

**CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DE MINAS GERAIS
DEPARTAMENTO DE COMPUTAÇÃO
INTELIGÊNCIA COMPUTACIONAL**

REGRESSÃO LINEAR

CRISTIANO NEIVA ABRANTES E CHRISTIAN MEDEIROS

**BELO HORIZONTE
NOVEMBRO DE 2021**

1 Introdução

Implementação da tarefa de Regressão Linear com Uma e Múltiplas Variáveis, os fontes (Algoritmos em Python) com as implementações estarão nos arquivos em anexo, juntamente às bases de dados.

2 Objetivo

O objetivo é reforçar o aprendizado da parte teorica sobre Regressão Linear com uma ou Múltiplas Variáveis.

3 Regressão linear com uma variável

Implementação de uma regressão linear com uma variável para prever o lucro que poderia ser obtido caso se tenha decidido montar um food truck, a partir de uma base de dados.

3.1 Dados

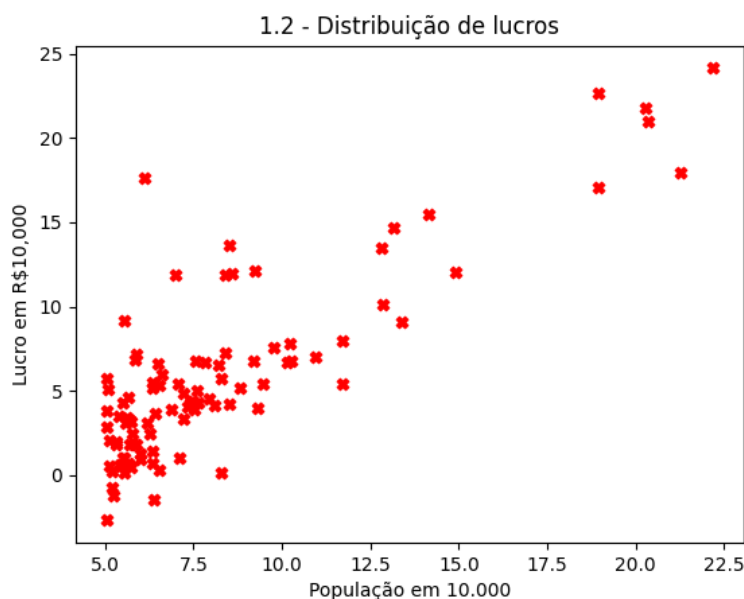


Figura 1: Exibição dos dados

Observando a distribuição dos dados vemos que temos uma concentração de dados onde há os menores valores de lucro e população.

3.2 Gradiente descendente

Calculo dos parâmetros da regressão linear e a função custo para o conjunto de dados, utilizando uma taxa de aprendizado de 0,01

3.2.1 Custo por número de iterações

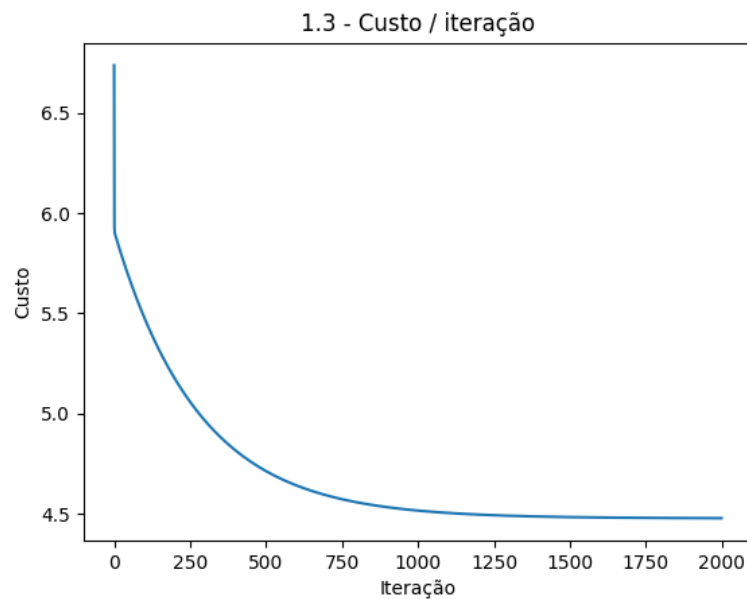


Figura 2: Custo por Iteração

Observando os resultados vemos que o desempenho é bom mesmo quando se cresce o número de iterações e é menos eficiente para um número baixo de iterações

