

Atividade 2

1. Monte um passo a passo para o algoritmo RF

1- Criar subsets com Bootstrap;

2- Definir o número de variáveis(colunas) a serem utilizadas nos subsets; as variáveis são selecionadas de forma aleatória;

3- Criar os Base Learners, que diferente do Bagging que são modelos, no RF são árvores de decisão;

4- Comparar os resultados com um dado real e através da matriz de frequência, e selecionar como resposta final a resposta mais frequente.

2. Explique com suas palavras o Randomforest

É uma técnica onde se utilizam os dados para a formação de subsets com variáveis aleatórias para a criação de árvores de decisão, que são geradas de forma aleatória e gerenciadas através de hiperparâmetros. Uma de suas vantagens é que sua aleatoriedade a torna robusta a overfitting devido a sua menor correlação na geração das amostras.

3.Qual a diferença entre Bagging e Random Forest?

O RF utiliza o bootstrap usado no Bagging, com um hiperparâmetro: o `max_features`, onde designamos o número de variáveis que queremos utilizar em cada subset para testagem, e são criados os subsets a serem utilizados com as variáveis sendo definidas em cada subset de forma aleatória, mas sempre com a quantidade definida no `max_features`;

Diferente do Bagging, onde os Base Learners são modelos, no RF eles são árvores de decisão.