Il classificatore J48nel metodo buildClassifier sulla base delle opzioni costruisce o un c45Prunable classifier tree o un PrunableClassifierTree. Sono questi oggetti che, in realtà svolgono classificano l’istanza. A questi vengono inoltre demandati altri compiti, come la visualizzazione tramite ToString e la capacità di disegnarsi tramite il metodo graph e graphType. I metodi Graph e GraphType sono implementati nella superclasse ClassifierTree

La calsse info Gain split criterion sembra essere quella più promettente. Devo capire chi la usa

Catena: modelSelection ->c4.5Split->InfoGainSplitCriterion

Il calcolo del guadagno informativo per un singolo attributo viene fatto nella classe C45split per gli attributi categorici alla riga 196 per quelli numerici sembra alla riga 308, anche se il codice è parecchio complicato. Devi quindi portare il peso fino a questo punto in modo da moltiplicare il guadagno informativo calcolato da InfoGain Split Criterion.

Cosa mi sembra faccia il codice al 19 gennaio

Il metodo buildClassifier della classe J48 crea un ModelCriterion, in particolare un bin c45ModelSelection se l’opzione di split binari è attiva, un c45ModelSelection altrimenti. Al model criterion vengono passate le istanze da classificare ed il numero minimo di oggetti che devono essere presenti in una singola foglia. Dopodichè se l’opzione di Reduced Error Pruning l’albero di classificazione viene inizializzato a un C45Prunable Classifier tree, altrimenti solo a un PrunableClassifierTree. In entrambi i casi agli alberi viene passato come parametro il ModelCriterion creato precedentemente. Dopodiché viene invocato il metodo buildClassifier sull’albero. Questo metodo invoca il metodo buildTree della superclasse ClassifierTree. Questo metodo decide prima su quale attributo fare lo split, poi lo esegue. Per il primo compito viene usato il metodo selectModel del campo modelSelection. Questo metodo è il centro della selezione dell’attributo in base a quale fare lo split. Crea un vettore di C4.5Split, uno per ogni attributo. Poi, invoca per ogni attributo il metodo BuildClassifier, che crea uno split, il risultato viene valutato sia in termine di validità sia in termine di guadagno informatico**. In particolare alle righe 161 e 163 di C45ModelCriterion viene richiesto il guadagno informativo del C45Split**. Questo viene aggiunto alla media. Infine si verifica se è stato trovato uno split buono e se ne decide il migliore.