3.2.2 Ajuste Linear

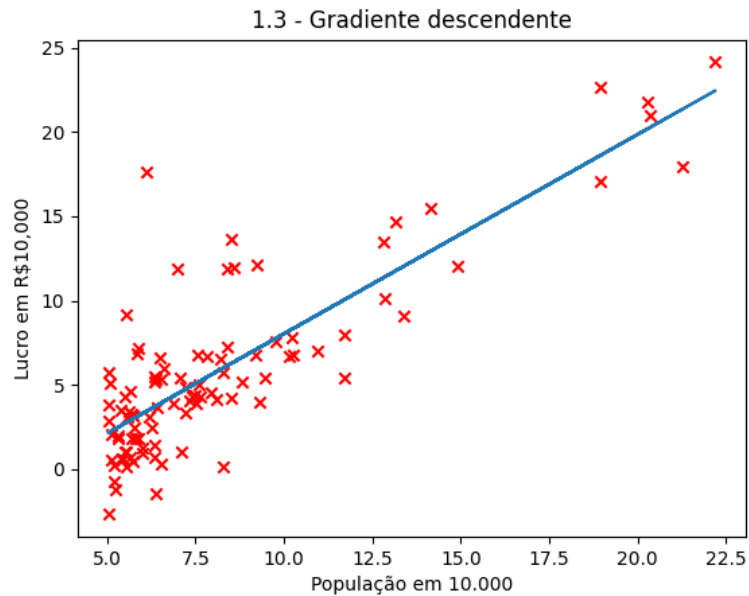


Figura 3: Ajuste Linear

Observando os resultados do ajuste linear obtido vemos que os dados da base se aglutinam próximos à linha da regressão linear em azul, observando este resultados concluímos que a regressão linear, neste caso se aproximou bem à modelagem deste problema, algumas instâncias distoantes das demais ficaram mais afastadas do ajuste.

4 Regressão linear com múltiplas variáveis

Implementação de uma regressão linear com múltiplas variáveis para prever o preço de uma casa, a partir de uma base de dados, e fazer um modelo de previsão de preços.

4.1 Dados

O arquivo em anexo “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.

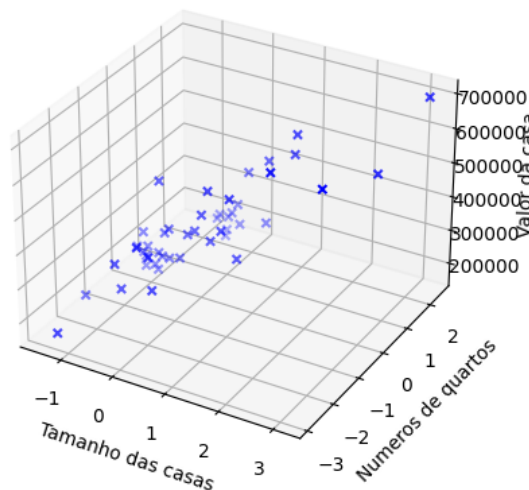


Figura 4: Exibição dos dados

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.

4.1.1 Normalização dos dados

Subtraindo o valor médio de cada feature do conjunto de dados e dividindo os valores das features pelos seus respectivos desvios-padrão

