

Aprofundando nas técnicas de combinação

Vimos nessa última Aula o que podemos chamar de técnicas de `ensemble`. Do inglês, `ensemble` significa misturar e, no caso, a ideia é justamente essa: combinar uma série de modelos não tão fortes com a intenção de, justamente, desenvolver um modelo final poderoso! Conforme vimos, árvores de decisão tendem a sofrer de `overfitting` com uma certa facilidade e, junto a isso, os principais erros que encontramos ao treinarmos modelos pode ocorrer por conta da falta de dados, por conta da variância ou do viés causado pelo modelo. Os modelos de `ensemble` partem de um modelo `base`. No `sklearn`, o padrão é justamente usar árvores de Regressão.

Com base no que vimos em sala, quais são as alternativas que mais se encaixam no que aprendemos?

Selecione 3 alternativas

A

As técnicas que vimos foram `Boosting` e `Bagging`. Ambos os métodos "criam" diversas amostras dos dados fazendo uma escolha aleatória do dataset original, com reposição. Logo, alguns dados podem ser repetidos em diferentes amostras.
No caso do `Bagging`, qualquer elemento tem chance de aparecer nas outras amostragens, enquanto no `Boosting`, há uma ponderação entre alguns dados.

B

AdaBoost é um exemplo de algoritmo de `Boosting`. O método "cria" diversas amostras dos dados, realizando escolhas aleatórias do dataset original, com reposição. Sendo assim, alguns dados podem ser repetidos em diferentes amostras.
O `Boosting` leva em consideração a premissa de que 'classificadores fracos combinados podem construir um modelo forte'. Sendo assim, o processo de `Boosting` é feito de forma iterativa, havendo uma ponderação de alguns dados com base nos erros.

C

`RandomForest` é um exemplo de algoritmo de `Boosting`. O método "cria" diversas amostras dos dados, realizando escolhas aleatórias do dataset original, com reposição. Sendo assim, alguns dados podem ser repetidos em diferentes amostras.
O `Boosting` leva em consideração a premissa de que 'classificadores fracos combinados podem construir um modelo forte'. Sendo assim, o processo de `Boosting` é feito de forma iterativa, havendo uma ponderação de alguns dados com base nos erros.

D

AdaBoost é um exemplo de algoritmo de `Bagging`. O método "cria" diversas amostras dos dados, realizando escolhas aleatórias do dataset original, com reposição. Sendo assim, alguns dados podem ser repetidos em diferentes amostras.
Os algoritmos de `Bagging` (também conhecidos pelo nome de `Bootstrap aggregating`) aplicam modelos de aprendizado em cima dessas diferentes 'subamostras' dos dados, gerando diferentes modelos de forma que a média simples destes resulta no nosso modelo final.

E

`RandomForest` é um exemplo de algoritmo de `Bagging`. O método "cria" diversas amostras dos dados, realizando escolhas aleatórias do dataset original, com reposição. Sendo assim, alguns dados podem ser repetidos em diferentes amostras.
Os algoritmos de `Bagging` (também conhecidos pelo nome de `Bootstrap aggregating`) aplicam modelos de aprendizado em cima dessas diferentes 'subamostras' dos dados, gerando diferentes modelos de forma que a média simples destes resulta no nosso modelo final.