escola britânica de artes criativas & tecnologia

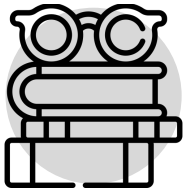
Profissão: Cientista de Dados

escola
britânica de
artes criativas
& tecnologia

GLOSSÁRIO

Combinação de modelos II

- Conheça o Boosting
- Conheça o AdaBoost
- Aplique Gradient Boosting Machine - GBM
- Conheça Stochastic Gradient Boosting Machine
- Utilize eXtreme Gradient Boosting - XGBoost
- Realize Boosting no Python



Dica: para encontrar rapidamente a palavra que procura aperte o comando CTRL+F e digite o termo que deseja achar.

Conheça o Boosting

Conheça o Boosting

Boosting

Técnica de aprendizado de máquina que combina vários modelos fracos para criar um modelo forte. Foi originalmente projetado para problemas de classificação, mas pode ser estendido para regressão também.

Weak Learner (Aprendiz Fraco)

Uma árvore com um nó de profundidade e duas folhas. O Boosting utiliza vários desses aprendizes fracos para criar um modelo forte.

Conheça o AdaBoost

Conheça o AdaBoost

AdaBoost

É um algoritmo de aprendizado de máquina que combina vários modelos fracos para criar um modelo forte. Ele constrói 'tocos' de árvores em vez de árvores completas, cada árvore é influenciada pela anterior e as árvores têm pesos diferentes.

Peso dos dados

No AdaBoost, cada linha de dados recebe um peso inicial. A performance de cada toco é calculada e o toco com a maior performance é selecionado, atualizando os pesos dos dados com base em suas previsões.

Conheça o AdaBoost

Tocos de árvores

No contexto do AdaBoost, são modelos simples de árvore de decisão que são usados para compor o modelo final. Cada toco é criado para cada variável explicativa.

Votação ponderada

É o método pelo qual o AdaBoost faz previsões. Ele soma as performances ponderadas de cada modelo para cada classe possível, selecionando a classe com a maior soma.

Aplique Gradient Boosting Machine - GBM

Aplique Gradient Boosting Machine - GBM

- **Floresta de árvores**
  É uma técnica de aprendizado de máquina que combina várias árvores de decisão para resolver um problema específico.

- **Gradient Boosting Machine (GBM)**
  É uma técnica de aprendizado de máquina para problemas de regressão e classificação, que produz um modelo de previsão na forma de um conjunto de modelos de previsão fracos, geralmente árvores de decisão.

Aplique Gradient Boosting Machine - GBM

Resíduos

Em estatística e otimização, os resíduos de um modelo de regressão são a diferença entre os valores observados do resultado a ser previsto e os valores previstos pelo modelo de regressão.

Conheça Stochastic Gradient Boosting Machine

Conheça Stochastic Gradient Boosting Machine

Função de perda

É uma função que mede o quão bem um modelo de aprendizado de máquina está fazendo seu trabalho. Ela é usada para otimizar o modelo durante o treinamento.

Subamostra

É uma parte menor de um conjunto de dados maior. No contexto do GBM, uma subamostra é selecionada aleatoriamente e sem reposição do conjunto de dados de treinamento.

Robustez

É a capacidade de um modelo de aprendizado de máquina de produzir resultados consistentes, mesmo quando os dados de entrada têm ruído ou outliers.

Utilize eXtreme Gradient Boosting – XGBoost

Utilize eXtreme Gradient Boosting - XGBoost

- **Computação paralela e distribuída**
  É uma forma de computação em que muitos cálculos são realizados simultaneamente. Os cálculos podem ser distribuídos em vários núcleos de um único computador ou em vários computadores em uma rede.

- **Procedimento de quartil ponderado**
  É um método para calcular quartis que leva em consideração a distribuição dos dados. Ele é usado para lidar com dados esparsos no XGBoost.

- **XGBoost (eXtreme Gradient Boosting)**
  É um algoritmo de aprendizado de máquina baseado em árvore que usa o princípio do boosting. Ele é conhecido por sua velocidade e eficiência, sendo capaz de rodar mais rápido e ter uma performance muito boa em classificação e regressão.

Realize Boosting no Python

Realize Boosting no Python

- **Critério de parada**
  É uma condição que determina quando o algoritmo de Boosting deve parar de adicionar novos modelos à série.

- **Métrica ROC AUC**
  É uma métrica de desempenho para modelos de classificação binária. Ela mede a capacidade do modelo de distinguir entre as classes positiva e negativa.

Realize Boosting no Python

Validação fora do tempo

É uma técnica de validação onde o conjunto de validação é um período mais recente do conjunto de dados. Isso é feito para garantir que o modelo seja capaz de generalizar para dados futuros.

Superestimação

É um problema que ocorre quando um modelo de aprendizado de máquina se ajusta demais aos dados de treinamento e tem um desempenho ruim nos dados de teste.

escola britânica de artes criativas & tecnologia

Bons estudos!