
Random Forest

Fabrício Barth

Setembro de 2018

Ensemble Learning

- Métodos que geram diversos modelos e agregam o seu resultado.
- No caso do Random Forest, são geradas diversas árvores e cada árvore é gerada considerando apenas um sub-conjunto do conjunto de treinamento.

Random Forest

- O algoritmo possui apenas dois parâmetros configuráveis:
 - ★ quantidade de atributos considerados em cada árvore (m_{try}), e;
 - ★ quantidade de árvores (n_{tree}).

Random Forest

Para problemas de classificação e regressão o algoritmo funciona da seguinte forma:

- Cria n_{tree} sub-conjuntos de exemplos a partir do dataset original.
- Para cada sub-conjunto de exemplos cria-se uma árvore de classificação ou regressão sem poda. A criação de cada árvore considera apenas um sub-conjunto de exemplos: m_{try} atributos selecionados aleatoriamente e $2/3$ dos exemplos também selecionados aleatoriamente.

-
- A predição para novos dados acontece pela agregação das predições das n_{tree} árvores.
 - Para problemas de **classificação** é considerado a maioria dos votos.
 - Para problemas de **regressão** é considerado a média dos votos.

Estimativa de erro

- Uma estimativa de erro, usando apenas o conjunto de treinamento, pode ser obtida através do conjunto de treinamento. Ao invés de ser utilizado algum outro método, como *cross-validation*.
- Para cada árvore construída é usado um sub-conjunto de exemplos. $1/3$ dos exemplos são mantidos fora do conjunto de treinamento. Estes exemplos mantidos fora do conjunto de treinamento são utilizados como teste.

Exemplo

<http://rpubs.com/fbarth/exemploRandomForest>

Material de **consulta**

- Liaw and Wiener. Classification and Regression by randomForest. R News 2 (3): 18–22 (2002)
- Breiman and Cutler. Random Forests. Acessado em <https://www.stat.berkeley.edu/~breiman/RandomForests/>