

EBAC

Aluna: Cristina Freitas

Módulo 23 – Bagging

Exercício 1: Monte um passo a passo de Bagging e o explique com suas palavras.

Passo a passo:

- 1) É gerado a partir de um dataframe alguns outros com o mesmo número de linhas com reposição;
- 2) A partir de cada conjunto de dados gerados é aplicado um modelo de classificação;
- 3) Com o resultado dos modelos se faz a agregação;

O **Bagging** é uma técnica em Ciência de Dados para determinar a variável alvo que pode ser fruto de uma árvore binária – como 0 ou 1 ou fruto de uma regressão logística – valores contínuos com o preço de alguma coisa.

A técnica consiste em pegar um conjunto de dados e gerar aleatoriamente outros com o mesmo número de linhas com a possibilidade de reposição. Neste contexto, algumas linhas podem ser repetidas. Esta fase é chamada de **Bootstrap**.

Cada conjunto de dados é submetido a um **modelo de classificação** que pode ser uma árvore binária, regressão logística, regressão linear etc. Cada um desses conjuntos submetidos aos modelos de classificação é chamado **base learner**.

Com o resultado de cada um dos modelos é feita uma agregação. No caso de valores binários a agregação tem como resultado a maior frequência. No caso de resultados contínuos a agregação tem como resultado a média. Esta fase da técnica **Bagging** é conhecida como **Aggregating**.