

Analisi dei Dati ed Estrazione della Conoscenza

Compito d'esame a.a. 2005/2006

Lo studente svolga (a scelta) una delle seguenti prove finali. Le prove possono anche essere svolte in gruppi di due persone.

Prova finale 1

1. Implementare in R il metodo di classificazione bayesiana naive per soli attributi categoriali. In pratica, si tratta di realizzare due funzioni:
 - a) `bayes.train(class, data)` che calcola un modello di classificazione bayesiano dato l'insieme di dati (fornito nel parametro `data` di tipo data frame) e l'indice dell'attributo classe (fornito nel parametro `class` di tipo intero o stringa);
 - b) `bayes.predict(model, newdata)` che classifica le nuove istanze fornite in `newdata`, utilizzando il modello fornito in `model`. Il risultato deve comprendere, per ogni istanza, la distribuzione di probabilità tra le possibili classi. Se in `newdata` è già presente l'attributo classe, utilizzarlo per calcolare il tasso di errore e la perdita quadratica medica.
2. Testare l'implementazione con l'insieme di dati `weather` (la versione con attributi categoriali) e controllare che i risultati siano gli stessi forniti da Weka con il metodo `NaiveBayesSimple`.

Prova finale 2

1. Leggere uno dei seguenti articoli (o combinazioni di articoli) a scelta:
 - a) J. R. Quinlan. *Learning with continuous classes*. Proceedings of AI'92.
Y. Wang, I. Witten. *Inducing Model Trees for Continuous Classes*. Proceedings of Poster Papers, 9 th European Conference on Machine Learning.
 - b) D.H. Fisher. *Knowledge Acquisition Via Incremental Conceptual Clustering*. Machine Learning 2.
2. Progettare e realizzare degli esperimenti in Weka (e/o in R) che mostrino il funzionamento dell'algoritmo studiato, evidenziandone i punti più importanti. Come minimo, è necessario ripetere gli esperimenti presentati nell'articolo, discutendo eventuali incongruenze tra questi e i risultati ottenuti in proprio.
3. Scrivere una relazione sugli esperimenti compiuti.