Trabalho prático 1 - Execução dos códigos

Julianny Favinha Donda 156059 julianny.favinha@gmail.com

José Henrique Ferreira Pinto 155976 joseh.fp@gmail.com

É necessário que a biblioteca numpy esteja instalada.

TrainLinearRegression.py

Este arquivo contém a implementação da descida do gradiente em modo batch. Para executá-lo, execute no terminal

\$ python3 TrainLinearRegression.py

O programa gera o arquivo "BGDModel", que contém os parâmetros gerados pela execução do último número de iterações usando a descida do gradiente em modo batch. Imprime na saída padrão, para cada número de iterações, os parâmetros encontrados, o custo (mean squared error) e o tempo de execução decorrido.

NormalEquation.py

Este arquivo contém a implementação da equação normal. Para executá-lo, execute no terminal

\$ python3 NormalEquation.py

O programa gera o arquivo "NormalEquationModel", que contém os parâmetros gerados pela execução. Imprime na saída padrão os parâmetros encontrados e o tempo de execução decorrido.

SKLearnSGDRegressor.py

Este arquivo contém a implementação da descida do gradiente em modo estocástico, utilizando a biblioteca Scikit Learn. Para executá-lo, execute no terminal

\$ python3 SKLearnSGDRegressor.py

O programa gera o arquivo "SGDRegressorModel", que contém os parâmetros gerados pela execução. Imprime na saída padrão os parâmetros encontrados, o custo (mean squared error) e o tempo de execução decorrido.

TestLinearRegression.py

Este arquivo contém a implementação para testar um modelo no conjunto de teste. Para executá-lo, execute no terminal

\$ python3 SKLearnSGDRegressor.py

Como entrada, deverá ser informado o nome do arquivo de coeficientes gerado anteriormente pelos programas descritos. A saída do programa é o custo (mean squared error).

Cost.py e ParseDiamondSet.py

Arquivos auxiliares. Arquivo com funções para cálculo do mean squared error e arquivo para parsear a base de dados vinda do Kaggle em conjuntos de treino, validação e teste, respectivamente.