

Esercizio 1

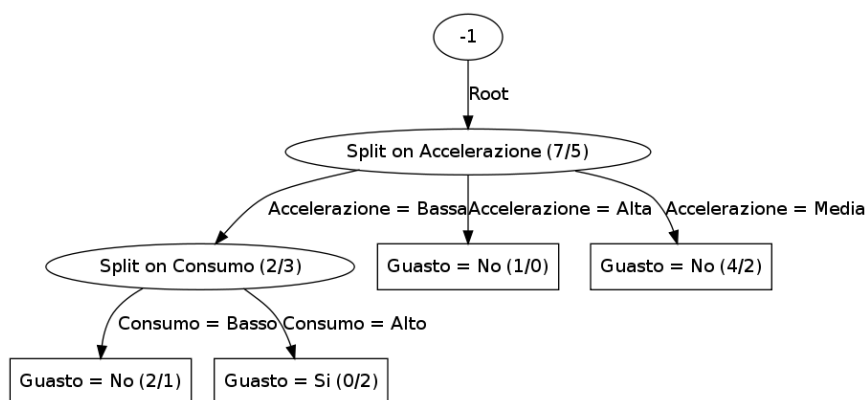
Si consideri il seguente insieme di transazioni (*training set*).

Consumo	Nuova	Accelerazione	Risposta	Guasto
Alto	No	Media	Lenta	Si
Alto	No	Bassa	Lenta	Si
Alto	Si	Media	Lenta	No
Alto	No	Alta	Veloce	No
Alto	No	Media	Lenta	No
Basso	No	Bassa	Veloce	Si
Alto	No	Media	Veloce	Si
Basso	No	Bassa	Veloce	No
Basso	Si	Bassa	Lenta	No
Alto	Si	Media	Veloce	No
Alto	No	Media	Veloce	No
Alto	Si	Bassa	Veloce	Si

- A) Si costruisca su tale dataset un albero di decisione per la variabile “Guasto”, utilizzando il criterio di split basato su “misclassification rate”, espandendo i nodi dell'albero fino a che la precisione non è più migliorabile localmente (ovvero nessuno split da' un guadagno).
- B) Si valuti l'accuratezza dell'albero ottenuto al punto A) tramite matrice di confusione, calcolata sia sul training set che sul test set riportato qui sotto. Confrontare i risultati.

Consumo	Nuova	Accelerazione	Risposta	Guasto
Alto	No	Bassa	Lenta	Si
Basso	Si	Bassa	Veloce	No
Alto	Si	Alta	Veloce	No
Alto	No	Alta	Veloce	Si
Basso	No	Alta	Lenta	No
Alto	No	Media	Lenta	No
Basso	Si	Media	Lenta	Si
Alto	Si	Media	Lenta	No
Basso	No	Alta	Lenta	Si
Alto	Si	Alta	Veloce	No

Soluzione



3 errori su 12 = precisione 75% su training
2 errori su 10 = precisione 80% su test

Esercizio 2

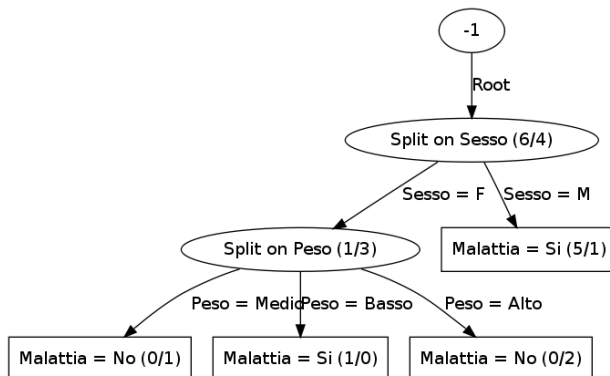
Si consideri il seguente insieme di transazioni (*training set*).

Altezza	Peso	Età	Sesso	Malattia
Bassa	Alto	Giovane	F	No
Bassa	Basso	Giovane	F	Si
Bassa	Basso	Anziano	M	No
Bassa	Medio	Giovane	M	Si
Bassa	Alto	Giovane	M	Si
Alta	Medio	Anziano	F	No
Bassa	Alto	Giovane	F	No
Alta	Basso	Anziano	M	Si
Alta	Basso	Anziano	M	Si
Bassa	Medio	Anziano	M	Si

- A) Si costruisca su tale dataset un albero di decisione per la variabile “Malattia”, utilizzando il criterio di split basato su “misclassification rate”, espandendo i nodi dell'albero fino a che la precisione non è più migliorabile localmente (ovvero nessuno split da' un guadagno).
- B) Si valuti l'accuratezza dell'albero ottenuto al punto A) tramite matrice di confusione, calcolata sia sul training set che sul test set riportato qui sotto. Confrontare i risultati.

Altezza	Peso	Età	Sesso	Malattia
Alta	Alto	Anziano	F	No
Bassa	Alto	Anziano	F	No
Bassa	Basso	Giovane	F	No
Alta	Basso	Giovane	M	Si
Alta	Medio	Anziano	M	Si
Alta	Basso	Giovane	M	No
Alta	Medio	Anziano	F	No
Alta	Medio	Anziano	F	No

Soluzione



1 errore su 10 = precisione 90% su training
2 errori su 8 = precisione 75% su test

	Si	No
Si	6	0
No	1	3

	Si	No
Si	2	0
No	2	4

Esercizio 3

Si consideri il seguente insieme di transazioni con attributi sia discreti che continui:

Cost	Duration	Class
High	<input type="checkbox"/>	No
High	<input type="checkbox"/>	No
High	<input type="checkbox"/>	No
High	<input type="checkbox"/>	Yes
Low	<input type="checkbox"/>	No
Low	<input type="checkbox"/>	Yes
Low	<input type="checkbox"/>	Yes
Low	<input type="checkbox"/>	Yes

Si costruisca un albero di decisione per la variabile target “Class”, utilizzando come criterio di split il “Misclassification Rate” e terminando la costruzione quando la precisione dell'albero non è più migliorabile. **(10 punti)**

Esercizio 4

Si consideri il seguente insieme di transazioni con un unico attributo continuo (oltre alla classe):

Value	Class
7	No
12	No
20	No
25	Yes
31	Yes
40	Yes
45	No
80	No
81	No
85	Yes
90	Yes

Si costruisca un albero di decisione per la variabile target “Class”, utilizzando come criterio di split il “Misclassification Rate” e terminando la costruzione quando la precisione dell'albero non è più migliorabile.

Soluzione

