

Data Mining - Corsi di Laurea Specialistica in
Informatica per l'Economia e l'Azienda

Verifica – 28 maggio 2009

Esercizio 1 – Classification – alberi di decisione (15 punti)

- a) I responsabili del settore di marketing considerano FIDELIZZATI i clienti i cui acquisti semestrali sono superiori a 1000 Euro e con acquisti mensili costantemente superiori a 100 Euro. Tutti gli altri sono considerati OCCASIONALI. Si illustri il processo di costruzione di un albero di decisione in modo di prevedere i clienti fidelizzati a partire dalle due tabelle sottostanti. La prima è relativa all'anagrafica clienti e la seconda allo storico degli acquisti fatti negli ultimi 6 mesi. Si costruisca un albero di decisione per il problema proposto utilizzando un opportuno training set definito al passo precedente.

Esercizio 1: Tabella Anagrafica Clienti

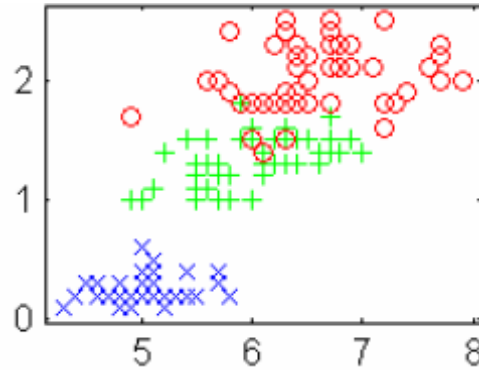
Client id	Study	INCOME	STATUS	AGE
1	Laurea	High	Married	20-30
2	Laurea	Medium	Married	20-30
3	Media Inferiore	Low	Married	20-30
4	Laurea	High	Married	40-50
5	Media Superiore	Low	Single	20-30
6	Media Superiore	Medium	Single	40-50
7	Media Inferiore	Low	Married	40-50
8	Media Inferiore	High	Married	20-30
9	Media Superiore	High	Single	20-30
10	Media Inferiore	Low	Single	20-30
11	Laurea	Medium	Single	20-30
12	Media Inferiore	High	Single	40-50
13	Media Inferiore	Medium	Married	40-50
14	Media Superiore	Low	Married	40-50

Esercizio1: Tabella Acquisti mensili

Cliente id	Gennaio	Febbraio	Marzo	Aprile	Maggio	Giugno	Totale
1	257	357	256	0	262	249	1381
2	173	106	284	157	177	139	1038
3	182	193	197	632	275	173	1655
4	104	50	160	23	289	289	919
5	103	589	157	156	182	247	1439
6	157	193	230	124	194	201	1105
7	29	89	179	0	86	80	470
8	304	284	190	89	314	248	1437
9	356	298	283	293	232	186	1657
10	192	163	176	154	553	249	1497
11	159	112	143	121	575	164	1285
12	267	261	287	273	258	327	1685
13	164	136	284	269	201	154	1221
14	59	84	0	57	999	74	1287

Esercizio 2 – Classificazione – Alberi di decisione (9 punti)

Si consideri il seguente training set bi-dimensionale (sia X l'attributo rappresentato sull'asse orizzontale, e Y l'attributo rappresentato sull'asse verticale.)



a) Si proponga un albero di decisione con attributo target a valori \circ , X e $+$, e se ne discuta l'accuratezza .

Esercizio 3 – Classificazione – Valutazione di performance (9 punti)

Si consideri il seguente albero di decisione e si costruisca la matrice di confusione relativamente al test set indicato in Tabella 2. Si forniscano, inoltre la misura di *misclassification rate* e l'accuratezza.

Albero di classificazione

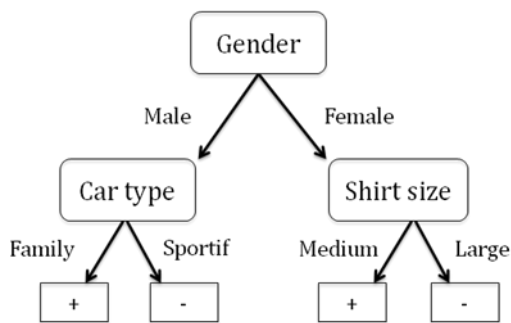


Tabella 2. Test set

Instance	Gender	Car type	Shirt size	Class
11	Male	Family	Medium	+
12	Male	Sportif	Large	+
13	Female	Sportif	Medium	+
14	Female	Family	Large	-
15	Female	Family	Medium	+