

Instituto Federal de Educação, Ciência e Tecnologia – IFB  
Campus Taguatinga  
Curso Tecnológico em Automação Industrial  
Disciplina: Aprendizagem de Máquina  
professor Lucas Moreira

## Trabalho 01

O objetivo do primeiro trabalho é ambientar o(a) aluno(a) com a programação em Matlab/Octave, com a Álgebra Linear e o desenvolvimento de algoritmos numéricos.

Nesse trabalho, o(a) aluno(a) irá implementar um programa computacional de regressão linear de múltiplas variáveis, utilizando o algoritmo do gradiente descendente.

### Dados

Foram coletados dados do mercado imobiliário acerca do preço de 47 imóveis e suas principais características. Dentre elas, serão usadas somente a área do imóvel e o número de quartos.

Esses dados estão disponíveis no arquivo “data.txt”. Eles estão compilados em uma matriz 47 x 3, onde a primeira coluna corresponde à área do imóvel (em pés quadrados “*squared feet*”), a segunda coluna corresponde ao número de quartos e a terceira o valor desse imóvel, avaliado por um corretor experiente. Cada linha da tabela corresponde a um imóvel.

Esses, e muitos outros dados de exemplo, estão disponíveis no site The Data and Story Library (<http://lib.stat.cmu.edu/DASL>).

### Problema

Deseja-se construir um modelo matemático que consiga estimar automaticamente o valor de imóveis a partir de suas características. Assim, o(a) aluno(a) deve usar a regressão linear para calibrar esse modelo, utilizando os dados de treinamento do arquivo “data.txt”.

O trabalho deve ser constituído por 3 (três) arquivos Matlab/Octave (com extensão .m) distintos.

São eles:

- Um script principal, que deverá ser executado para realizar a regressão linear;
- Um script de função para cálculo da função custo;
- Um script de função para realizar o algoritmo do gradiente descendente.

## Dicas

O primeiro script, pode ter qualquer nome permitido pelo padrão Matlab/Octave, mas deve necessariamente conter a extensão `.m`. É sugerido para as primeiras linhas de comando desse script, os comandos `clear` e `clc`. O primeiro apaga todas as variáveis do Ambiente de Trabalho, evitando bugs devido ao uso anterior de outros scripts. O segundo, limpa a tela de comando, facilitando a visualização das saídas do script executado pela última vez.

A função custo deve receber como parâmetros a matriz de dados de entrada  $\mathbf{X}$ , o vetor de valores alvo  $\mathbf{y}$  (lembrando que a regressão linear é um algoritmo de treinamento supervisionado) e o vetor de parâmetros  $\boldsymbol{\theta}$ . Para facilitar, e agilizar, o cálculo da função custo, é sugerido o uso de operações algébricas no lugar de loops de repetição, como o `for`. Por isso, a matriz  $\mathbf{X}$  deve conter uma coluna adicional na primeira posição, formada somente pelo valor 1. O nome do arquivo deve ser o mesmo nome da função, acrescida da extensão `.m`.

A função do gradiente descendente, deve receber como parâmetros a matriz de dados de entrada  $\mathbf{X}$ , o vetor de valores alvo  $\mathbf{y}$ , o vetor de valores iniciais dos parâmetros  $\boldsymbol{\theta}$ , o valor da taxa de aprendizagem  $\alpha$  e o número total de iterações. Não é necessário criar critérios de convergência para o algoritmo. Essa função deve retornar, obrigatoriamente, o vetor de parâmetros  $\boldsymbol{\theta}$  com os valores ótimos, ajustados ao conjunto de dados de treinamento. Sugere-se, também, retornar um vetor de valores da função custo para cada iteração.

## Entrega

Os 3 (três) arquivos citados acima, devem ser compactados em um único arquivo `.zip` e enviados para o endereço de e-mail [lucas.moreira@ifb.edu.br](mailto:lucas.moreira@ifb.edu.br) até às 23h55min do dia 09 de setembro de 2018.

O trabalho é individual, e deve ser entregue um arquivo compactado, contendo os 3 (três) scripts solicitados, por aluno(a).

Cópias de trabalhos receberão nota 0 (zero).

Em caso de dúvidas, pede-se que entrem em contato pelo endereço de e-mail acima.