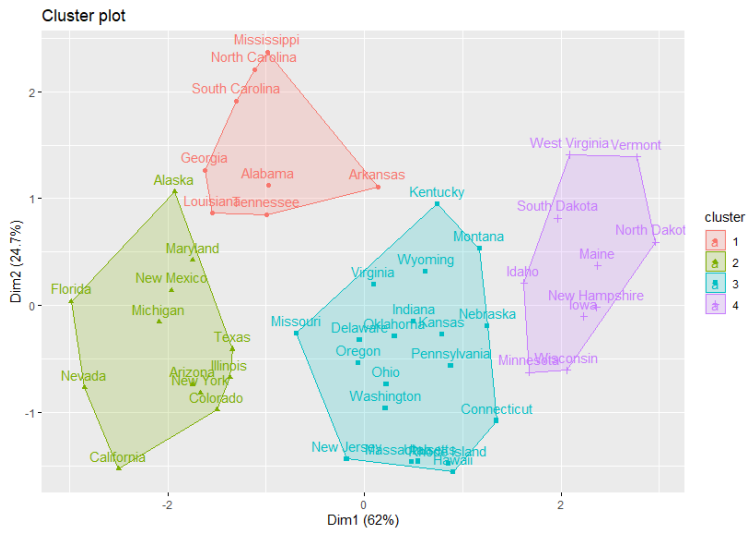


Figura 15.13 - Clustering con PAM

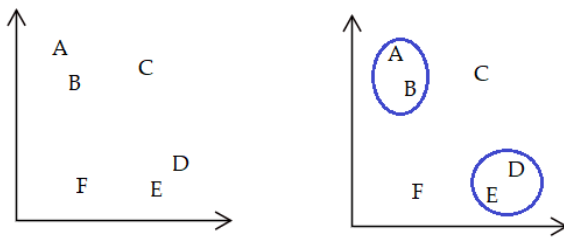


Figura 15.14 - Configurazione iniziale e prima selezione



Figura 15.15 - Prima iterazione

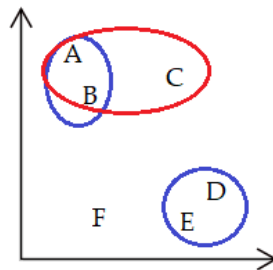


Figura 15.16 - Il cluster C è il più vicino al cluster (A,B)

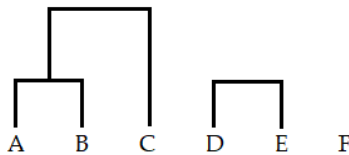


Figura 15.17 - Interazione successiva

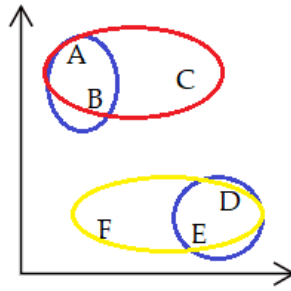


Figura 15.18 – Il cluster F è il più vicino al cluster (D,E)

