



15 de julho de 2013

1 Weka

2 Floresta Randômica

- Random Forest
- Random Forest

- Programa que possui uma coleção de algoritmos de aprendizagem de máquina para usar em tarefas de mineração de dados
- Desenvolvido pelo Machine Learning Group da Universidade de Waikato
- O Weka permite que se trabalhe sobre o dataset e traz ferramentas para: pré-processamento, classificação, regressão, clusterização, regras de associação e visualização
- Todos os nossos resultados foram feitos com o apoio do Weka

Floresta Randômica (Random Forest)

- Tudo começou em 1996, com o surgimento do meta-algoritmo de Bagging (Bootstrap Aggregating), por Leo Breiman
- Bootstrap ajuda a reduzir variância e overfitting
 - Escolhe aleatoriamente amostras, D_i , de um conjunto de treino, D
 - Esse conjunto de treino D tem reposição (ou seja, uma amostra pode ser escolhida mais de uma vez)
 - Tamanho do conjunto de treino: n
 - Tamanho do conjunto de amostra: n'
 - Geralmente $n' < n$
 - Quando n' for praticamente n , nota-se uma chance de 63% de aparecer amostras não repetidas. Daqui que surgiu o nome de Bootstrap

Floresta Randômica (Random Forest)

- Em 2001, Breiman lançou o Random Forest
- É um método usado para classificação (e regressão)
- O método faz a construção de várias árvores de decisão, com a diferença de que não usa pruning (poda)
- Algoritmo:
 - Escolhe-se aleatoriamente a amostra (bootstrap) D_i dentre D
 - Constrói a árvore T_i usando amostra D_i
 - Em cada árvore T_i , escolhe aleatoriamente M variáveis e encontra a melhor divisão (split)
- No fim, pode-se
 - pegar o voto majoritário (classificação)
 - calcular a média dos resultados (regressão)