Aula5 – Modelos múltiplos, Bagging, Boosting, Decomposição Viés / variância.

1. Compara os ganhos em termos de rácio de erro no dataset soybean do Bagging sobre J48 contra Bagging sobre NaiveBayes. Tentar conjeturar uma justificação!



Relativamente ao clasificador J48, a utilização de um processo de Bagging levou a um ganho de aproximadamente 14%. A aplicação de Bagging trás em geral vantagens em algoritmos instáveis, que sejam suscetíveis a alterações durante o processo de treino. Por oposição, este processo não irá trazer resultados vantajoso em algoritmos lineares e estáveis como o caso do NaiveBeyes.

Como o algoritmo naiveBayes, por utilizar conjuntos de grandes dimensões para treino, apresenta um sensibilidade reduzida, e portanto a sua variância será praticamente inalterada ao utilizar Bagging.

1. Avalie os ganhos de adaboost sobre J48 contra adboost sobre naivebayes.



3) Faço um estudo análogo às alíneas anteriores mas para as diferentes classes do dataset hypothyroid.

4) Apresente a decomposição do erro para as duas alíneas anteriores e articule os argumentos para esses resultados.

5) Considere o dataset Kaggle-creditcard (a descarregar do blackboard). Desenvolva um modelo de previsão para identificar as transações fraudulentas descritas no dataset.

 Deve estudar e selecionar os atributos adequados ao processo. Os atributos apresentados correspondem aos atributos selecionados por um método de redução da dimensão do espaço de representação (ACP).

 O dataset apresenta um pesado desequilíbrio da proporção das classes.

 Apresente resultados para diferentes algoritmos com validação cruzada.