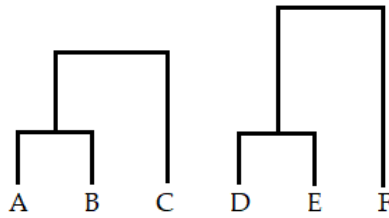


Figura 15.19 - Iterazione successiva

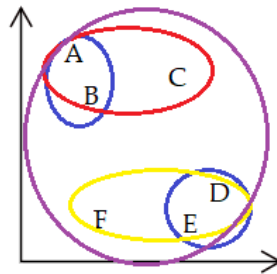


Figura 15.20 - Viene individuato il cluster finale

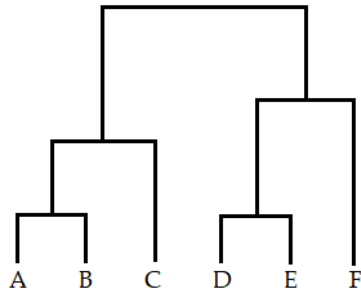


Figura 15.21 - Clustering gerarchico dei dati esaminati

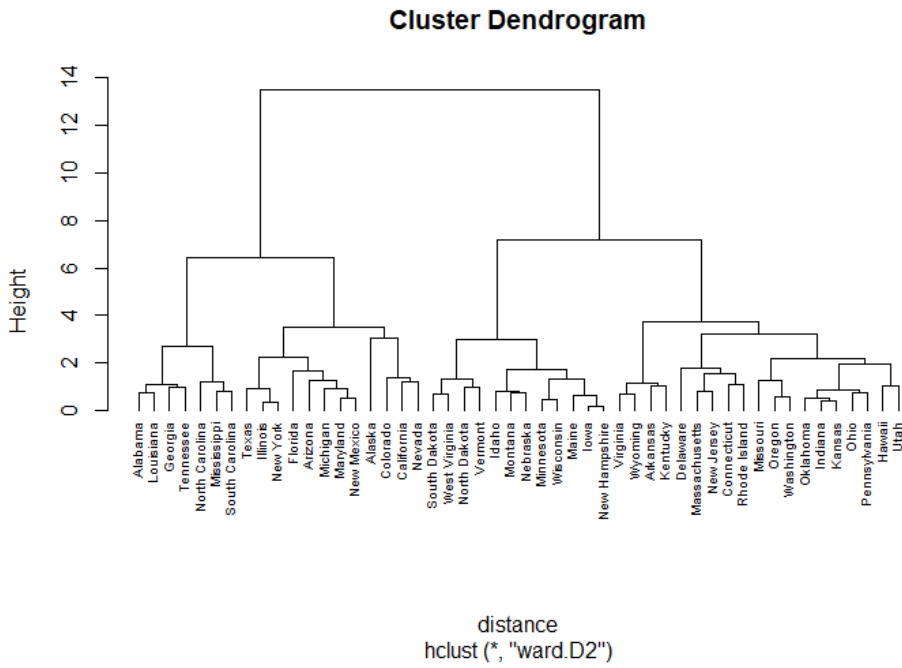


Figura 15.22 - Dendrogramma del clustering gerarchico

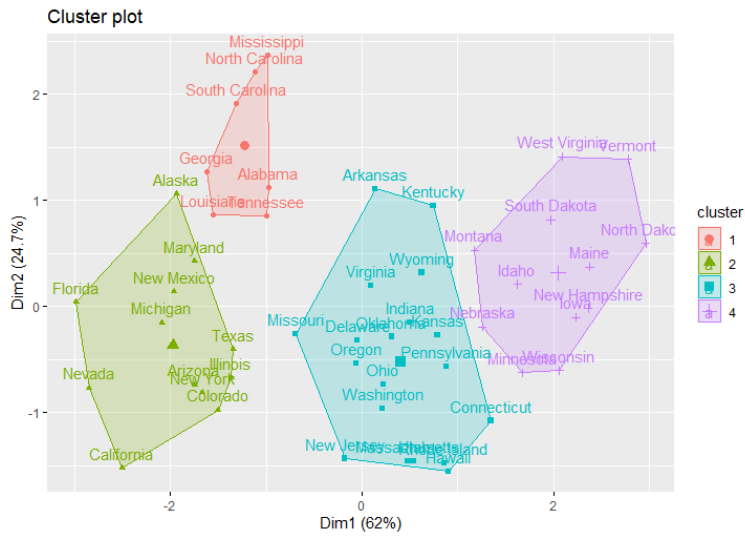


Figura 15.23 - I cluster individuati

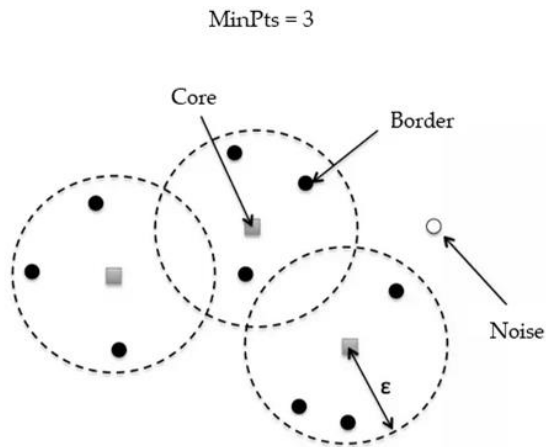


Figura 15.24 - Classificazione dei punti nell' algoritmo DBSCAN

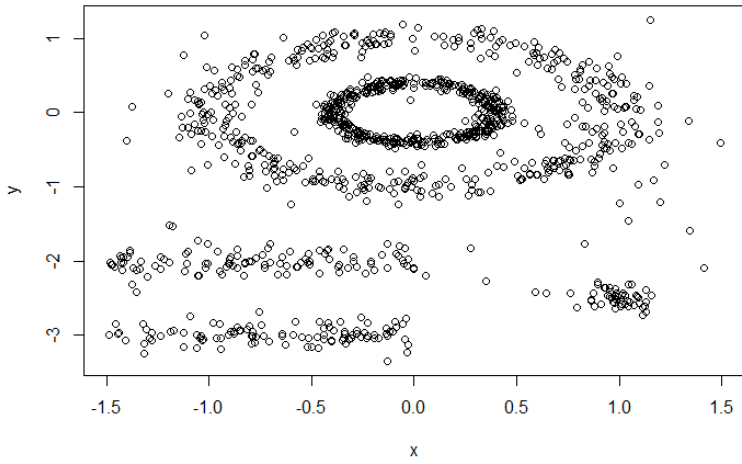


Figura 15.25 - I dati del dataset multishapes

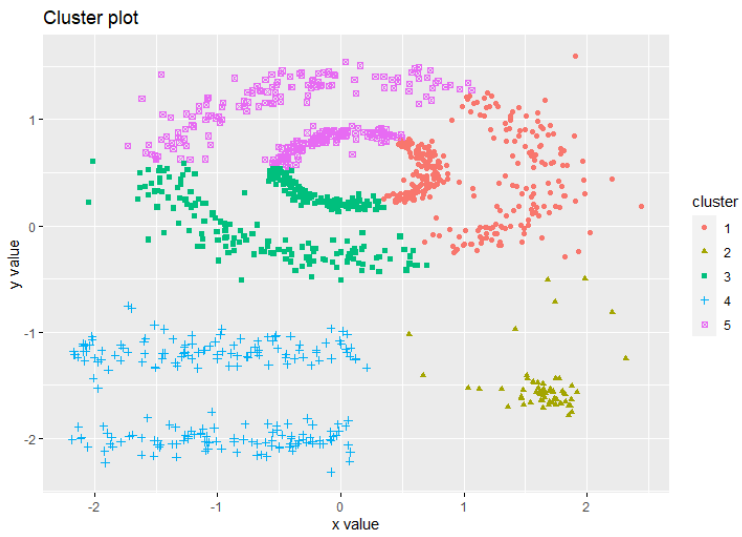


Figura 15.26 - Risultato dell'algoritmo k-means

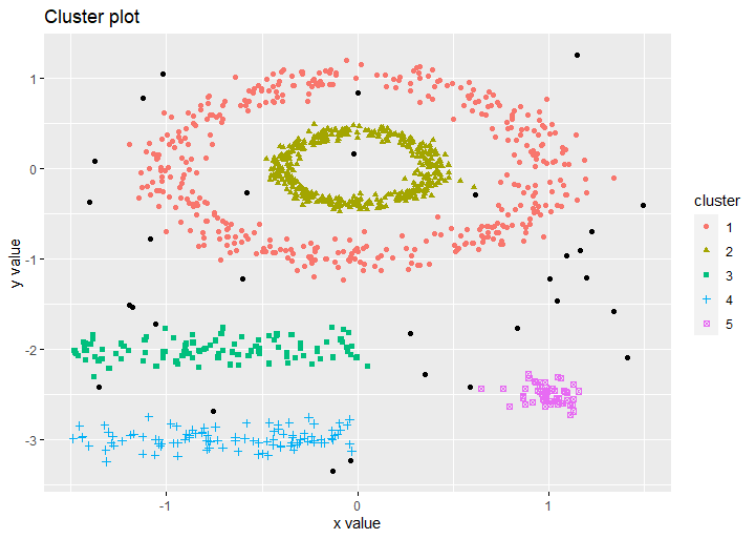


Figura 15.27 - Clustering con DBScan

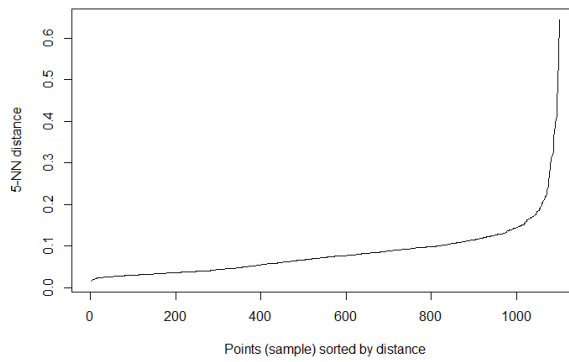


Figura 15.28 - Grafico delle distanze medie per epsilon

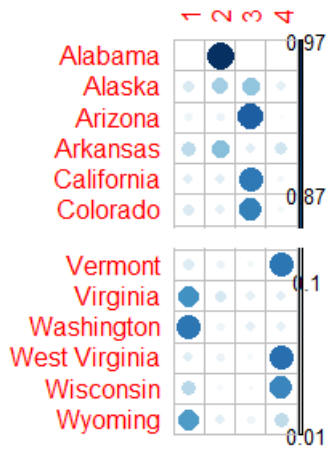


Figura 15.29 - Rappresentazione grafica dei gradi di appartenenza

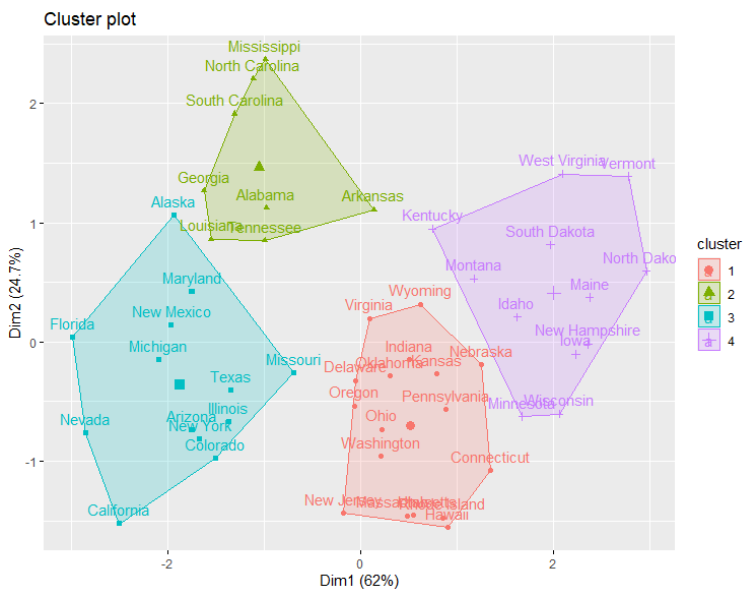


Figura 15.30 - Il risultato del clustering

16. RIDUZIONE DELLA DIMENSIONALITÀ

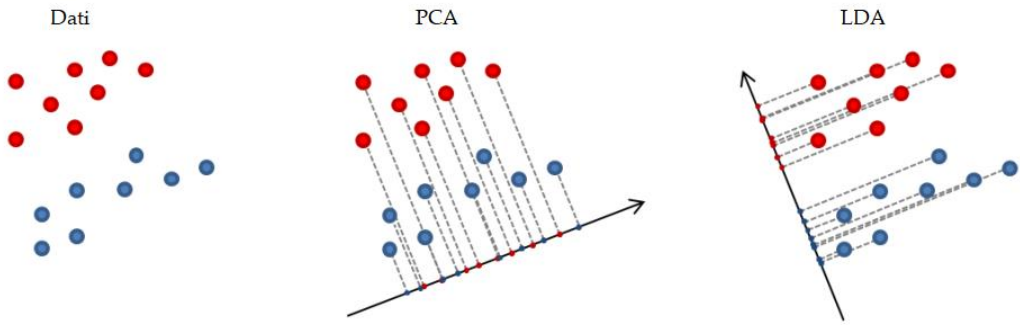


Figura 16.1 - PCA vs LDA

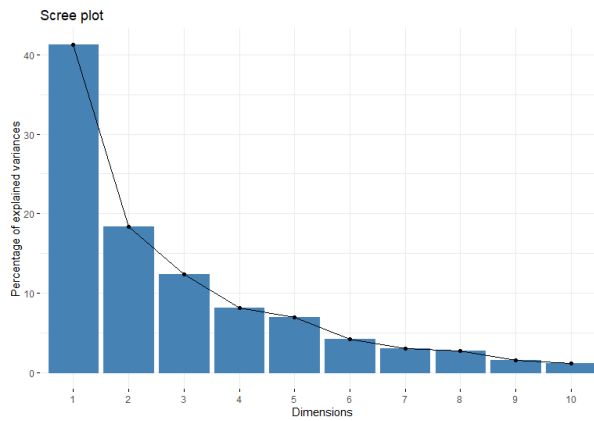


Figura 16.2 - Grafico delle percentuali di varianza spiegate per componente

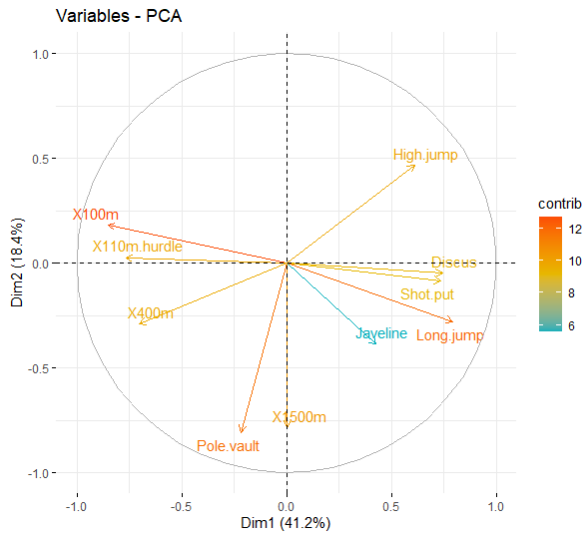


Figura 16.3 - Grafico delle correlazioni

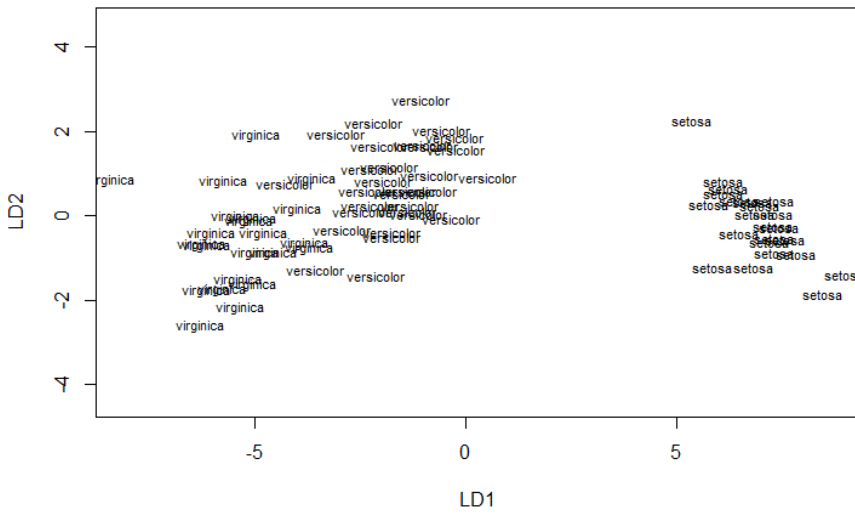


Figura 16.4 - Classificazione con LDA

17. ALGORITMI SEMI SUPERVISIONATI

Feature 1	Feature 2	Feature 3	Feature 4	Classe
6	27	47	9	SI
10	43	26	13	SI
24	17	42	43	SI
35	41	21	1	SI
45	27	19	25	-
46	5	22	36	-
46	3	23	38	-
42	33	4	41	-
45	26	49	6	-
17	14	35	47	-
46	39	15	43	-
22	33	6	28	-
50	7	43	14	-
32	8	2	37	-
18	20	16	40	-
7	40	1	47	-
43	1	47	16	-
1	10	50	9	-
14	13	2	37	-

