Machine Learning – Anotações do Curso de Python Fundamentos para Data Science

Scikit-Learn é uma biblioteca do Python que fornece ferramentas simples e eficientes paraData Mining e Análise de Dados, fornecendo modelos, é usado junto com o NumPy (Computação Matemática), SciPy e Matplotlib (Visualização de Dados) e é usado para criar modelos e não para leitura, manipulação e sumarização de dados, isso é papel do NumPy, Pandas (Manipulação de dados)

Tenta prever propriedades em dados desconhecidos dentro de amostra de dados.

# Tipos de Aprendizado

## Supervisionada

É o termo usado sempre que o programa é “treinado” sobre um conjunto de dados pré-definido. Baseado no treinamento com os dados pré-definidos, o programa pode tomar decisões precisas quando recebe novos dados. Exemplo: Pode-se usar um conjunto de dados de recursos humanos para treinamento da Machine Learning, que tenha tweets marcados como positivos, negativos e neutros e assim treinar um classificador de análise de sentimento.

Exemplo pode ser o reCaptcha do Google.

**Classificação –** É o processo de tomar algum tipo de entrada e atribuir um rótulo a ela. Sistemas de classificação são usados geralmente quando as previsões são de natureza distinta, ou seja um simples “Sim” ou “Não”.

Exemplo: Mapeamento de uma imagem de uma pessoa e classificação como masculino e feminino.

**Regressão –** Usada quando o valor que está sendo previsto difere de um “Sim” ou “Não” e que siga um espectro contínuo. Sistemas de regressão poderiam ser usados, por exemplo, para responder às perguntas: “Quanto Custa?” ou “Quantos Existem?”.

## Não Supervisionada

É o termo usado quando um programa pode automaticamente encontrar padrões e relações em um conjunto de dados.

Exemplo: Análise de um conjunto de dados de e-mails e agrupamento automático de e-mails relacionados ao tema, sem que o programa possuía qualquer conhecimento prévio sobre os dados.