



UNIVERSITÀ DEGLI STUDI ROMA TRE
DIPARTIMENTO DI INFORMATICA E
AUTOMAZIONE

Classificazione e Predizione

Data: 07/05/2009

- 1) Supponiamo di eliminare l'attributo *nome* dall'insieme di dati descritto nella tabella, e di addestrare in questo modo un classificatore Bayesiano naive.

Nome	Numero Zampe	Vola	Tipo
corvo	2	Y	uccello
tirannosauro	2	N	dinosauro
cane	4	N	mammifero
pegaso	4	Y	animale legendario

- a. Data una generica istanza “numero zampe = n, vola = v”, qual è la probabilità che la classe predetta sia un mammifero?
- b. Poiché probabilità nulle sono spesso indesiderate, quale metodo può essere utilizzato per eliminare questo inconveniente?

2) Si consideri il seguente insieme di dati:

ID	X	Y
A	0	0
B	0	0.5
C	2	1
D	0.2	0.5
E	0	-0.8
F	2	2

Applicare l'algoritmo k-means. Il raggruppamento deve avvenire solo sulla base degli attributi x e y. Lo studente può scegliere a suo piacimento il tipo di misura di distanza da utilizzare.

3) Sia dato il seguente insieme di dati

Animale	Vola	Acquatico	Zampe	Mammifero
Colomba	S	N	S	N
Mosca	S	N	S	N
Delfino	N	S	N	S
Squalo	N	S	N	N
Lucertola	N	N	S	N
Cane	N	N	S	S
Pipistrello	S	N	S	S

Si supponga di addestrare con questi dati un classificatore Bayesiano naive, considerando l'attributo mammifero come attributo di classe e ignorando l'attributo animale. Quale classe verrebbe predetta in presenza della seguente istanza?

Animale	Vola	Acquatico	Zampe
Negumi	S	S	S

4) Si consideri il seguente insieme di dati:

ID	X	Y
A	0	0
B	0	0.5
C	2	1
D	0.2	0.5
E	0	-0.8
F	2	2

Applicare l'algoritmo k-means con $k=2$, scegliendo come centri iniziali i punti A e B. Il raggruppamento deve avvenire solo sulla base degli attributi x e y. Lo studente può scegliere a suo piacimento il tipo di distanza da utilizzare.

- 5) Sia dato il seguente insieme di dati, nel quale l'attributo *stipendio* è da considerare l'attributo di classe:

Nome	Età	Anni servizio	Dipartimento	Stipendio
Gianluca	30	10	Ricerca	basso
Carla	50	20	Ricerca	alto
Lucia	50	25	Vendite	alto
Michela	40	10	Vendite	alto

Dopo aver normalizzato i dati di tipo numerico, calcolare il tasso di errore utilizzando il metodo di leave one out cross validation e l'algoritmo 1-nearest neighbour

6) Sia dato il seguente insieme di dati:

ID	X	Y
A	0	0
B	0	1
C	2	1
D	2	2
E	2	3

Applicare l'algoritmo k-medoids con $k=2$, considerando i punti C ed E come mediodi iniziali.

7) Descrivere gli algoritmi di bagging e boosting usati per la combinazione di classificatori.