Supervisionato

Non supervisionato

Figura 17.1 - Esempio di problema semi supervisionato

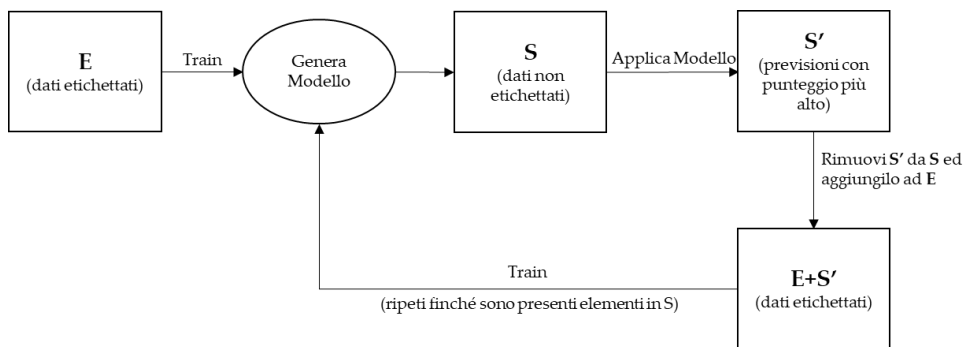


Figura 17.2 - Il metodo del self training

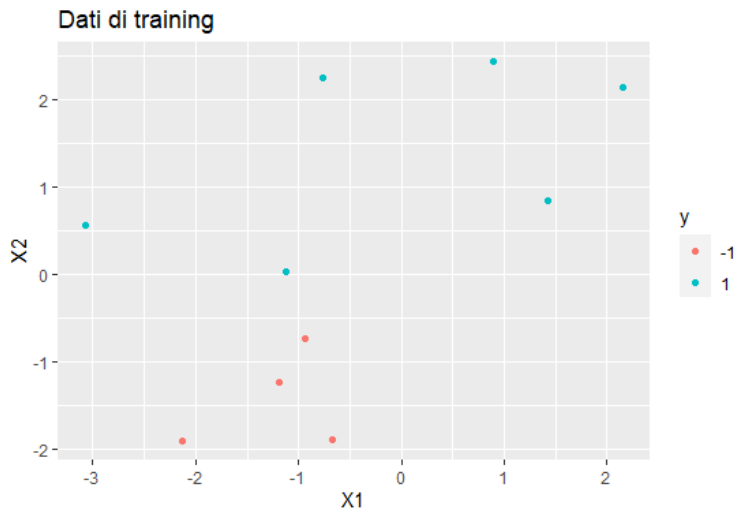


Figura 17.3 - I dati di training etichettati

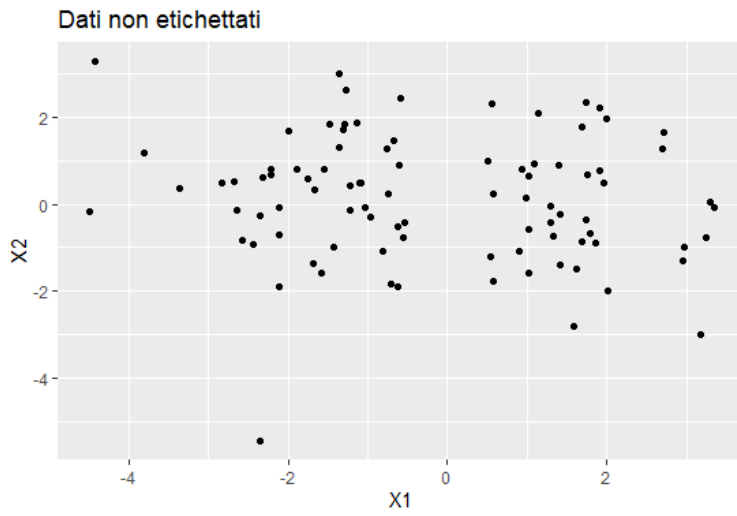


Figura 17.4 - I dati non etichettati



Figura 17.5 - I dati di test

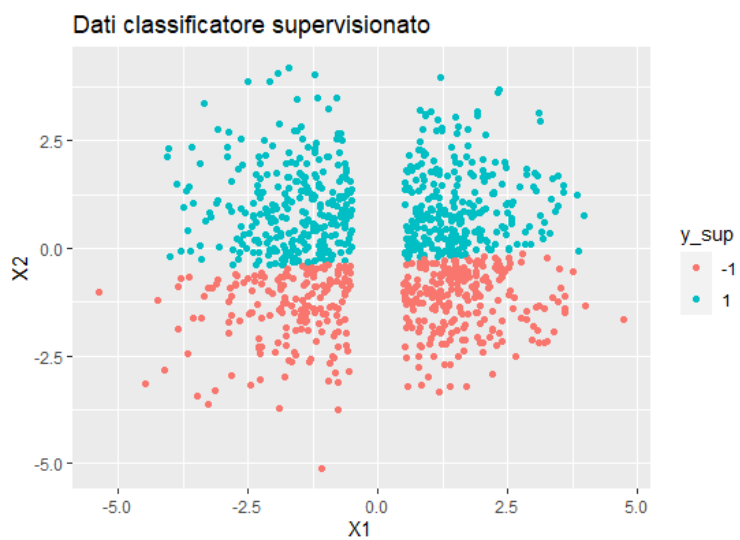


Figura 17.6 - Risultati del classificatore supervisionato

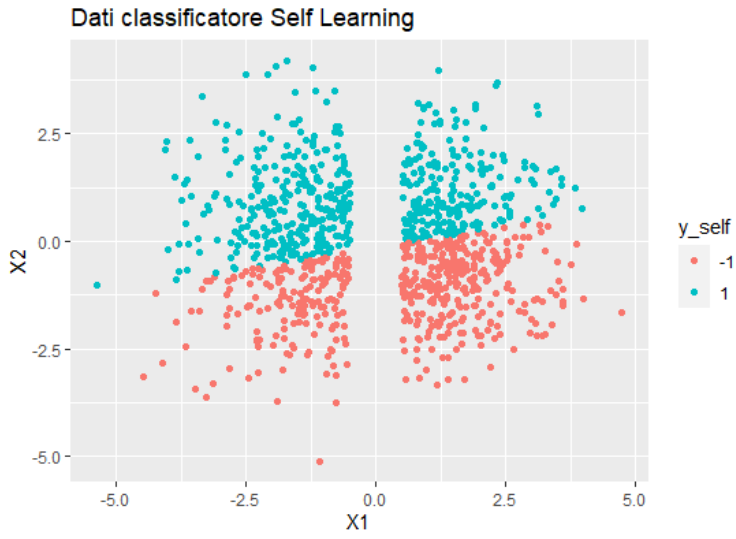


Figura 17.7 - Il classificatore self training

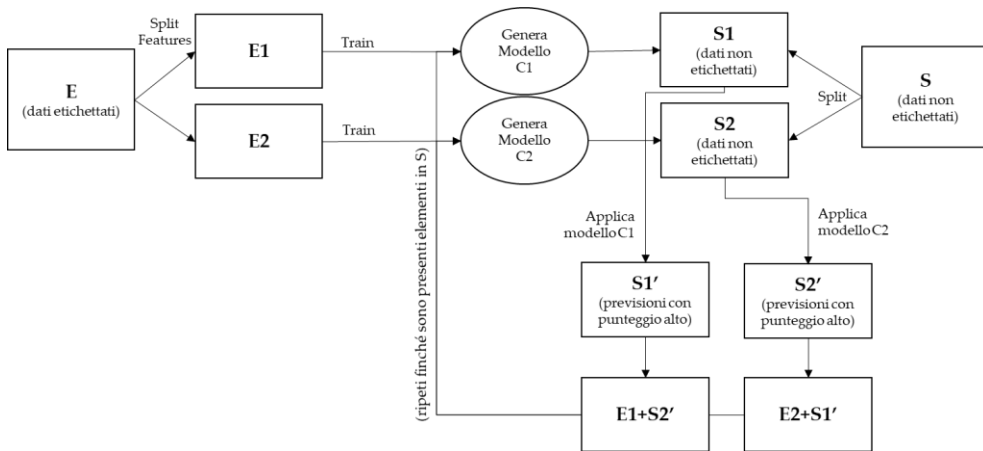


Figura 17.8 - Il metodo del Co-training

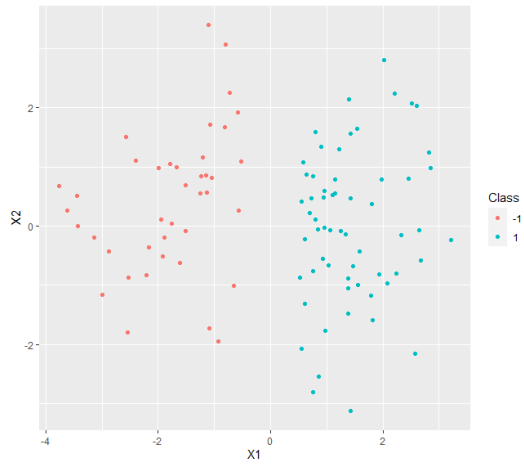


Figura 17.9 - Il dataset generato

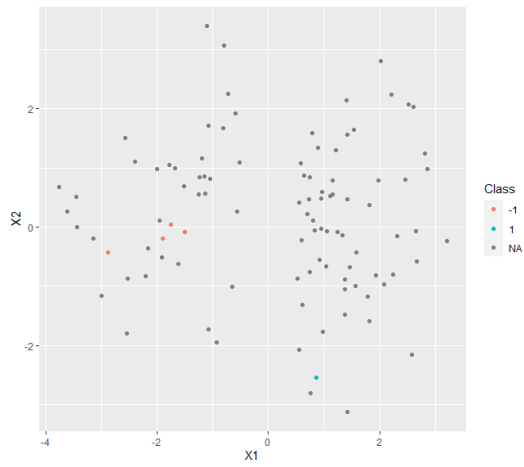


Figura 17.10 - Il dataset con etichette missing

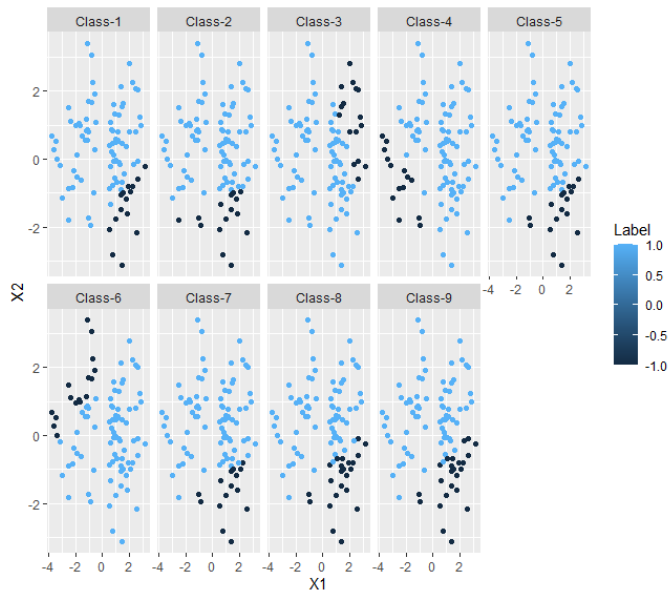


Figura 17.11 – Etichette considerate

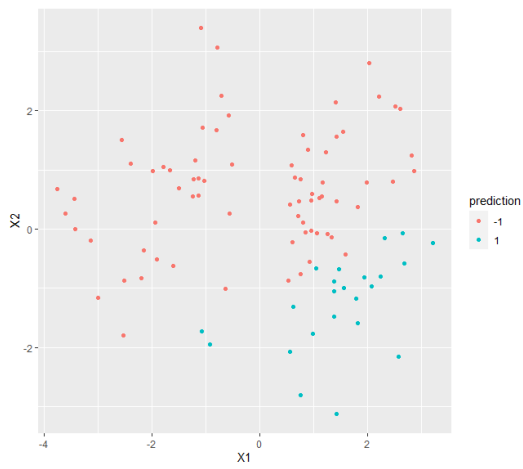


Figura 17.12 - Etichettatura Finale

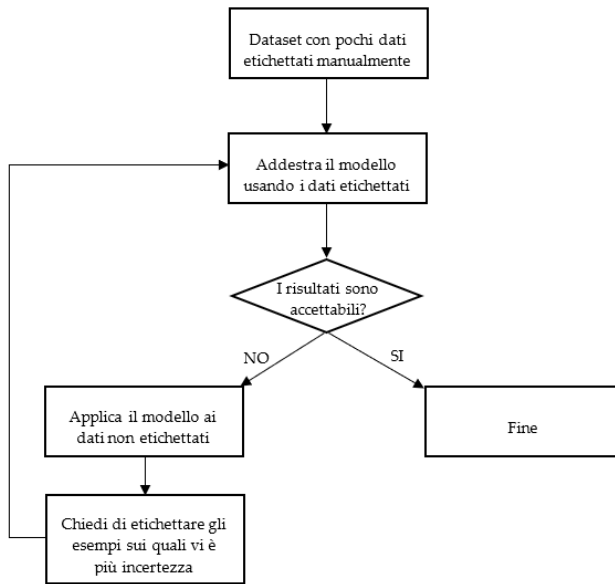


Figura 17.13 – Il processo iterativo nell'Active Learning

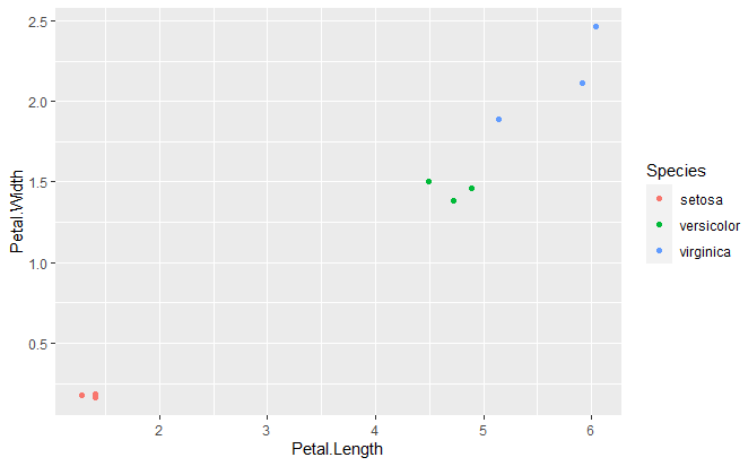


Figura 17.14 - I dati etichettati

18. LAZY LEARNING

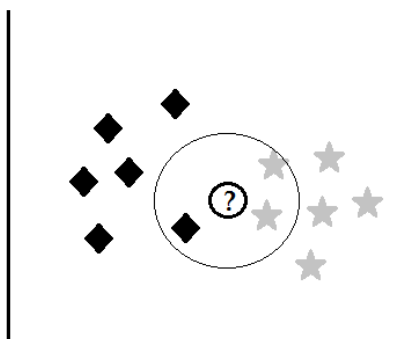


Figura 18.1 - Esempio di algoritmo KNN

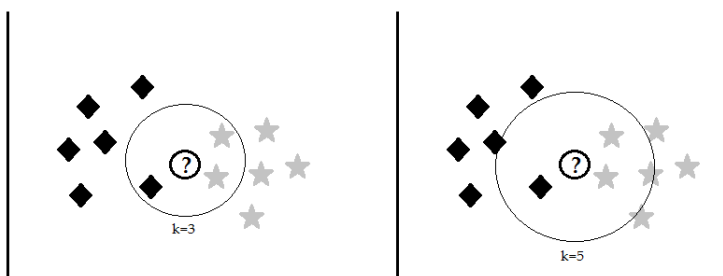


Figura 18.2 - La scelta di k nell'algoritmo KNN

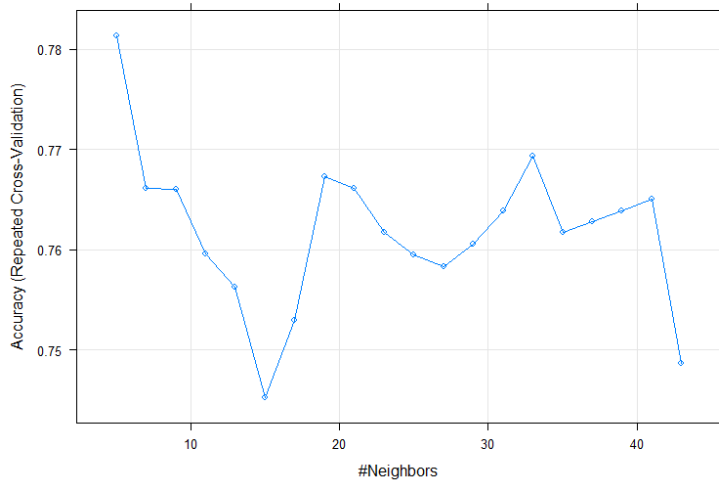


Figura 18.3 - Performance del modello per i diversi valori di K

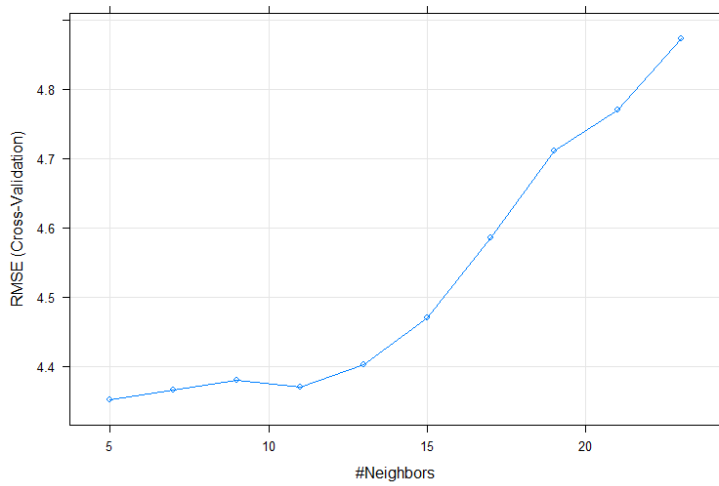


Figura 18.4 - Grafico dell'RMS per diversi valori di k

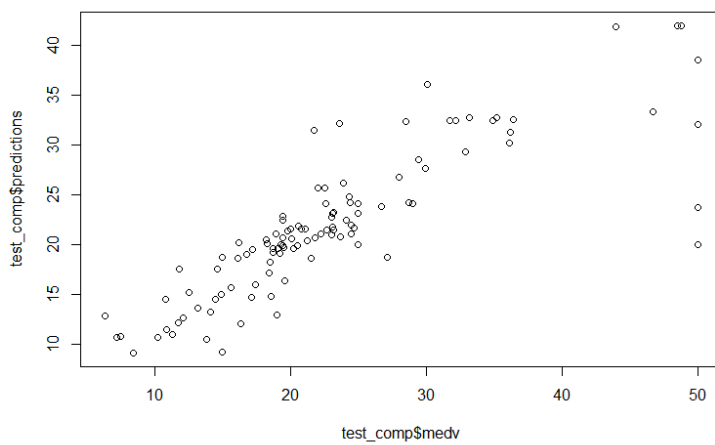


Figura 18.5 - Valore previsto vs valore reale

19. METODI ENSEMBLE

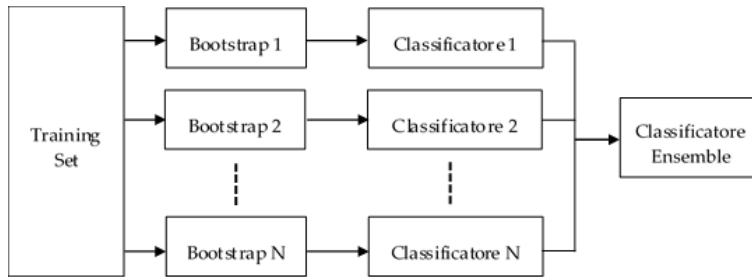


Figura 19.1 - Bagging

PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Survived
1	3	Kelly, Mr. James	male	34.50	0	0	330911	7.8292		Q	0
2	3	Wilkes, Mrs. James (Ellen Needs)	female	47.00	1	0	363272	7.0000		S	0
3	2	Myles, Mr. Thomas Francis	male	62.00	0	0	240276	9.6875		Q	0
4	3	Wirz, Mr. Albert	male	27.00	0	0	315154	8.6625		S	0
5	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.00	1	1	3101298	12.2875		S	0
6	3	Svensson, Mr. Johan Cervin	male	14.00	0	0	7538	9.2250		S	0
7	3	Connolly, Miss. Kate	female	30.00	0	0	330972	7.6292		Q	1
8	2	Caldwell, Mr. Albert Francis	male	26.00	1	1	248738	29.0000		S	0
9	3	Abraham, Mrs. Joseph (Sophie Halaut Easu)	female	18.00	0	0	2657	7.2292		C	1
10	3	Davies, Mr. John Samuel	male	21.00	2	0	A/4 46871	24.1500		S	0
12	1	Jones, Mr. Charles Cresson	male	46.00	0	0	694	26.0000		S	0
13	1	Snyder, Mrs. John Pillsbury (Nelle Stevenson)	female	23.00	1	0	21228	82.2667	B45	S	1
14	2	Howard, Mr. Benjamin	male	63.00	1	0	24065	26.0000		S	0

