

Bootstrap: O que é

É um sorteio de linhas de um dataset com reposição. Ou seja, uma linha de um dataset é sorteada aleatoriamente para a formação de um novo dataframe, e esta linha sorteada estará disponível para ser sorteada novamente aleatoriamente.

Demonstração

	num_legs	num_wings	spec_seen
falcon	2	2	10
dog	4	0	2
spider	8	0	1
fish	0	0	8

Com o dataframe acima, sortearmos as linhas aleatoriamente, até preencher um novo dataframe Com o mesmo número de linhas, formando como exemplo o dataframe abaixo:

	num_legs	num_wings	spec_seen
falcon	2	2	10
dog	4	0	2
falcon	2	2	10
dog	4	0	2

- Podemos repetir este processo para criar quantos dataframes forem necessários para realizarmos o ‘Base Learners’.
- Base Learners são modelos que vão utilizar os dataframes criados com o Bootstrap para chegar a uma resposta, seja ela de classificação ou regressão. Por exemplo, se quisermos identificar qual dos animais acima é uma ave:
 - O modelo 1 classificou o ‘falcon’ como ‘sim’;
 - O modelo 2 identificou ‘falcon’ como ‘não’;
 - O modelo 3 identificou ‘falcon’ como ‘sim’:

Tomaremos como verdadeira a resposta apontada pelo maior número de modelos. No exemplo acima, dois modelos apontaram ‘falcon’ como ‘sim’ e um como ‘não’; Então consideraremos ‘falcon’ como ‘sim’. Esta técnica é chamada de Bagging – abreviação de ‘Bootstrap Aggregating’.

Nota: E se for um modelo de regressão?

No caso da regressão, tiraremos a média das respostas apontadas pelos 3 modelos e a consideraremos como nossa resposta.