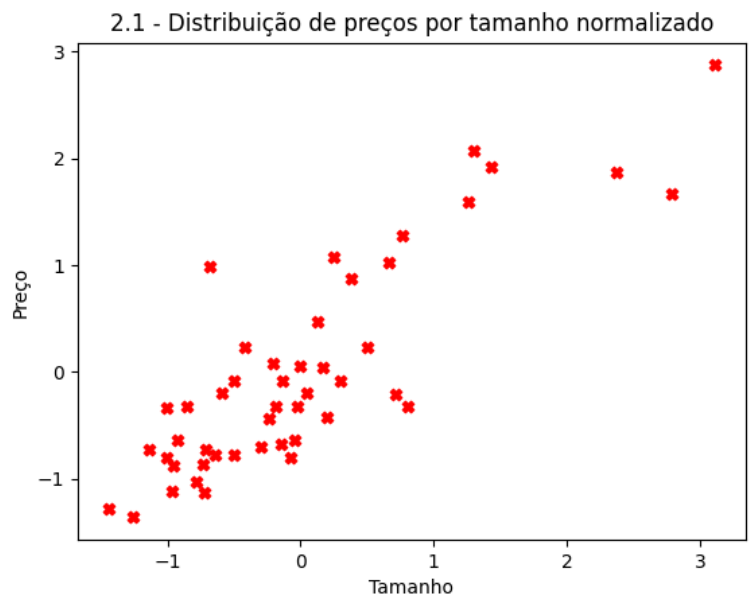


Figura 5: Distribuição de preços por tamanho normalizado

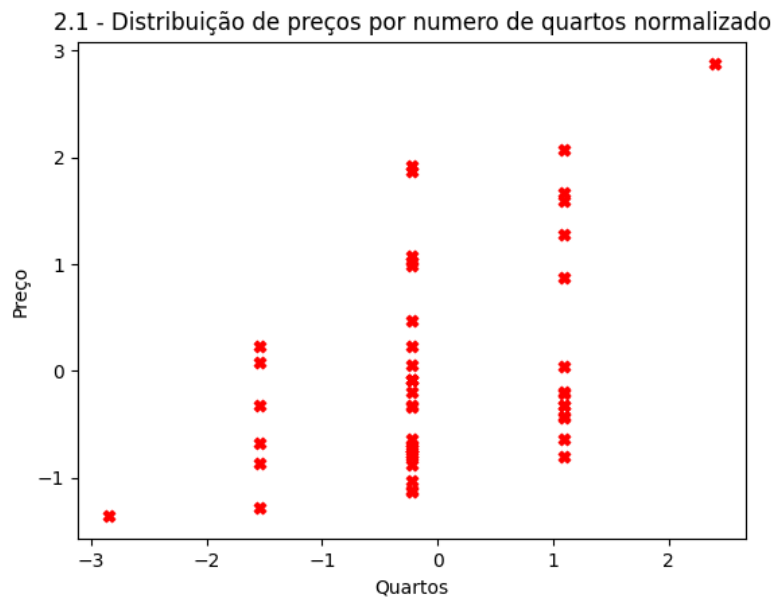


Figura 6: Distribuição de preços por numero de quartos normalizado

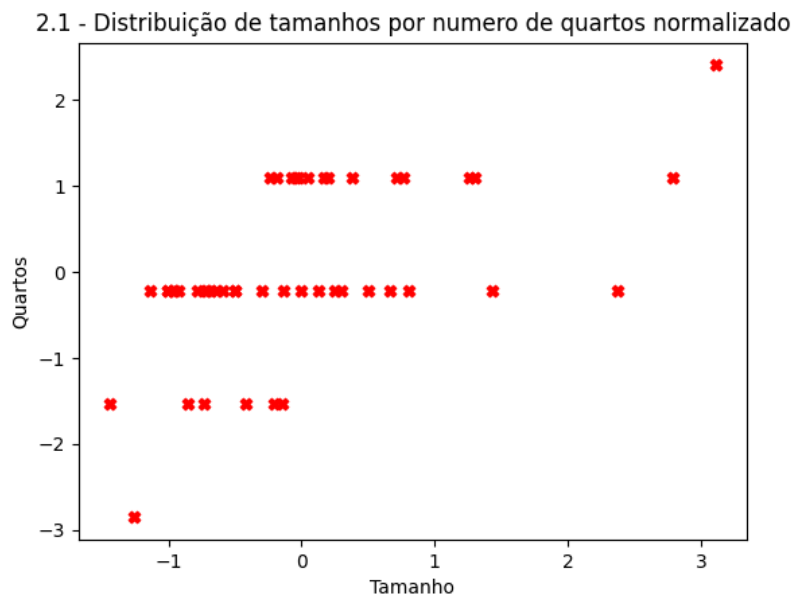


Figura 7: Distribuição de tamanhos por numero de quartos normalizado

4.2 Gradiente descendente

Calculo dos parâmetros da regressão linear e a função custo para o conjunto de dados, utilizando uma taxa de aprendizado de 0,001