Figura 19.2 - Il dataset di test

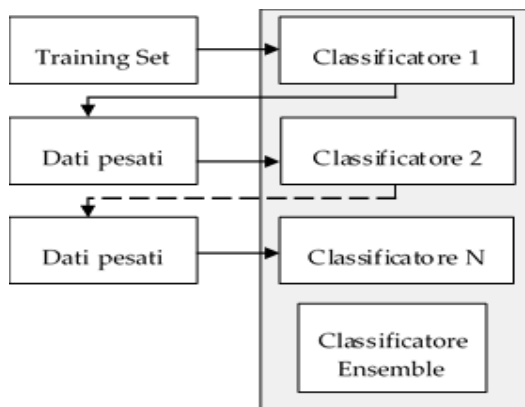


Figura 19.3 - Boosting

20. ALGORITMI DI ASSOCIAZIONE

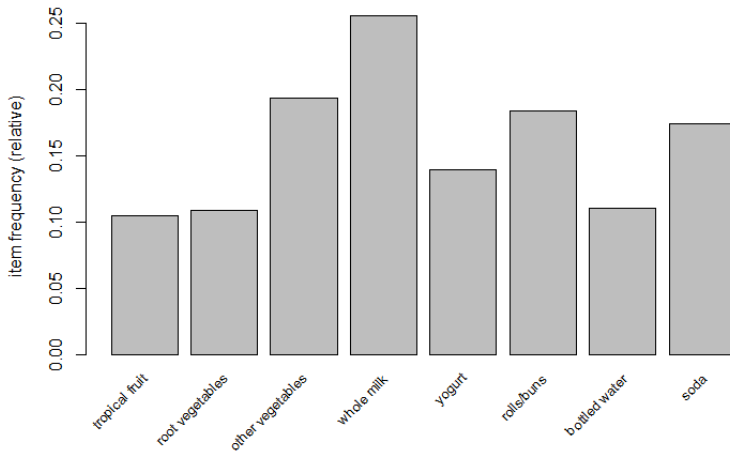


Figura 20.1 - Itemset con supporto almeno pari a 0,1

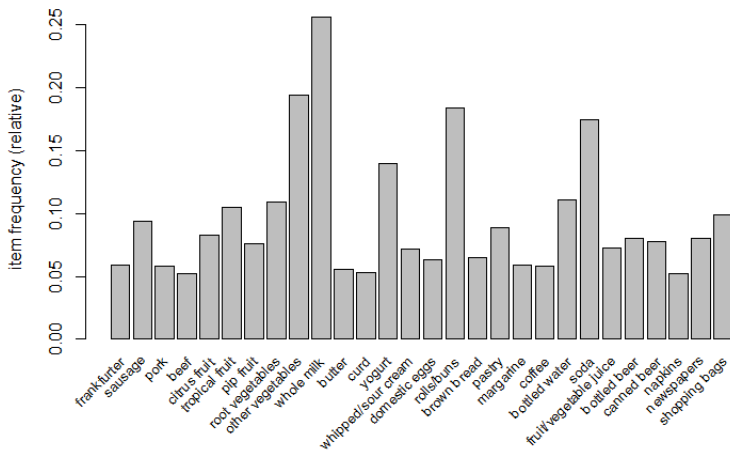


Figura 20.2 - Itemset con supporto almeno pari a 0,05

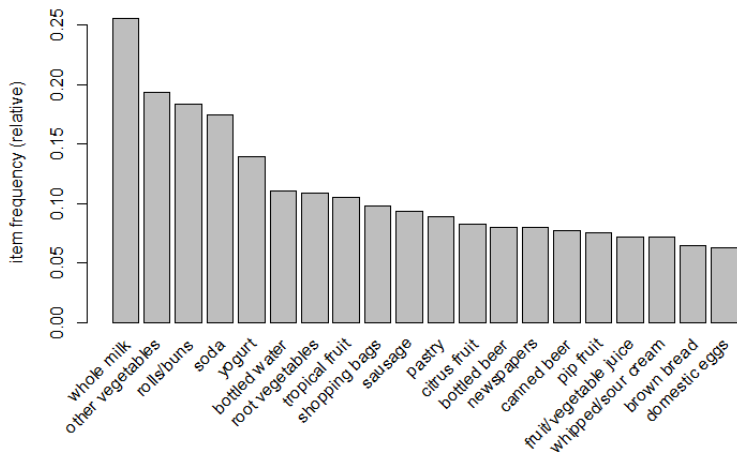


Figura 20.3 - I 20 itemset più frequenti

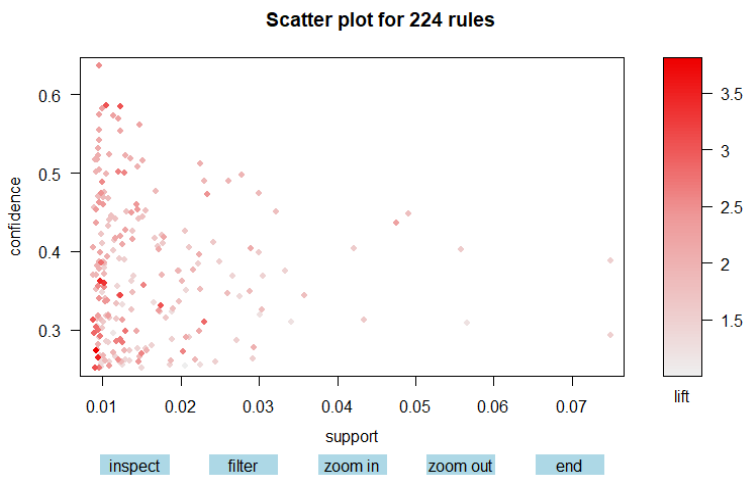


Figura 20.4 - Grafico di supporto e confidenza

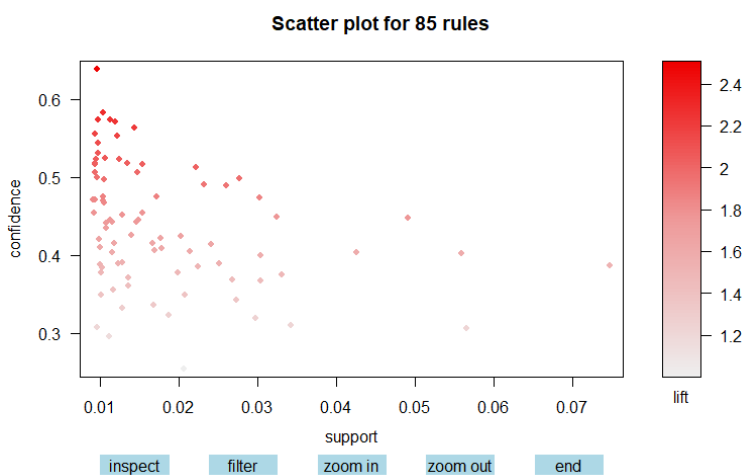


Figura 20.5 - Regole per prodotti collegati al latte

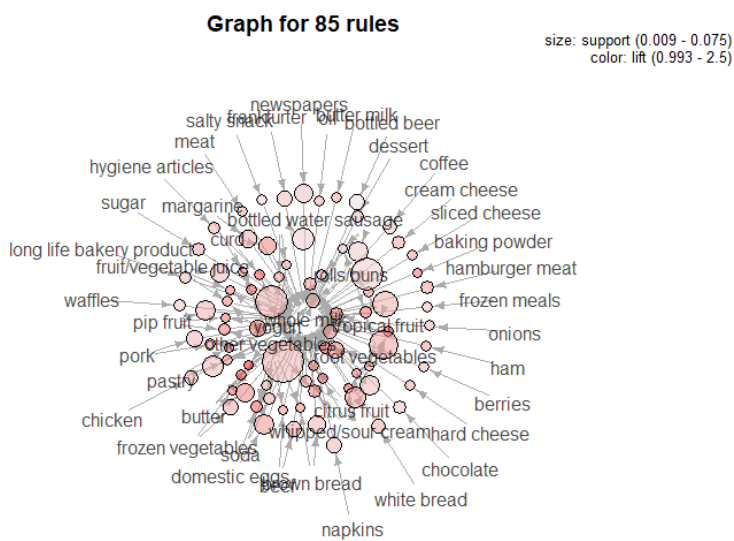


Figura 20.6 - Grafico delle regole

21. SERIE TEMPORALI

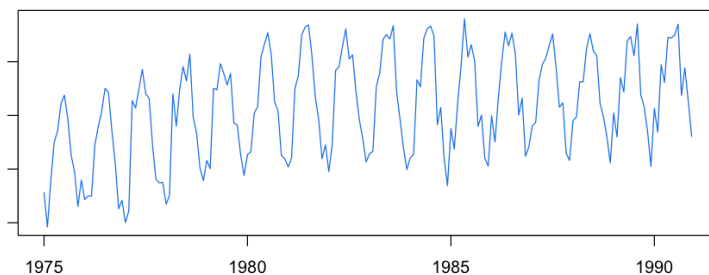


Figura 21.1 - Esempio di serie temporale

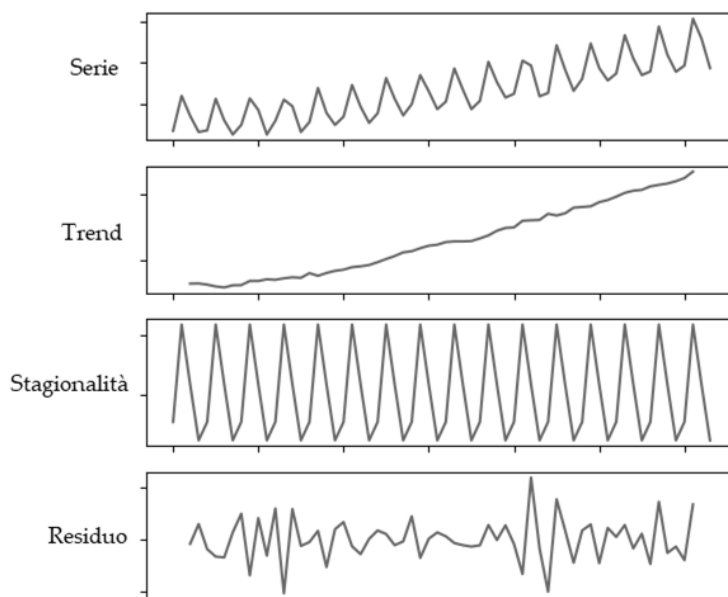


Figura 21.2 - Scomposizione di una serie temporale

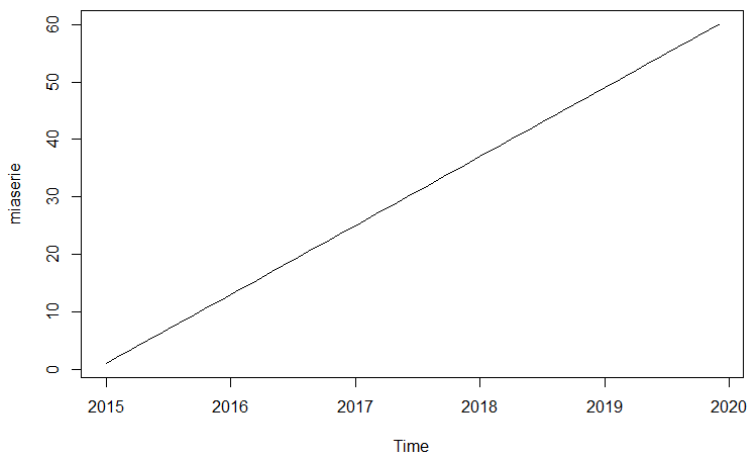


Figura 21.3 - La nostra semplice serie temporale

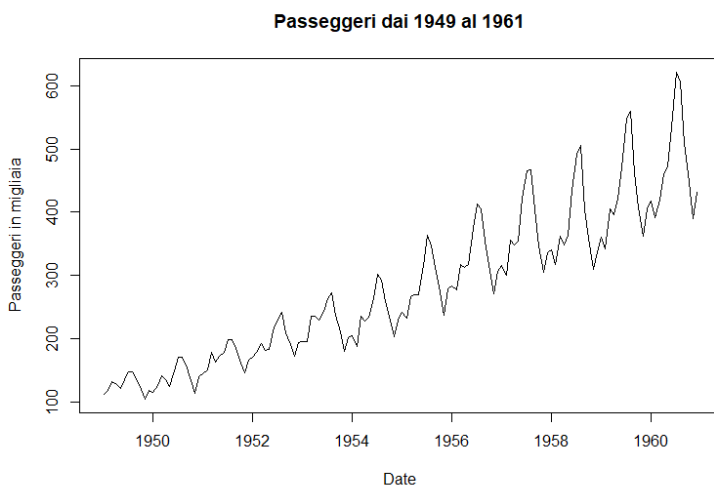


Figura 21.4 - La serie temporale AirPassengers

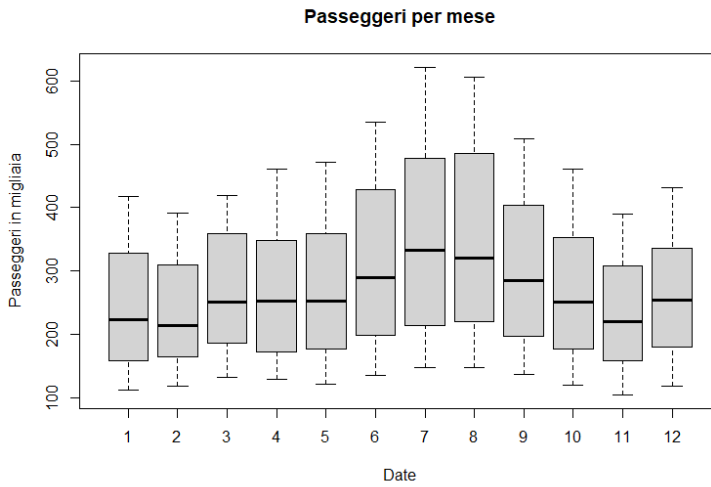


Figura 21.5 - Boxplot per mese

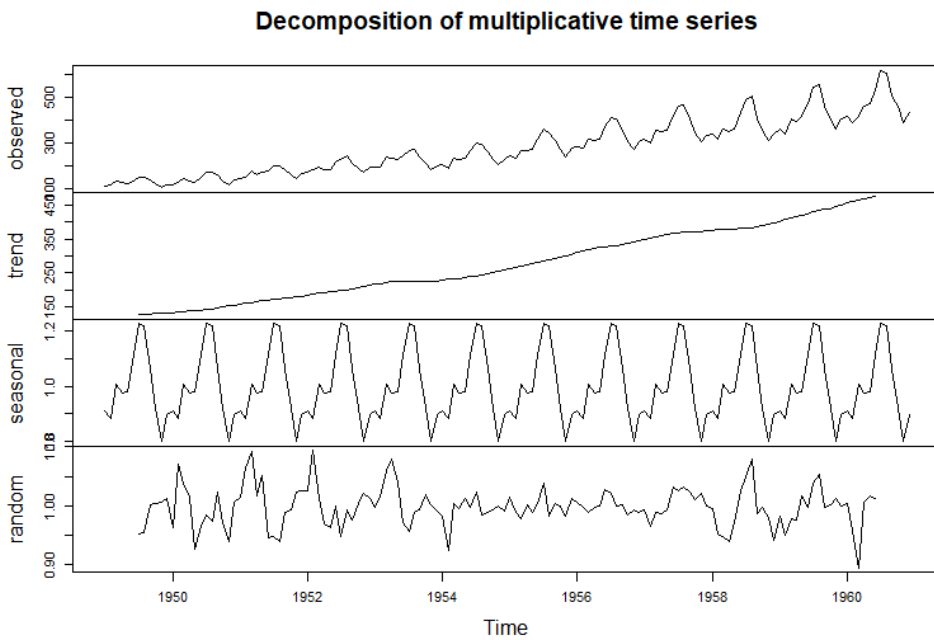


Figura 21.6 - Le componenti della serie temporale



Figura 21.7 - Analisi dei residui

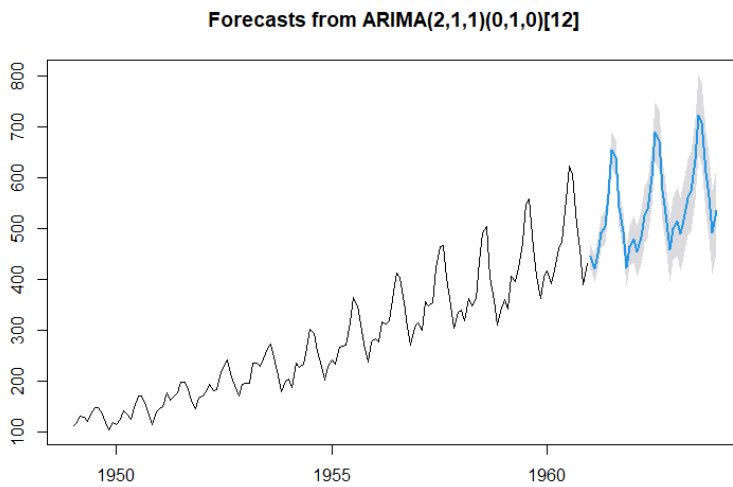


Figura 21.8 - Plot del forecast

22. RETI NEURALI

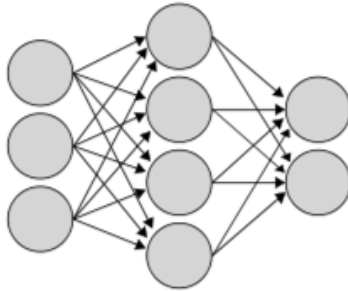


Figura 22.1 - Rappresentazione di una rete neurale

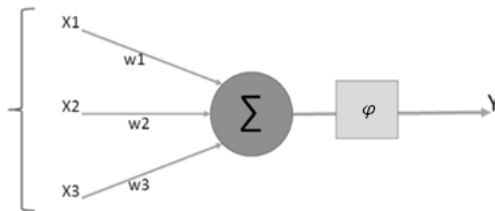


Figura 22.2 - Struttura del neurone

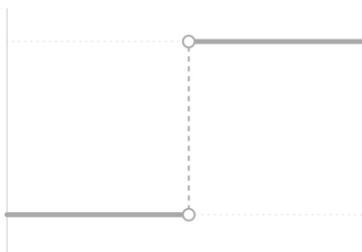


Figura 22.3 - Funzione soglia

$$f(x) = x$$

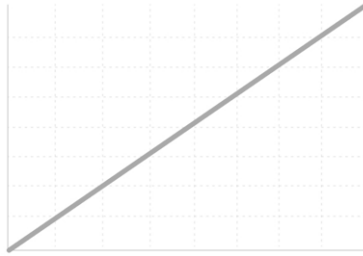


Figura 22.4 - Funzione lineare



Figura 22.5 - Funzione ReLu

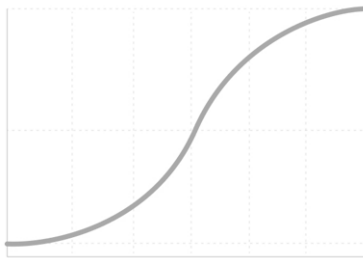


Figura 22.6 - Funzione Logistica

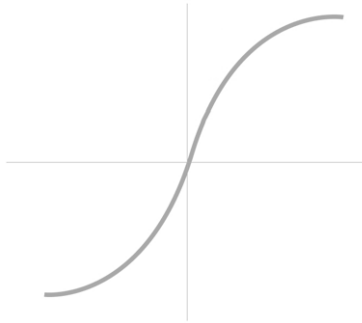


Figura 22.7 - Funzione tangente iperbolica

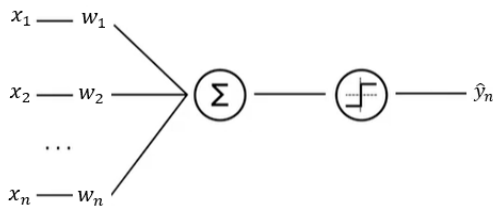


Figura 22.8 - Struttura del percettrone

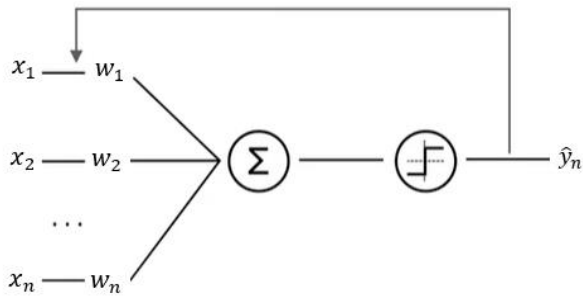


Figura 22.9 - Algoritmo del percettrone

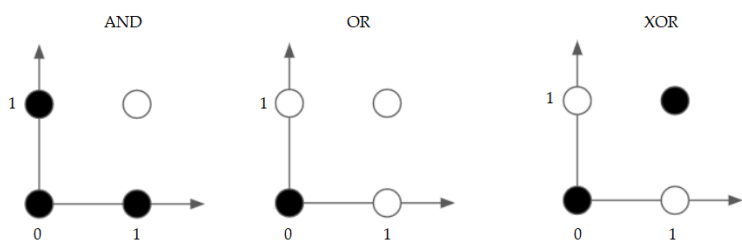


Figura 22.10 - Rappresentazione delle funzioni AND, OR e XOR

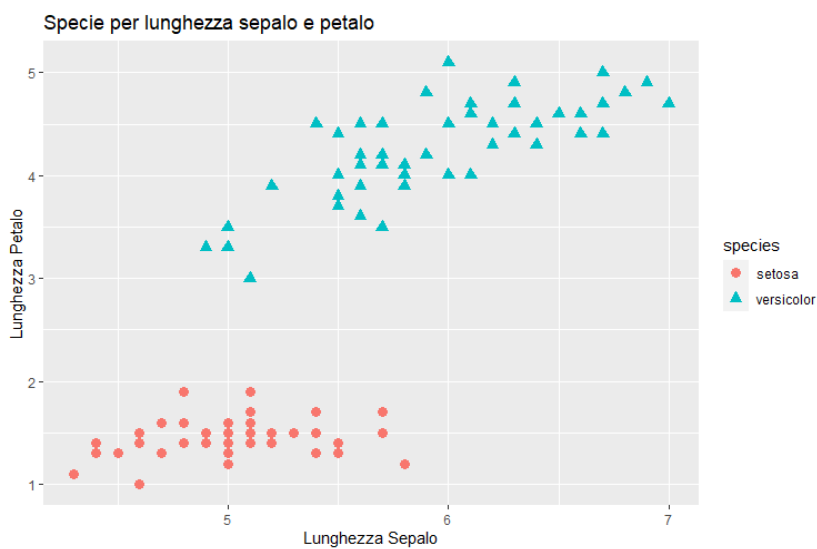


Figura 22.11 - Le classi linearmente separabili

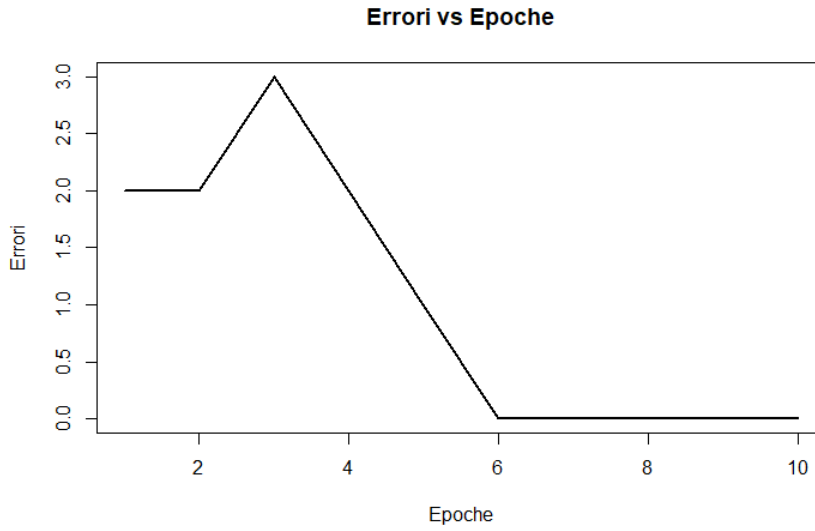


Figura 22.12 - Grafico degli errori

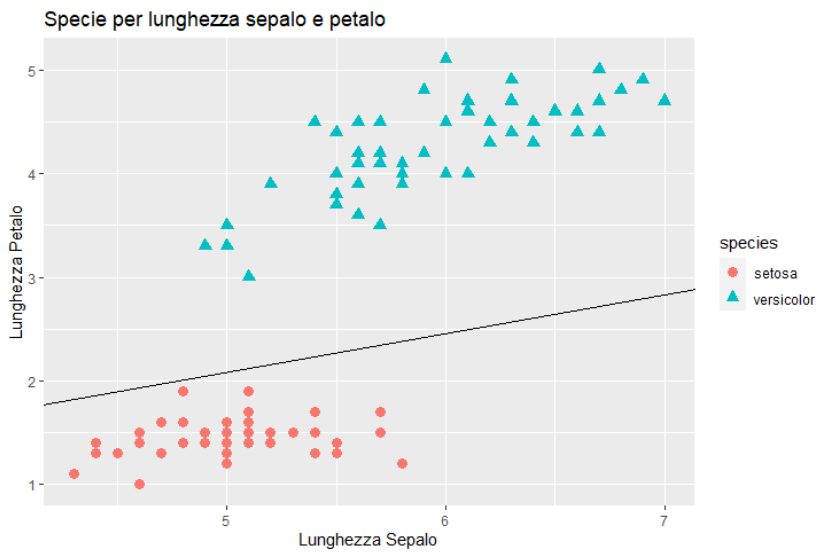


Figura 22.13 - I risultati del perceptrone

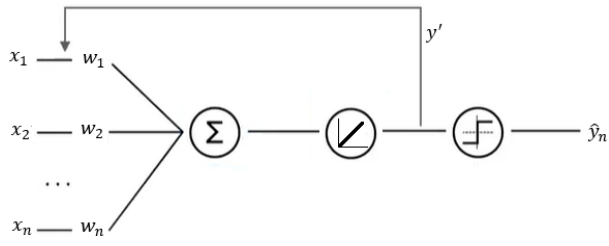


Figura 22.14 - Algoritmo Adaline

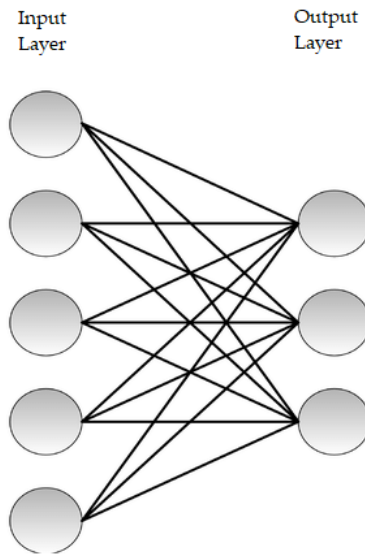


Figura 22.15 - Rete neurale SLP

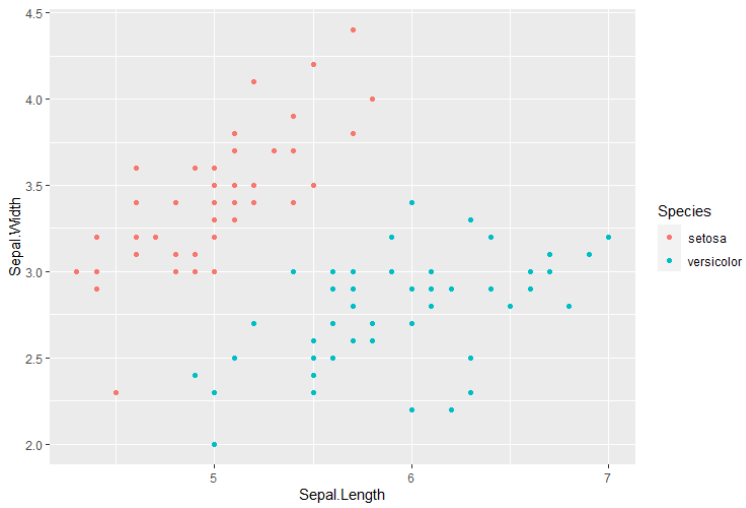
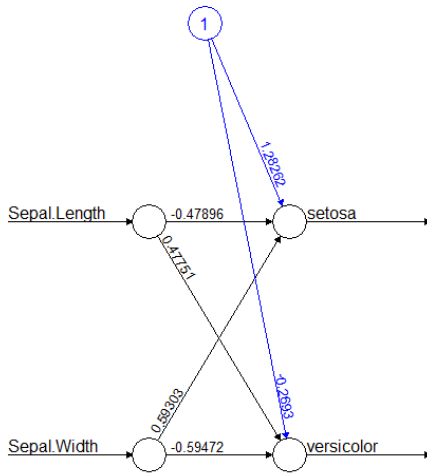


