

# Exercício de programação: Regressão Linear

## 1. Regressão linear com uma variável

Neste exercício, você implementará uma regressão linear com uma variável para prever o lucro que poderia ser obtido caso você tenha decidido montar um *food truck*.

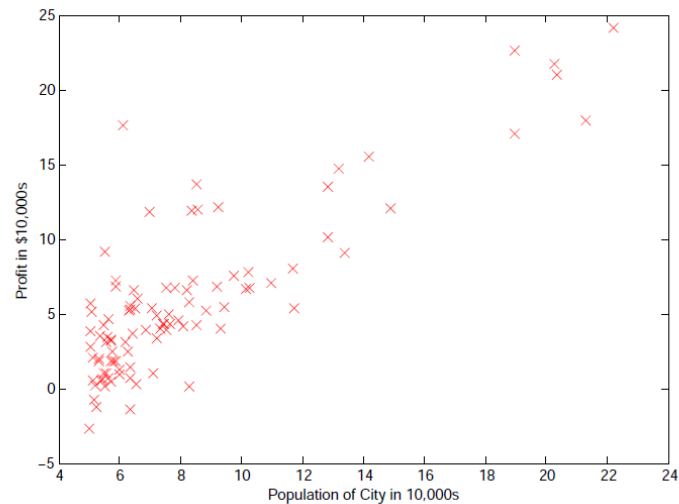
### 1.1 O problema:

Suponha que você seja o CEO de uma franquia de restaurantes e esteja procurando por um lugar (cidade) para abrir uma nova loja (*food truck*). A rede já tem caminhões em várias cidades e você tem os dados sobre lucro e tamanho da população em cada uma delas. Você gostaria então de usar esses dados para ajudá-lo a selecionar qual deveria ser a próxima cidade a ser escolhida dentro do seu projeto de expansão da marca.

O arquivo “data1.txt” contém o conjunto de dados para o esse problema de regressão linear. A primeira coluna deste arquivo representa a população de uma cidade e a segunda coluna, por sua vez, o lucro de um *food truck* naquela cidade. Um valor negativo para o lucro indica uma perda.

### 1.2 Plotando os dados

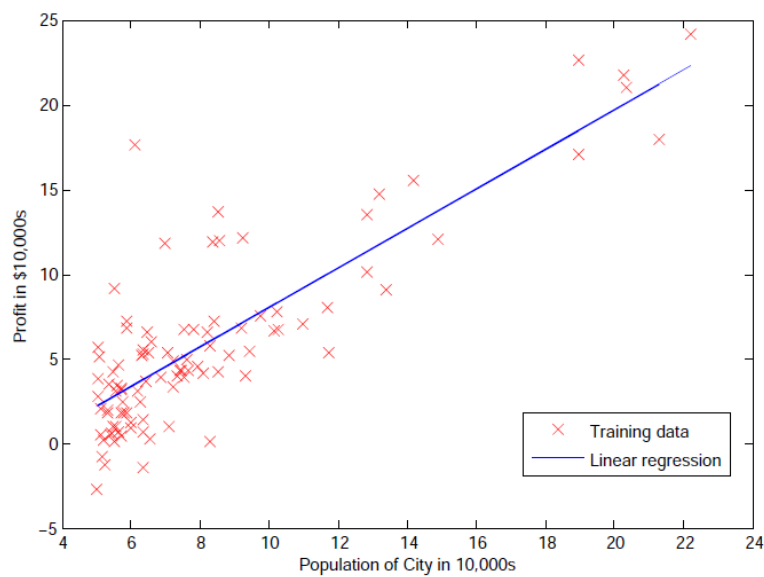
Antes de iniciar qualquer tarefa, muitas vezes é útil entender os dados visualizando-os. Para este conjunto de dados, você deve plotar um gráfico para visualizá-los, uma vez que ele possui apenas duas propriedades para traçar (lucro e população). (Muitos outros problemas que você encontrará na realidade são multidimensionais e não podem ser plotados em 2D.)



### 1.3 Gradiente descendente

Nesta parte, você deverá calcular os parâmetros da regressão linear e a função custo para o conjunto de dados, utilizando uma taxa de aprendizado de 0,01.

- Plote a função custo em relação ao número de iterações para ver o seu decaimento;
- Depois que terminar de calcular os parâmetros da regressão, traçar o ajuste linear, conforme a figura a seguir.



## 2. Regressão linear com múltiplas variáveis

Nesta parte, você implementará a regressão linear com múltiplas variáveis para prever o preço de uma casa. Suponha que você está vendendo sua casa e você queira saber qual seria um bom preço de acordo com o mercado. Dessa forma, uma maneira de fazer isso seria coletar informações sobre as casas que foram vendidas recentemente e fazer um modelo de previsão de preços.

O arquivo “data2.txt” contém um conjunto de treinamento de preços de casas em Portland, Oregon. A primeira coluna representa o tamanho da casa (em metros quadrados), a segunda o número de quartos e terceira o preço da casa.

### 2.1 *Feature Normalization*

Comece carregando e exibindo alguns valores a partir deste conjunto de dados. Olhando para os valores, note que a dimensão relativa ao tamanho das casas é cerca de 1000 vezes maior que o número de quartos. Quando as features diferem por uma elevada ordem de magnitude, reescalonar a dimensão das features pode fazer a descida do gradiente convergir muito mais rapidamente.

- Subtraia o valor médio de cada feature do conjunto de dados.
- Após subtrair a média, divida os valores das features pelos seus respectivos desvios-padrão

### 2.2 Gradiente descendente

Nesta parte, você deverá calcular os parâmetros da regressão linear e a função custo para o conjunto de dados, utilizando diferentes taxas de aprendizado para estudar o seu efeito na convergência.

- Plote a função custo em relação ao número de iterações para ver o seu decaimento;

**2.3 Veja que agora não é possível traçar o ajuste linear como no exercício anterior. Por quê?**

### 3. Equação Normal

Você aprendeu nos vídeos que a solução matemática para a regressão linear pode ser dada pela equação normal:

$$\theta = (X^T X)^{-1} X^T \vec{y}.$$

É importante disser que, o uso desta fórmula, não requer nenhum redimensionamento das *features*. Além disso, você obterá uma solução exata (não há *loops* até a convergência, como no gradiente descendente).

Calcule os parâmetros da regressão linear utilizando a equação normal e compare o resultado com os parâmetros obtidos por meio do método do gradiente descendente.

**Sugestão:** você poderia utilizar a base de dados para prever o valor das casas utilizando os parâmetros da regressão encontrados com os dois métodos (gradiente e equação normal). Após esta etapa, basta calcular o erro quadrático médio das previsões em relação aos valores reais para cada um dos métodos e fazer a comparação.