4.2.1 Custo por número de iterações

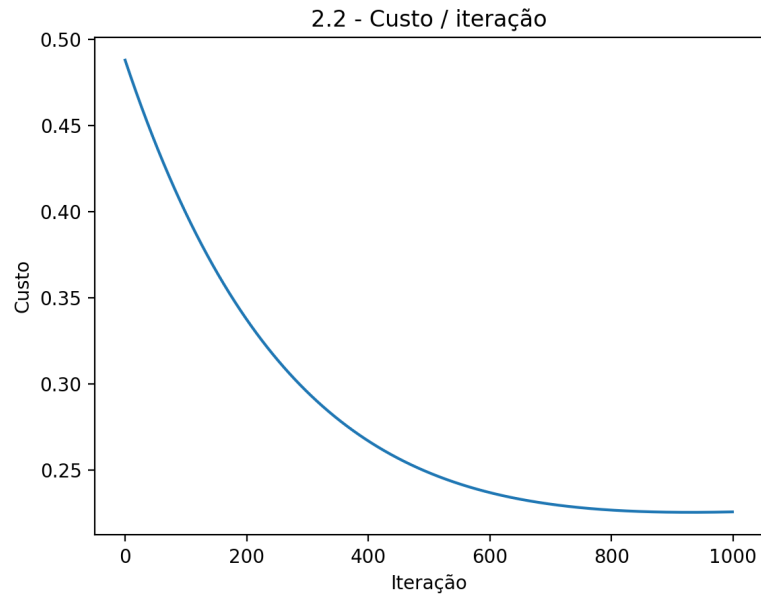


Figura 8: Custo por Iteração

Diferentemente da sessão anterior não podemos traçar a reta da equação, visto que agora não predizemos o valor a partir de três parâmetros, sendo dois linearmente dependentes dos parâmetros.

4.3 Equação normal

Implementação de uma regressão linear com múltiplas variáveis da mesma base de dados a partir da equação normal

$$\theta = (X^T X)^{-1} X^T Y$$

e comparação com os resultados do gradiente descendente.

4.3.1 Comparação do percentual de erro por valor

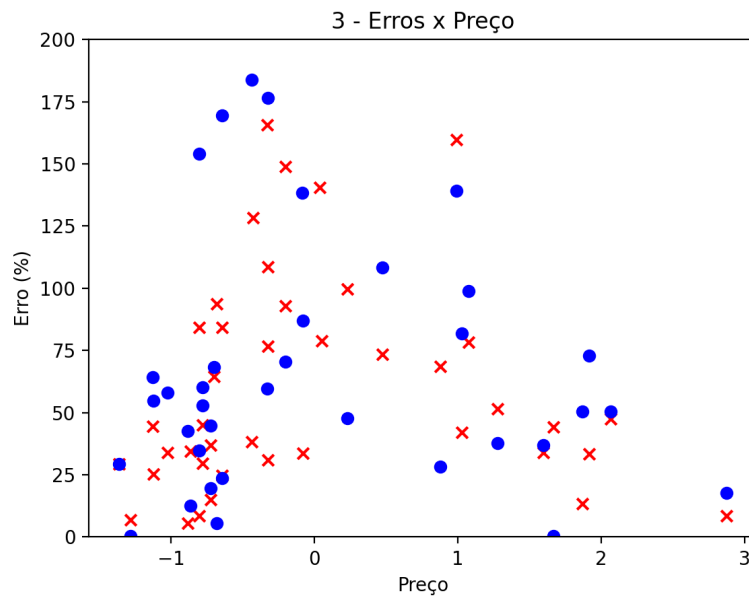


Figura 9: Comparação de erros

4.3.2 Comparação do erro quadrático médio das previsões

Utilizando o método da equação normal tivemos o erro médio de 0.5112 enquanto que ao utilizar o método do gradiente tivemos o erro medio de 0.6717.