Figura 22.16 - Le due classi linearmente separabili



Error: 4.107052 Steps: 2023

Figura 22.17 - La rete SLP del nostro esempio

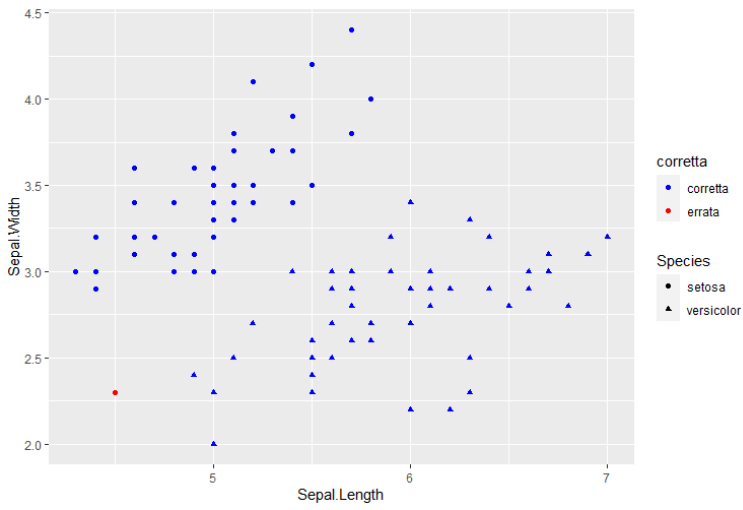


Figura 22.18 - Classi correttamente individuate

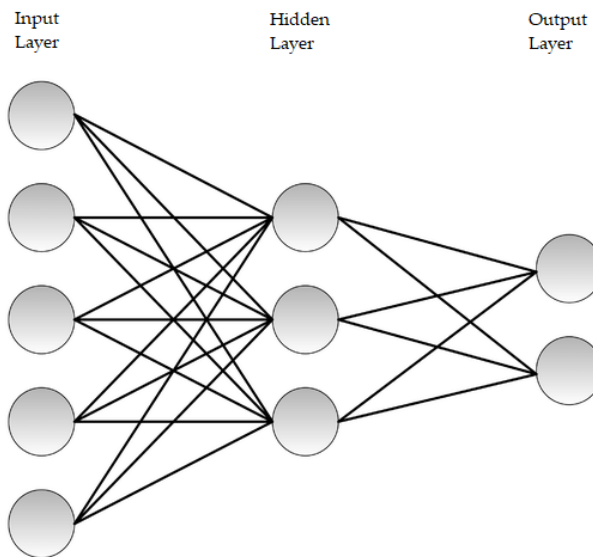


Figura 22.19 - Rete Neurale MLP

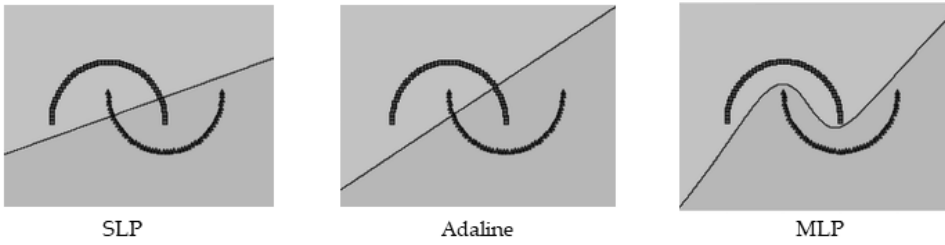


Figura 22.20 - Risoluzione di problemi non lineari con MLP

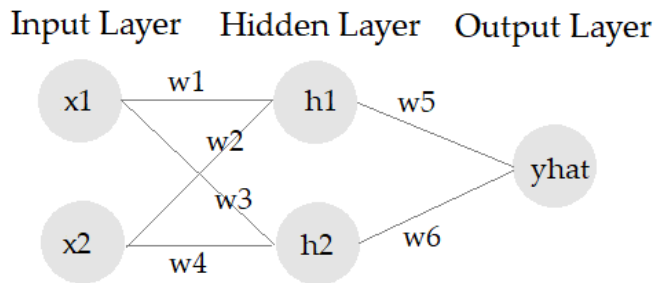


Figura 22.21 - La rete del nostro esempio

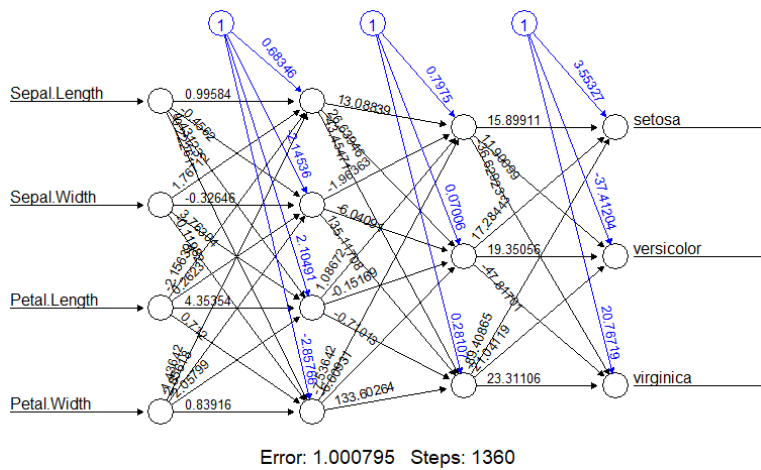


Figura 22.22 - Il modello di rete neurale

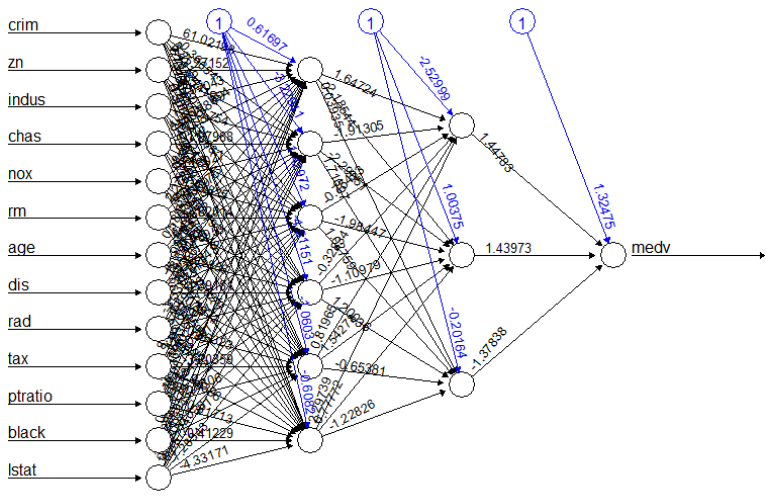


Figura 22.23 - Rete neurale per il dataset boston

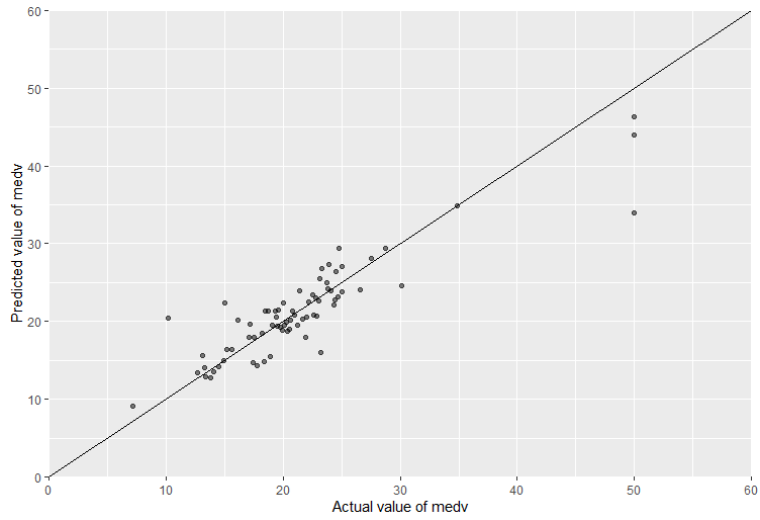


Figura 22.24 - Valore previsto vs Actual

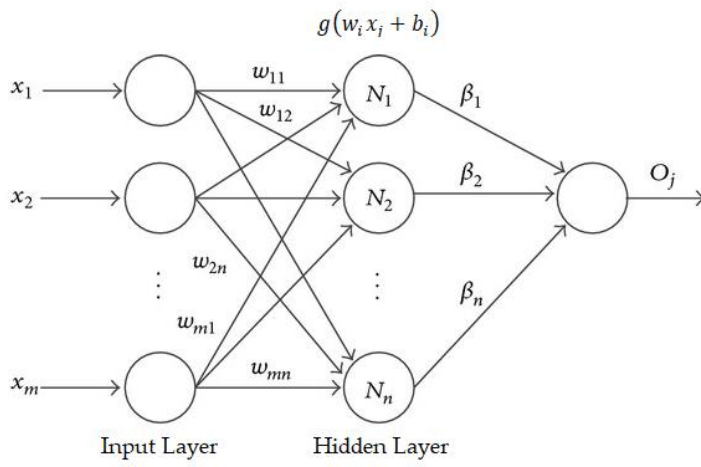


Figura 22.25 - Struttura di una ELM

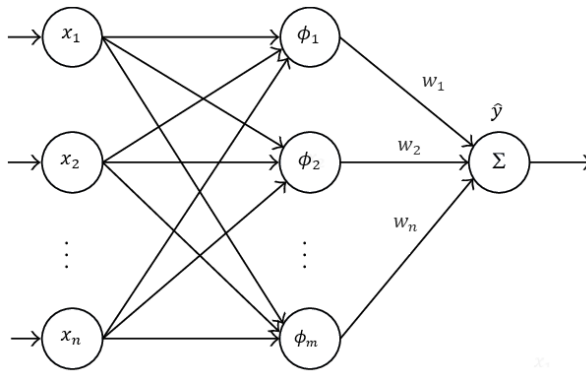


Figura 22.26 - Struttura di una rete RBF

23. DEEP LEARNING

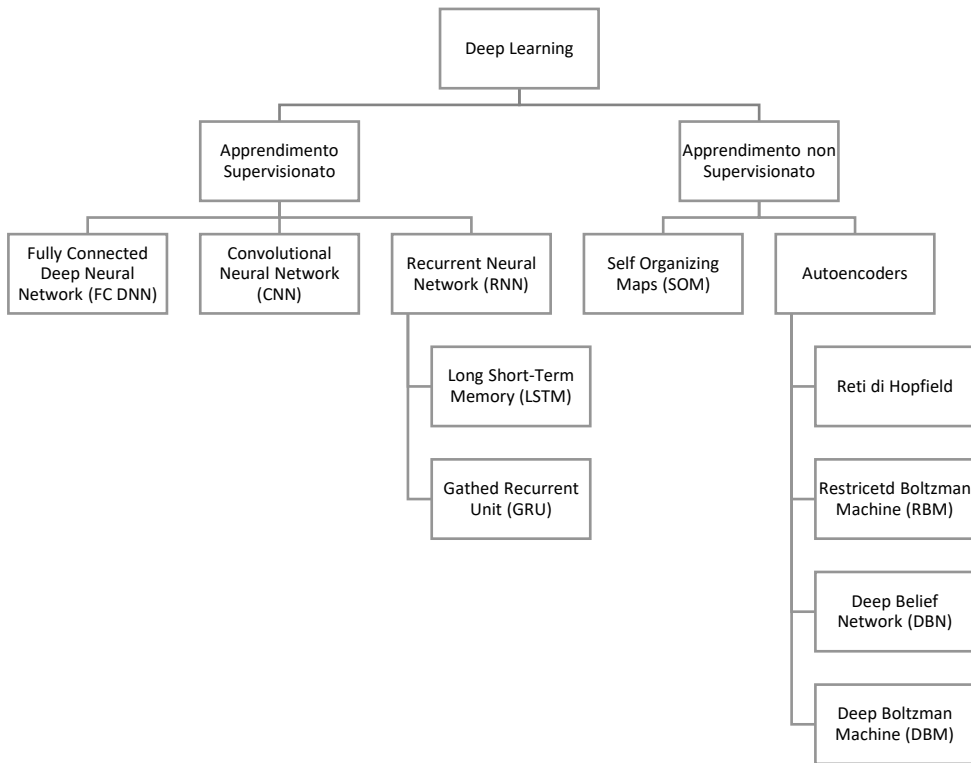


Figura 23.1 - Alcuni modelli di Deep Learning

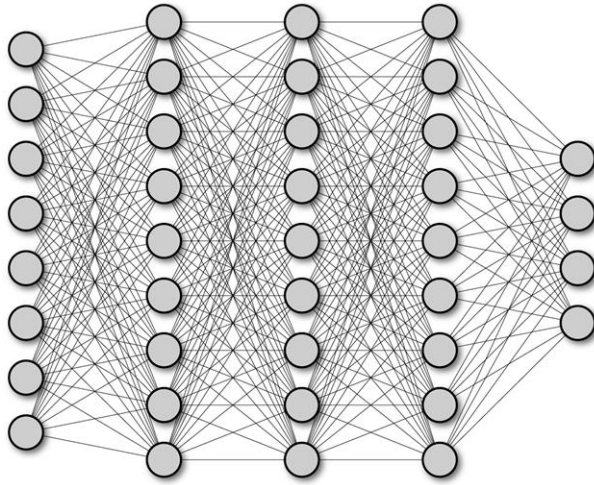


Figura 23.2 - Rappresentazione di una rete profonda completamente connessa



Figura 23.3 - Alcuni esempi dal dataset MNIST

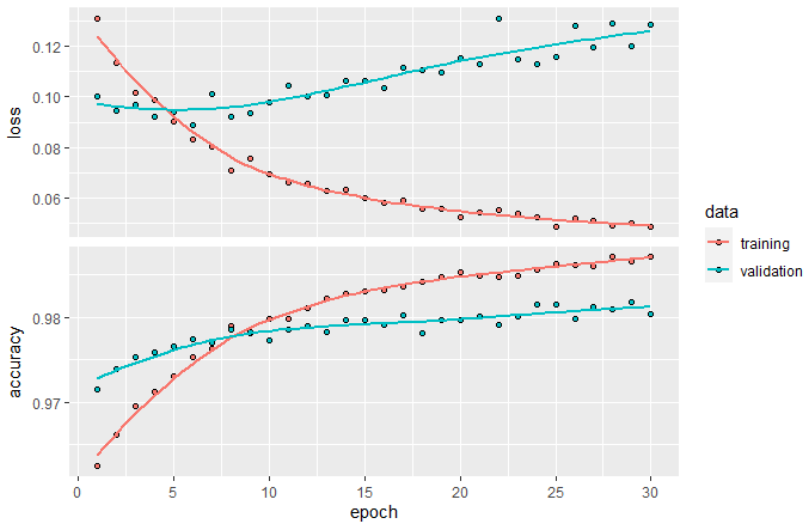


Figura 23.4 - loss e accuracy del modello

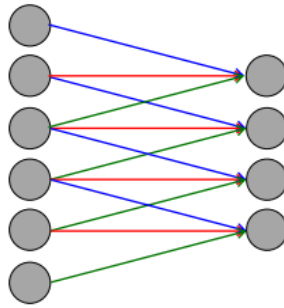


Figura 23.5 - Esempio di pesi condivisi

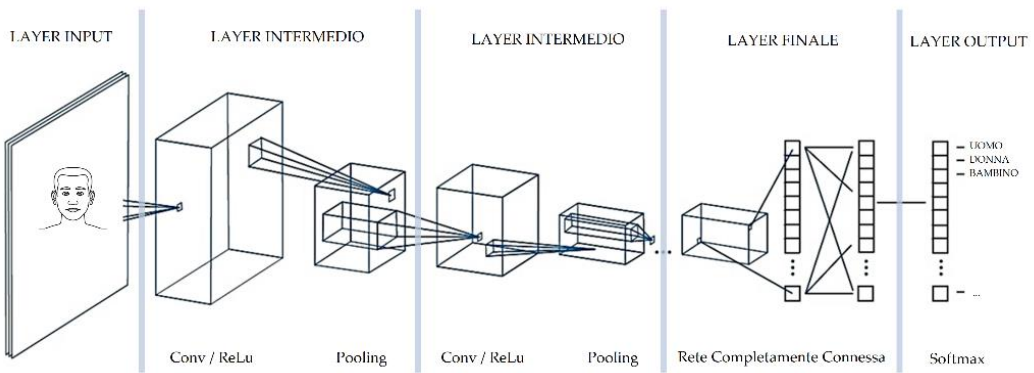


Figura 23.6 - Architettura di una CNN

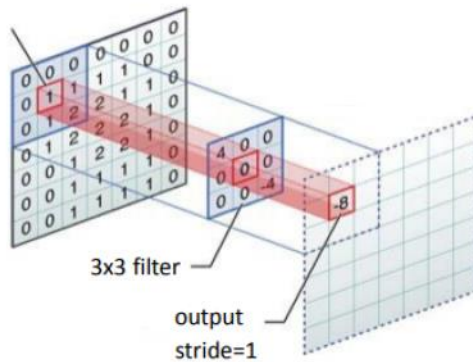


Figura 23.7 - Applicazione del filtro all'immagine

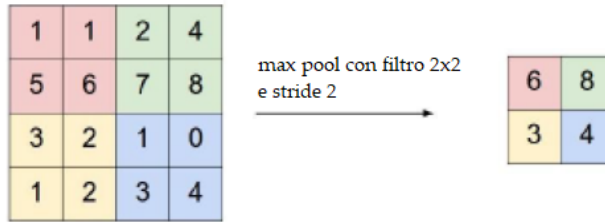


Figura 23.8 - Max pool

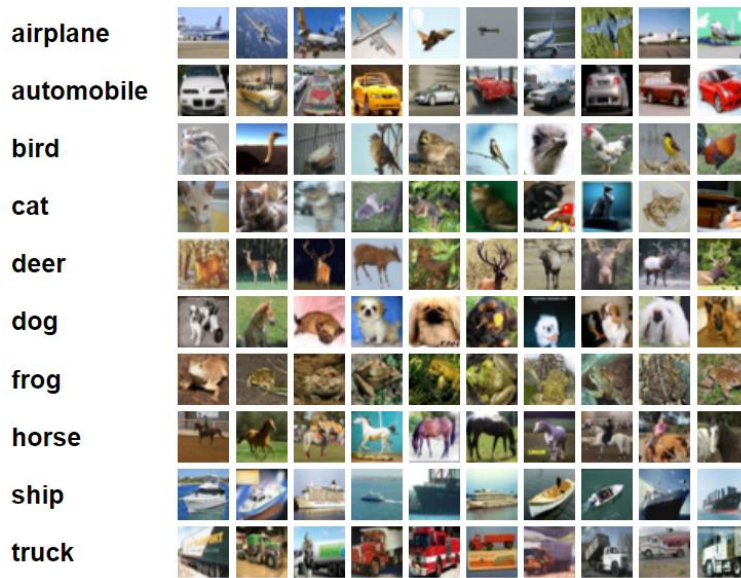


Figura 23.9 - Le classi del dataset, con alcune immagini di esempio

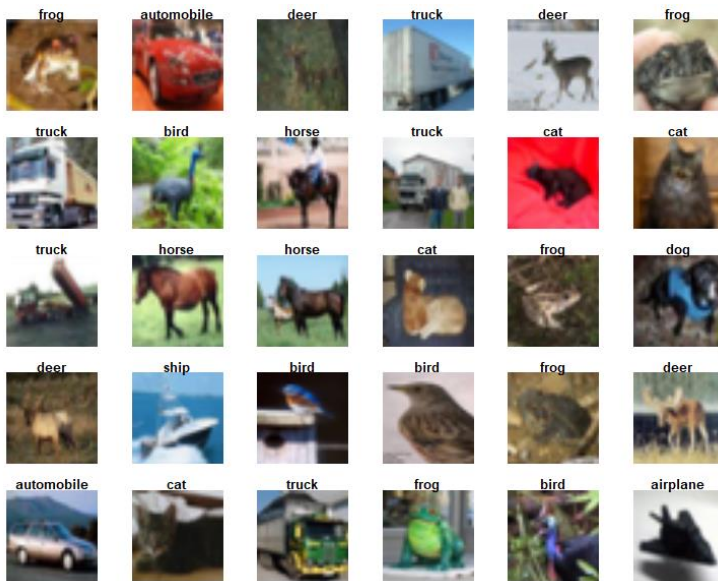


Figura 23.10 - Alcune immagini del dataset

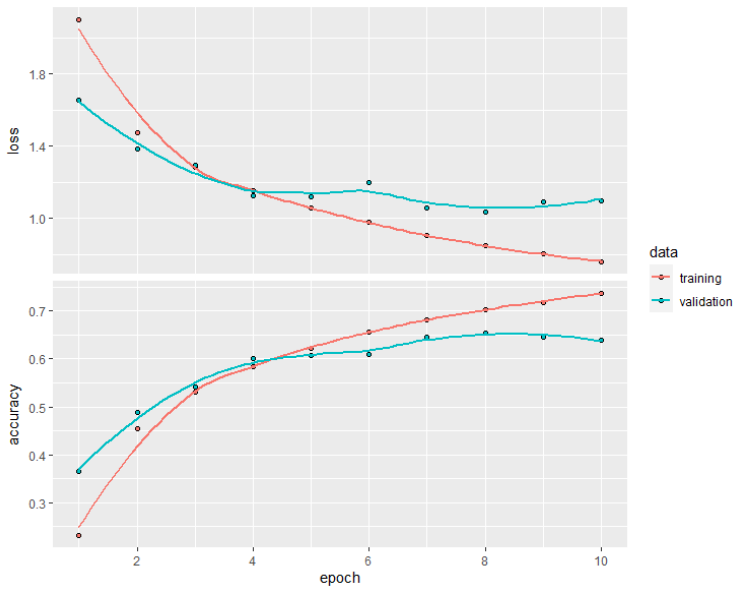


Figura 23.11 - I risultati del modello

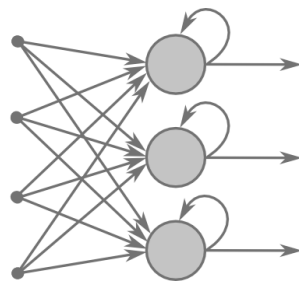


Figura 23.12 - Rete Neurale Ricorrente

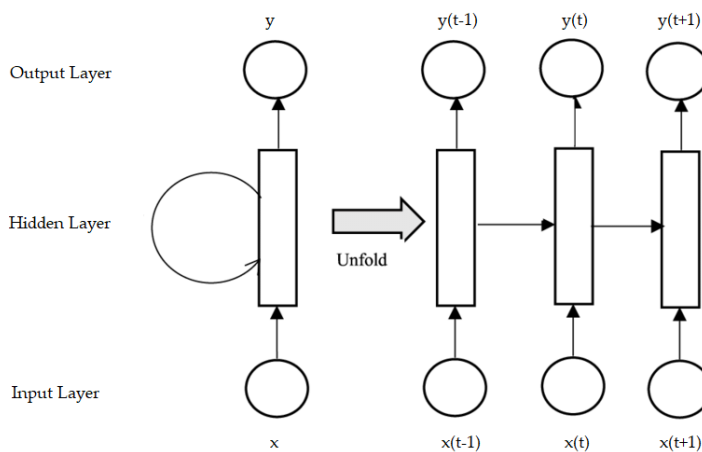


Figura 23.13 - Unfold della rete

	DATE	PRCP	TMAX	TMIN	RAIN
10	1948-01-10	0.74	43	34	TRUE
11	1948-01-11	0.01	42	32	TRUE
12	1948-01-12	0.00	41	26	FALSE
13	1948-01-13	0.00	45	29	FALSE
14	1948-01-14	0.00	38	26	FALSE
15	1948-01-15	0.00	34	31	FALSE
16	1948-01-16	0.00	34	28	FALSE

Figura 23.14 - Porzione di dataset

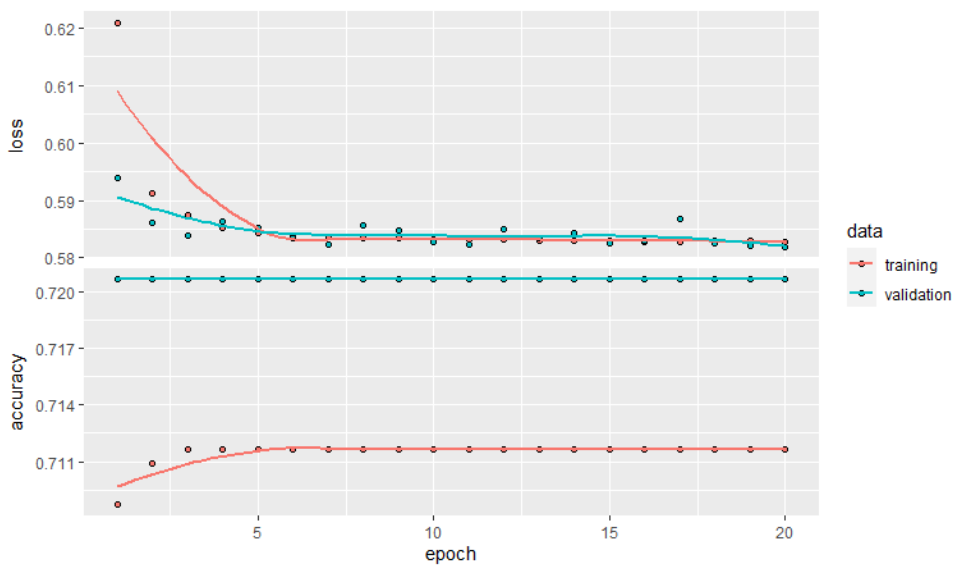


Figura 23.15 - Loss e Accuray durante il training

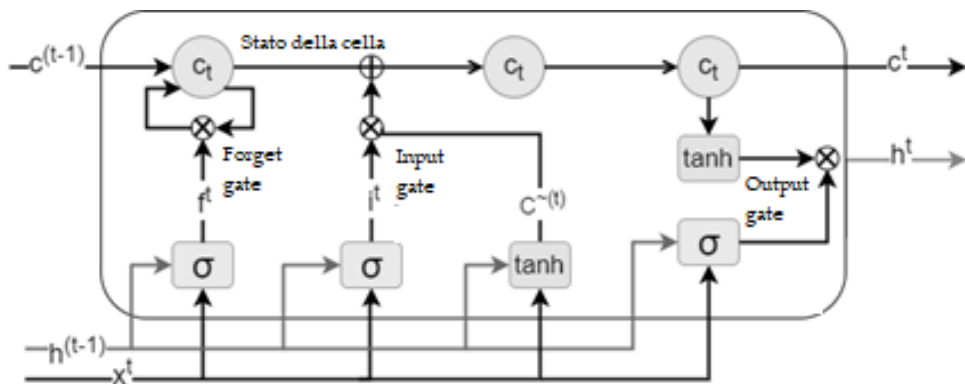


Figura 23.16 - Architettura di una cella di LSTM

review	sentiment
<p>One of the other reviewers has mentioned that after watching just 1 Oz episode you'll be hooked. They are right, as this is exactly what happened with me.</p> <p>The first thing that struck me about Oz was its brutality and unflinching scenes of violence, which set in right from the word GO. Trust me, this is not a show for the faint hearted or timid. This show pulls no punches with regards to drugs, sex or violence. Its is hardcore, in the classic use of the word.</p> <p>It is called OZ as that is the nickname given to the Oswald Maximum Security State Penitentiary. It focuses mainly on Emerald City, an experimental section of the prison where all the cells have glass fronts and face inwards, so privacy is not high on the agenda. Em City is home to many..Aryans, Muslims, gangstas, Latinos, Christians, Italians, Irish and more....so scuffles, death stares, dodgy dealings and shady agreements are never far away.