

Prática: Predição e a Base de Aprendizado de Máquina (II)

Eduardo Prasniewski

1 Seção 3

Regressão linear trata-se de traçar uma linha entre os pontos do conjunto, sendo assim com a função determinada é possível gerar novos valores ainda não inseridos. Foi um dos conteúdos estudados também em Cálculo Numérico, utilizando a função de erro dos mínimos quadrados, porém pode ser usado também gradiente descendente. Regressão polinomial tem o mesmo propósito que a regressão, mas agora para dados que apresentam um comportamento não-linear. Assim como a múltipla regressão, que agora trabalha com muitas dimensões.

2 Seção 4

Machine Learning são algoritmos que podem aprender com dados e fazer predições, há dois tipos: aprendizado supervisionado (20% dos dados para teste e 80% para treino) e não-supervisionado. Desse modo foi passado uma prática para analisar se uma regressão polinomial não estava fazendo um overfitting.

Em segundo momento o professor explica como o teorema de Bayes se aplica a desvendar se um email é um spam ou não, logo a seguir passa uma prática do mesmo assunto.

K-means clustering trata de separar os dados a partir de centroides, uma técnica não-supervisionada, que agrupa os dados em K buckets.

Árvores de decisão tratam-se de decisões dicotômicas em que cada passo buscam reduzir a entropia do sistema, a fim de ficar apenas com uma decisão. E assim foi passado um exemplo prático com dados de contratação.

Ensemble learning é a tática de combinar vários modelos juntos para conseguir um resultado melhor que um único modelo. E um dos modelos mais usados em Big Data e machine learning é o XGBoost, o qual o professor passou um pequeno exemplo.

Support Vector Machine Classification é um método usado para classificar dados com muitas dimensões.