

Esercizio 1 (punti 6 su 30)

Disegnare un percettrone, eventualmente multilivello, con indicazione dei pesi e delle step function, che calcoli la funzione booleana $(A \text{ AND } B) \text{ OR } (C \text{ XOR } D)$.

Esercizio 2 (punti 7 su 30)

Dato il seguente dataset di training, dove l'attributo *Gioca* è quello dipendente:

Tempo	Temperatura	Umidità	Vento	Gioca
Soleggiato	Calda	Alta	NO	NO
Soleggiato	Calda	Alta	SI	NO
Nuvoloso	Calda	Alta	NO	SI
Piovoso	Mite	Alta	NO	SI
Piovoso	Fredda	Normale	NO	SI
Piovoso	Fredda	Normale	SI	NO
Nuvoloso	Fredda	Normale	SI	SI
Soleggiato	Mite	Alta	NO	NO
Soleggiato	Fredda	Normale	NO	SI
Piovoso	Mite	Normale	NO	SI
Soleggiato	Mite	Normale	SI	SI
Nuvoloso	Mite	Alta	SI	SI
Nuvoloso	Calda	Normale	NO	SI
Piovoso	Mite	Alta	SI	NO

Costruire un albero di decisione ternario su 4 livelli (radice ed altri 3 livelli) scegliendo, per ogni livello, l'attributo che più di tutti migliora l'indice Gini, ipotizzando di utilizzare per ogni attributo un criterio di splitting binario o ternario, a seconda del numero di valori dell'attributo. Inoltre, per ogni nodo dell'albero indicare la distribuzione dei campioni ed il valore dell'indice Gini.

Esercizio 3 (punti 4 su 30)

Facendo riferimento al dataset dell'esercizio 2 ed usando il concetto di RFD, determinare quale attributo predittivo implica maggiormente l'attributo dipendente *Gioca*. Spiegare il perché.

Esercizio 4 (punti 5 su 30)

Facendo riferimento all'albero di decisione scaturito dall'esercizio 2, indicare supporto e confidenza delle regole di classificazione ottenute.

Esercizio 5 (punti 8 su 30)

Dati i seguenti punti in uno spazio bidimensionale: $(5,3)(1,1)(4,2)(3,6)(1,4)(2,1)(3,2)(1,2)$

Usare l'algoritmo *K-means* per suddividere tali punti in 3 cluster, utilizzando la distanza del coseno e svolgendo 3 iterazioni dell'algoritmo. Per la scelta dei 3 centroidi alla prima iterazione utilizzare il metodo che prevede la selezione del solo primo centroide in modo casuale (scegliere il punto più vicino a quello le cui coordinate si ottengono sommando tutte le prime componenti e poi le seconde degli 8 punti forniti), mentre i restanti 2 il più lontano possibile dai centroidi già scelti.