</p> <p>would say the main appeal of the show is due to the fact that it goes where other shows wouldn't dare. Forget pretty pictures painted for mainstream audiences, forget charm, forget romance...OZ doesn't mess around. The first episode I ever saw struck me as so nasty it was surreal, I couldn't say I was ready for it, but as I watched more, I developed a taste for Oz, and got accustomed to the high levels of graphic violence. Not just violence, but injustice (crooked guards who'll be sold out for a nickel, inmates who'll kill on order and get away with it, well mannered, middle class inmates being turned into prison bitches due to their lack of street skills or prison experience) Watching Oz, you may become comfortable with what is uncomfortable viewing....thats if you can get in touch with your darker side.</p>	positive
<p>A wonderful little production.</p> <p>The filming technique is very unassuming- very old-time-BBC fashion and gives a comforting, and sometimes discomforting, sense of realism to the entire piece.</p> <p>The actors are extremely well chosen- Michael Sheen not only "has got all the polari" but he has all the voices down pat too! You can truly see the seamless editing guided by the references to Williams' diary entries, not only is it well worth the watching but it is a terrifically written and performed piece. A masterful production about one of the great master's of comedy and his life.</p> <p>The realism really comes home with the little things: the fantasy of the guard which, rather than use the traditional 'dream' techniques remains solid then disappears. It plays on our knowledge and our senses, particularly with the scenes concerning Orton and Halliwell and the sets (particularly of their flat with Halliwell's murals decorating every surface) are terribly well done.</p>	positive
<p>I thought this was a wonderful way to spend time on a too hot summer weekend, sitting in the air conditioned theater and watching a light-hearted comedy. The plot is simplistic, but the dialogue is witty and the characters are likable (even the well bread suspected serial killer). While some may be disappointed when they realize this is not Match Point 2: Risk Addiction, I thought it was proof that Woody Allen is still fully in control of the style many of us have grown to love.</p> <p>This was the most I'd laughed at one of Woody's comedies in years (dare I say a decade?). While I've never been impressed with Scarlet Johanson, in this she managed to tone down her "sexy" image and jumped right into a average, but spirited young woman.</p> <p>This may not be the crown jewel of his career, but it was wittier than "Devil Wears Prada" and more interesting than "Superman" a great comedy to go see with friends. Basically there's a family where a little boy (Jake) thinks there's a zombie in his closet & his parents are fighting all the time.</p> <p>This movie is slower than a soap opera... and suddenly, Jake decides to become Rambo and kill the zombie.</p> <p>OK, first of all when you're going to make a film you must Decide if its a thriller or a drama! As a drama the movie is watchable. Parents are divorcing & arguing like in real life. And then we have Jake with his closet which totally ruins all the film! I expected to see a BOOGEYMAN similar movie, and instead i watched a drama with some meaningless thriller spots.</p> <p>3 out of 10 just for the well playing parents & descent dialogs. As for the shots with Jake: just ignore them.</p>	negative

