

## Aula4 – Custso, Regressão, Modelos múltiplos

- 1) Considere o dataset “cpu”. Apresente uma avaliação com 10-XVal para árvores de Modelos versus árvores de regressão (Model trees versus Regression trees). Avalie predição com e sem smoothing. Indique usando treino/test o efeito de pruning nos modelos.
- 2) Apresente um estudo sobre o dataset “breast-cancer” com custos. Defina uma matriz de custos que se ajuste a uma realidade e desenvolva um classificador que otimize os resultados para esta interpretação de custos.
- 3) Compare a mesma avaliação com modelos de regras. Invente um dataset de teste com 3 ou 4 casos por forma a conseguir elaborar um trace da previsão efetuada (tanto no modelo de regras como no model tree).
- 4) Compara os ganhos em termos de rácio de erro no dataset soybean do Bagging sobre J48 contra Bagging sobre NaiveBayes. Tentar conjecturar uma justificação!
- 5) Considere o dataset Kaggle-Consultas (a descarregar do blackboard). Desenvolva um modelo de previsão para estimar as faltas a consultas nos centros de saúde descritos no dataset.
  - a. Deve estudar e seleccionar os atributos adequados ao processo
  - b. Avalie cada atributo por forma a seleccionar os atributos indicados para obter boas previsões para o problema descrito.
  - c. Apresente resultados para diferentes algoritmos com validação cruzada.