Figura 23.17 - Alcune recensioni dal dataset IMDB

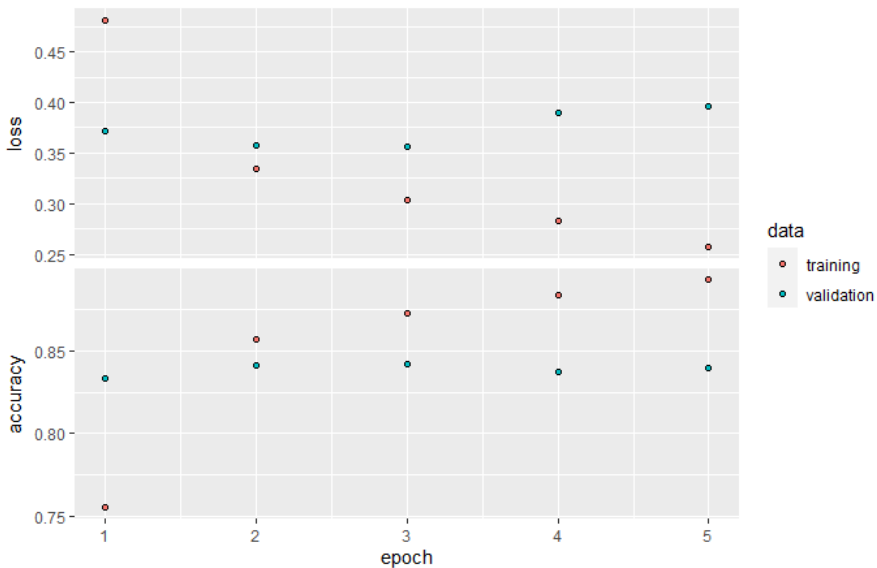


Figura 23.18 - Risultati del training

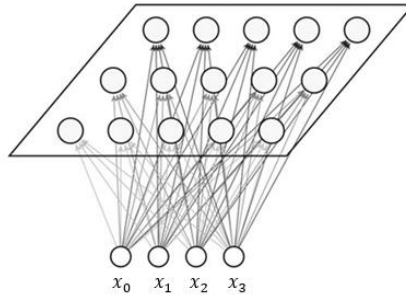


Figura 23.19 - Struttura di una SOM



Figura 23.20 - Una rappresentazione del neurone BMU e i suoi neighbors

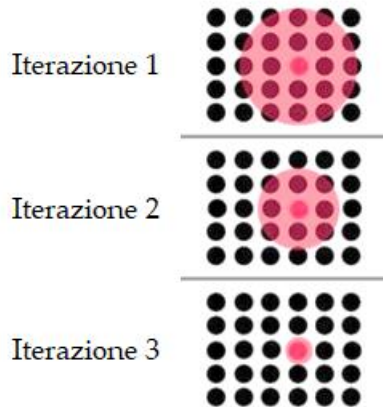


Figura 23.21 - Effetto della contrazione di $h_{i,j}(n)$

	alcohol	malic acid	ash	ash alkalinity	magnesium	tot. phenols	flavonoids	non-flav. phenols	proanth	col. int.	col. hue	OD ratio	proline
1	13.20	1.78	2.14	11.2	100	2.65	2.76	0.26	1.28	4.38	1.050	3.40	1050
2	13.16	2.36	2.67	18.6	101	2.80	3.24	0.30	2.81	5.68	1.030	3.17	1185
3	14.37	1.95	2.50	16.8	113	3.85	3.49	0.24	2.18	7.80	0.860	3.45	1480
4	13.24	2.59	2.67	21.0	118	2.80	2.69	0.39	1.82	4.32	1.040	2.93	735
5	14.20	1.76	2.45	15.2	112	3.27	3.39	0.34	1.97	6.75	1.050	2.85	1450
6	14.39	1.87	2.45	14.6	96	2.50	2.52	0.30	1.98	5.25	1.020	3.58	1290
7	14.06	2.15	2.61	17.6	121	2.60	2.51	0.31	1.25	5.05	1.060	3.58	1295
8	14.83	1.64	2.17	14.0	97	2.80	2.98	0.29	1.98	5.20	1.080	2.65	1045
9	13.86	1.35	2.27	16.0	98	2.98	3.15	0.22	1.85	7.22	1.010	3.55	1045
10	14.10	2.16	2.30	18.0	105	2.95	3.32	0.22	2.38	5.75	1.250	3.17	1510
11	14.12	1.48	2.32	16.8	95	2.20	2.43	0.26	1.57	5.00	1.170	2.82	1280
12	13.75	1.73	2.41	16.0	89	2.60	2.76	0.29	1.81	5.60	1.150	2.90	1320
13	14.75	1.73	2.39	11.4	91	3.10	3.69	0.43	2.81	5.40	1.250	2.73	1150

Figura 23.22 - Il dataset wine

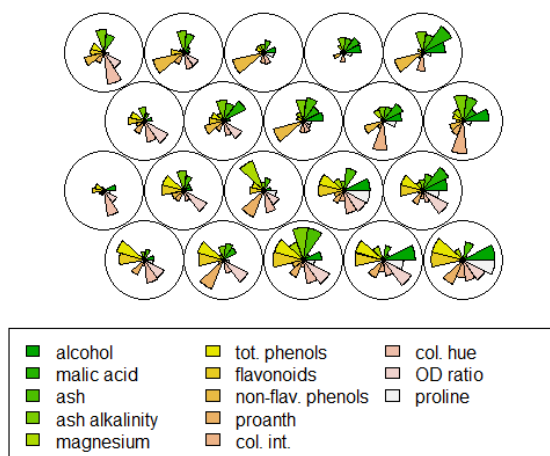


Figura 23.23 - Plot della SOM

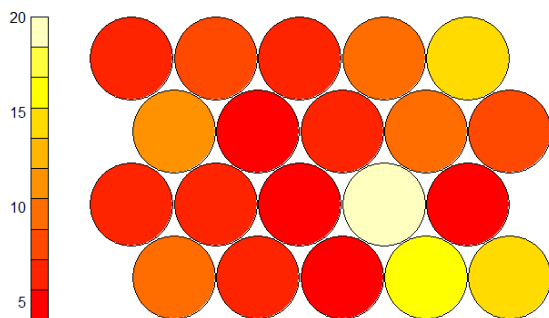


Figura 23.24 - Numero di oggetti mappati per ogni unità

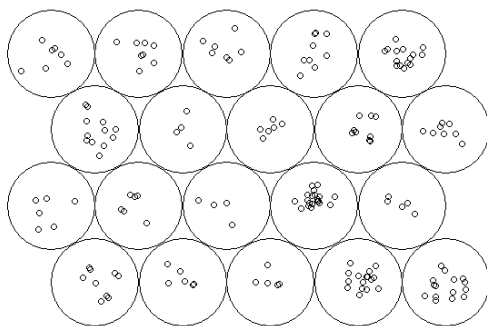


Figura 23.25 - Mapping degli oggetti

mapping plot

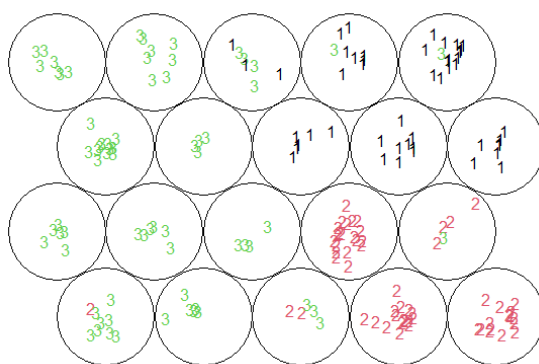


Figura 23.26 - Mapping delle classi

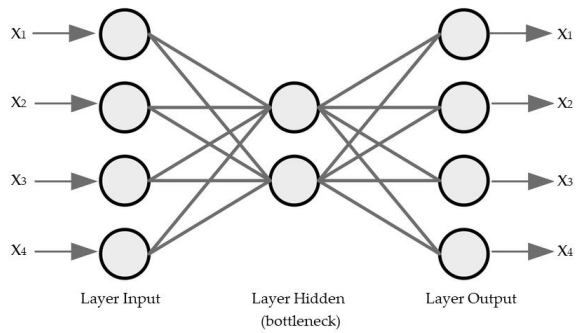


Figura 23.27 - Struttura dell'autoencoder

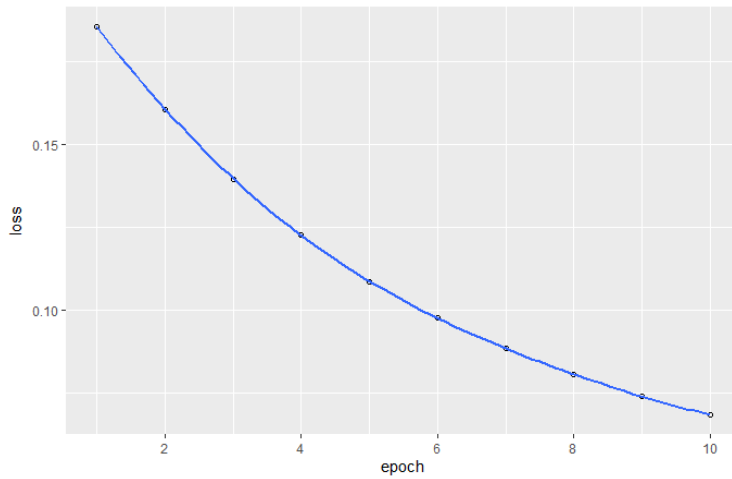


Figura 23.28 - Loss del modello

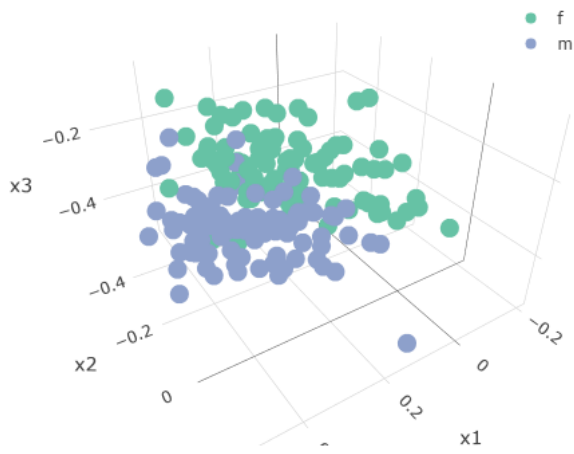


Figura 23.29 - Risultati dell'encoding

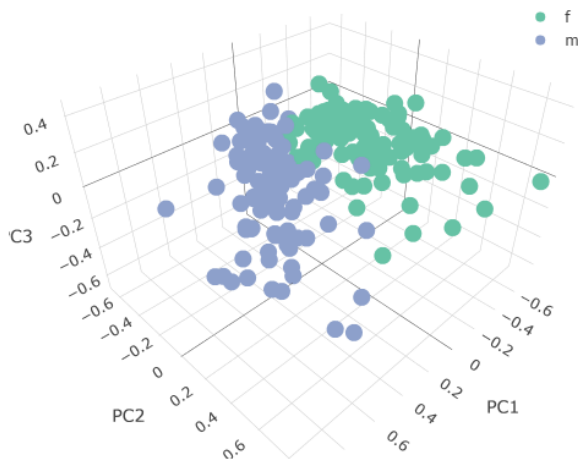


Figura 23.30 - Risultati della PCA

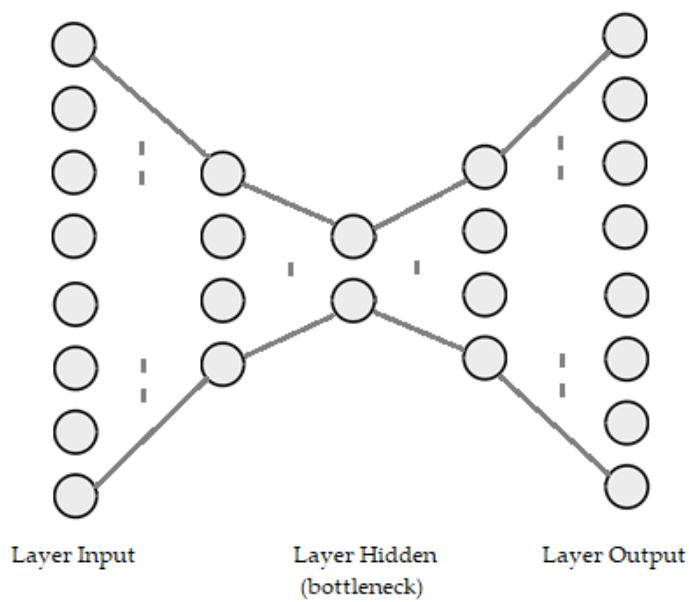


Figura 23.31 - Esempio di Stacked Autoencoder

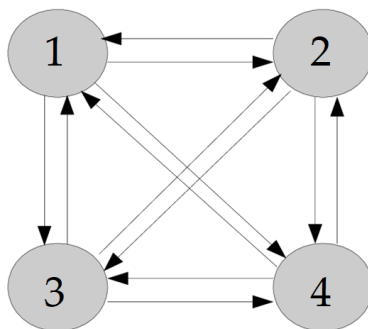


Figura 23.32 - Esempi di rete di Hopfield

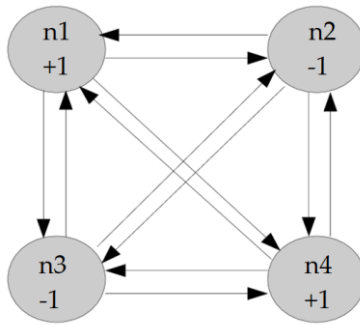


Figura 23.33 - Stato della rete

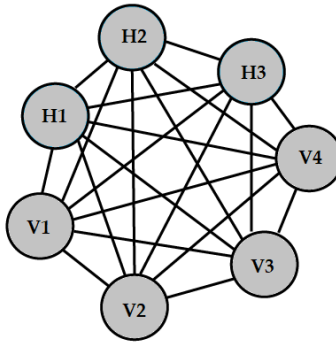


Figura 23.34 - Struttura di una Boltzmann Machine

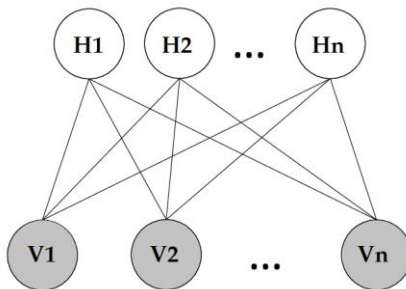


Figura 23.35 - Struttura di una RBM

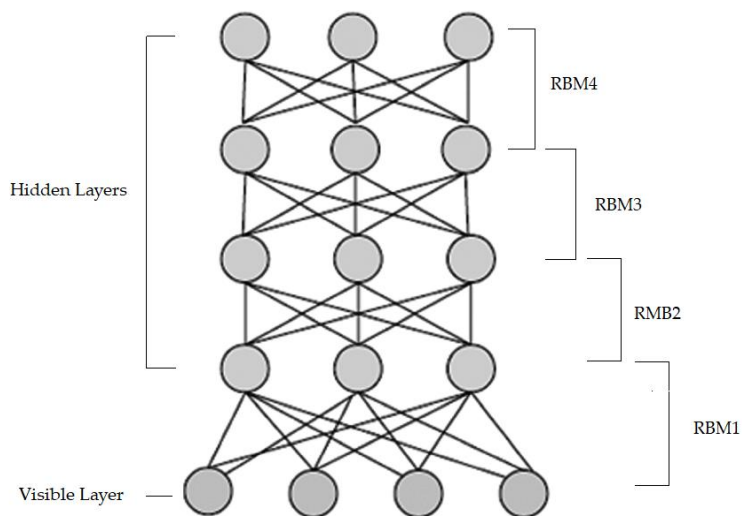


Figura 23.36 - Architettura di una rete DBN

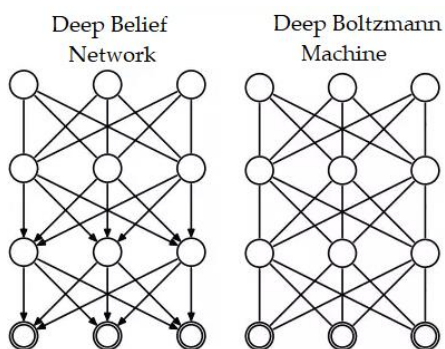


Figura 23.37 - DBN vs DBM

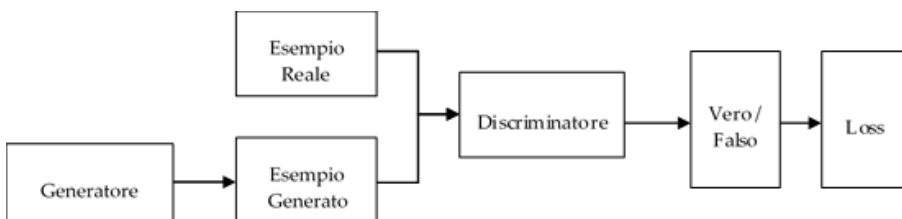


Figura 23.38 - Esempio di Rete Generativa Avversaria



Figura 23.39 - Un'immagine rappresentante un elefante africano

24. AUTOMATED MACHINE LEARNING

25. VISUALIZZARE I DATI



Figura 25.1 - Grafico a barre verticale e orizzontale

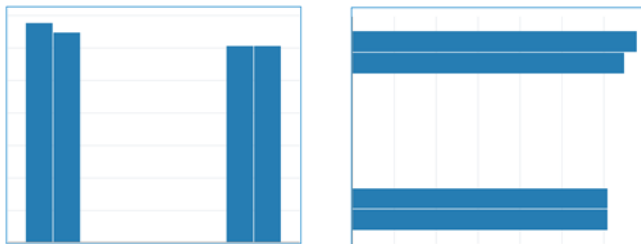


Figura 25.2 - Grafici a barre raggruppati

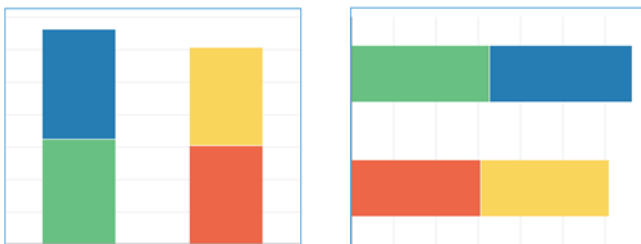


Figura 25.3 - Grafici a barre impilati

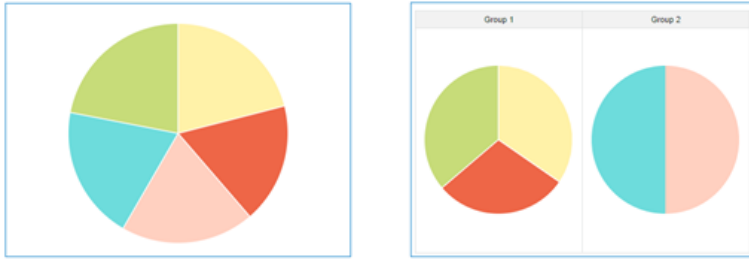


Figura 25.4 - Grafico a torta in configurazione singola e multipla

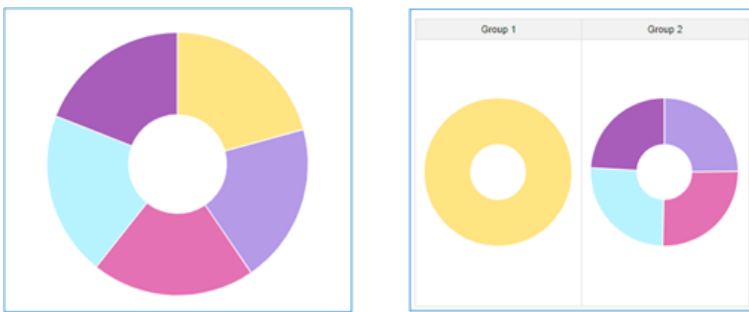


Figura 25.5 - Grafico a ciambella



Figura 25.6 - Grafici stacked al 100%

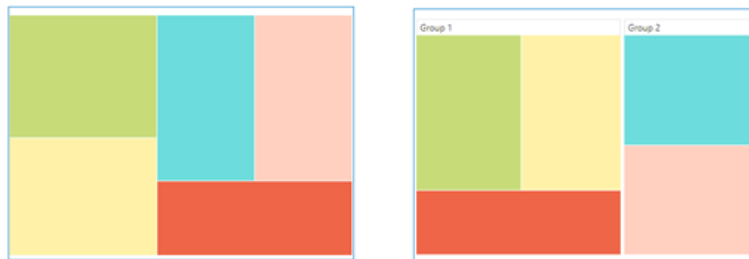


Figura 25.7 - Grafici treemap

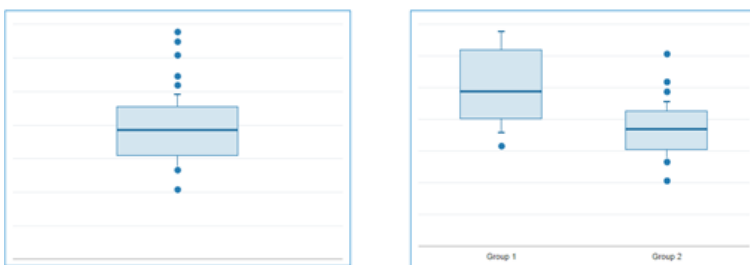


Figura 25.8 - Grafici boxplot

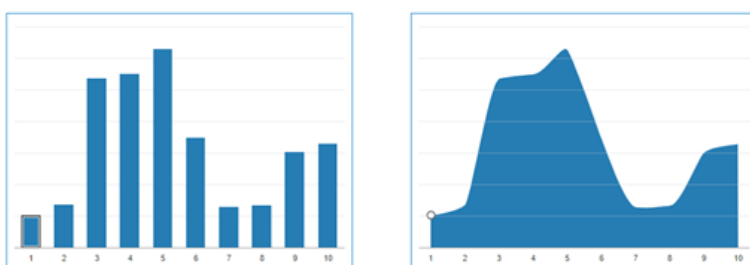


Figura 25.9 - Istogramma e density plot

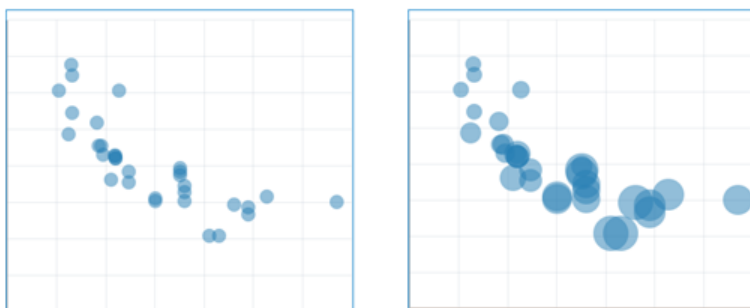


Figura 25.10 - Scatterplot e Bubble Chart

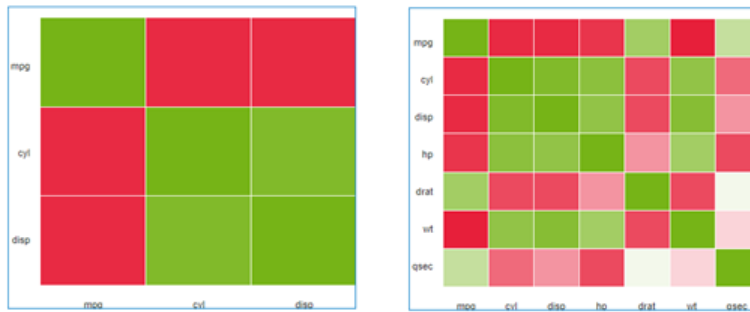


Figura 25.11 - Correlogramma per quattro e sette variabili

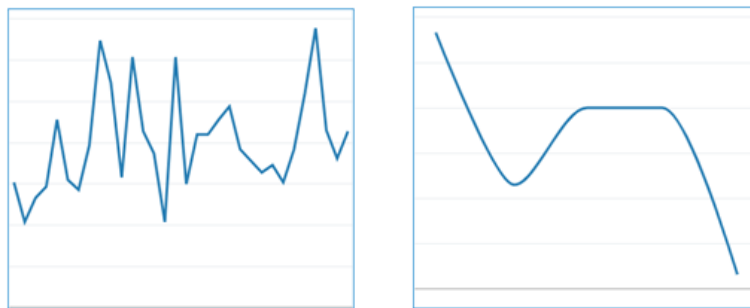


Figura 25.12 - Grafici a linee spezzate e curve

26. COMUNICARE I DATI

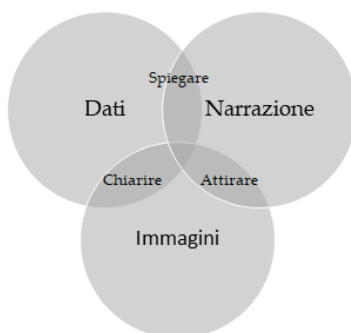


Figura 26.1 - Data Storytelling

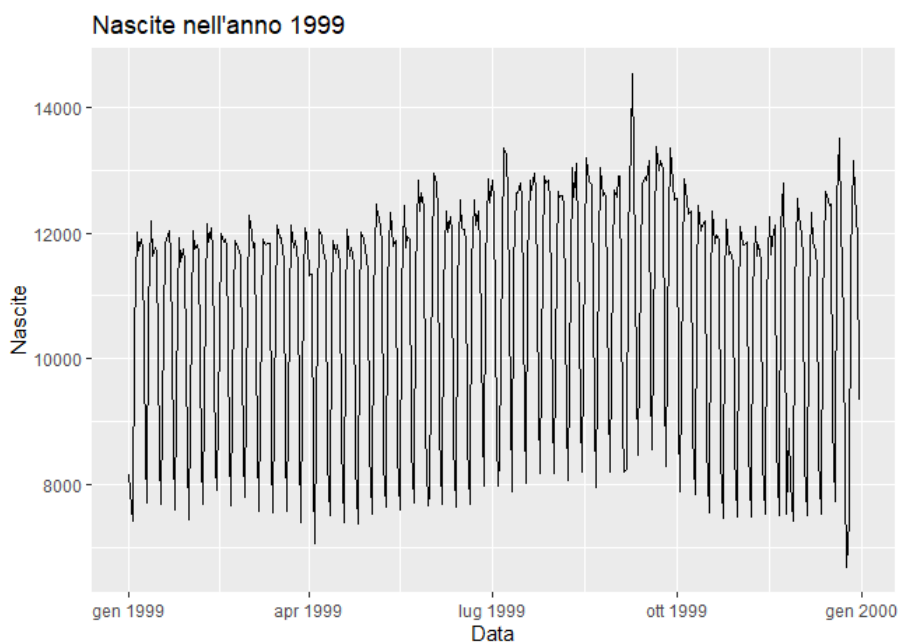


Figura 26.2 - Nascite negli Stati Uniti nell'anno 1999

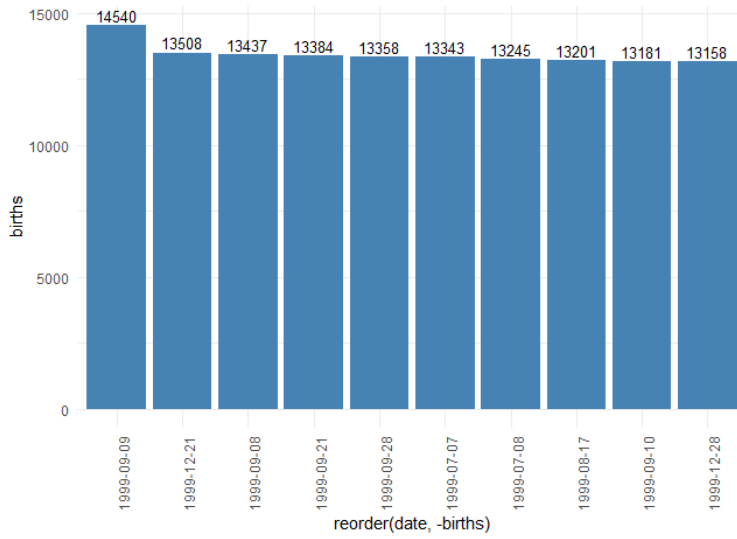


Figura 26.3 - I 10 giorni con più nascite

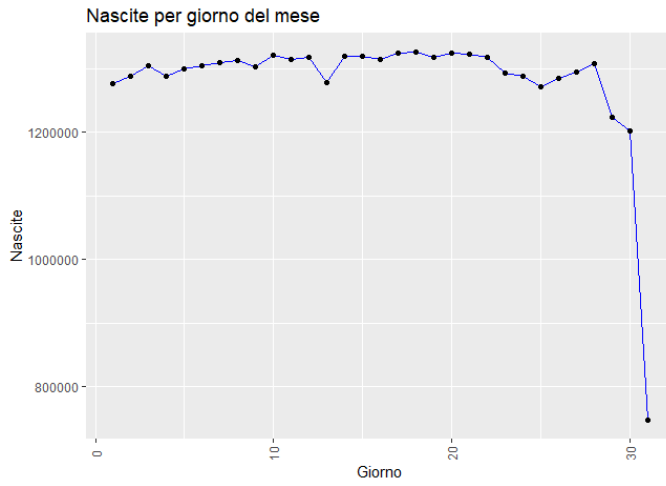


Figura 26.4 - Nascite per giorno del mese

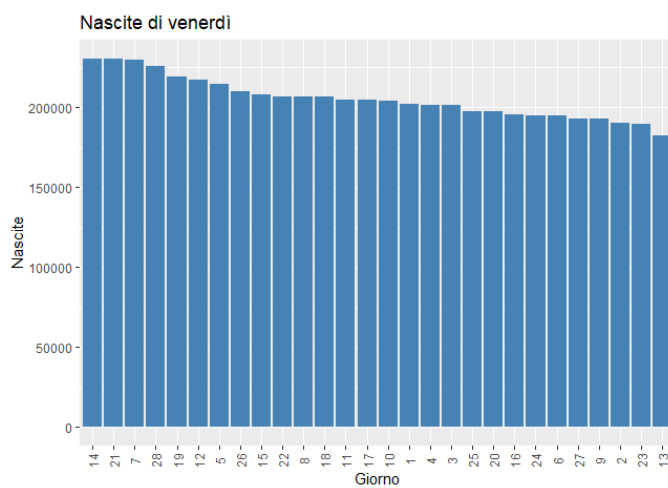


Figura 26.5 - Nascite di venerdì per giorno del mese