Aprendizagem Automática - Lista 1

Regressão Linear

Israel de Castro Vidal – 370019

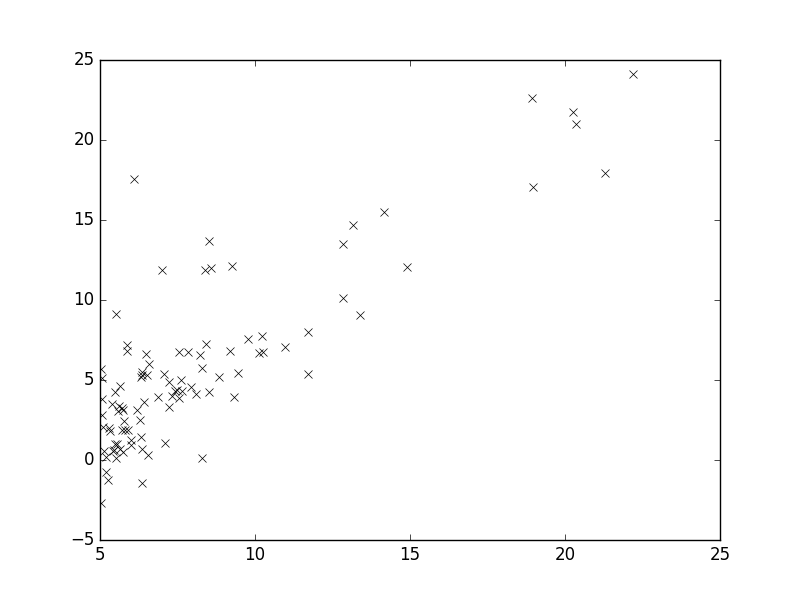
Deverá ser enviado ao professor, um arquivo texto contendo os gráficos, resultados e comentários requeridos em cada item.

1. **Regressão Linear Univariada**

- Carregue os dados contidos no arquivo ex1data1.txt.

O arquivo contem 97 linhas e 2 colunas de dados. Cada coluna se refere a uma variável. Neste problema, deve-se desenvolver um modelo onde a variável da coluna 2 será estimada a partir da variável da coluna 1.

**Apresentar**: Figura com os dados



Dataset1

**Comentários**: Um modelo de regressão linear parece ser adequado para os dados em questão? Comente.

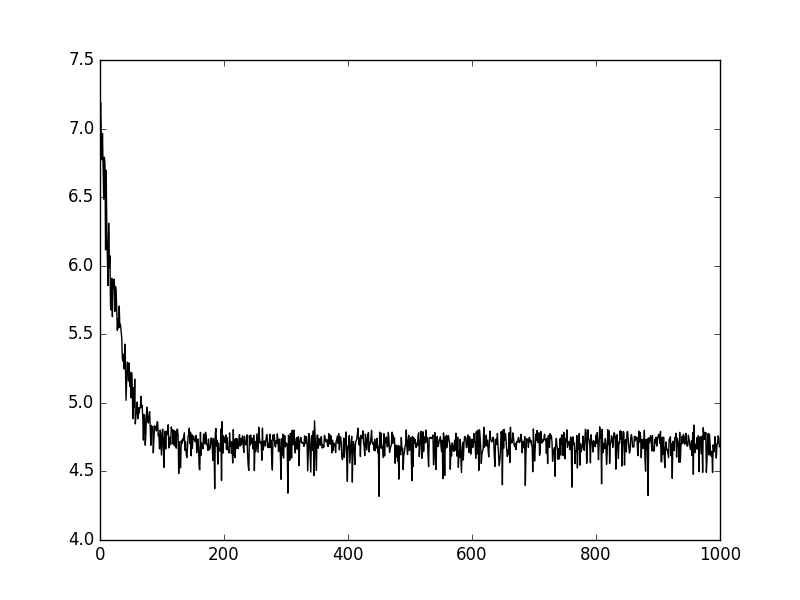
*Sim, um modelo de regressão linear parece adequado, pois, pelo gráfico anterior, parece existir uma reta que aproxima a função original minimizando o erro.*

- Implemente o algoritmo do gradiente descendente estocástico para encontrar os coeficientes da regressão.

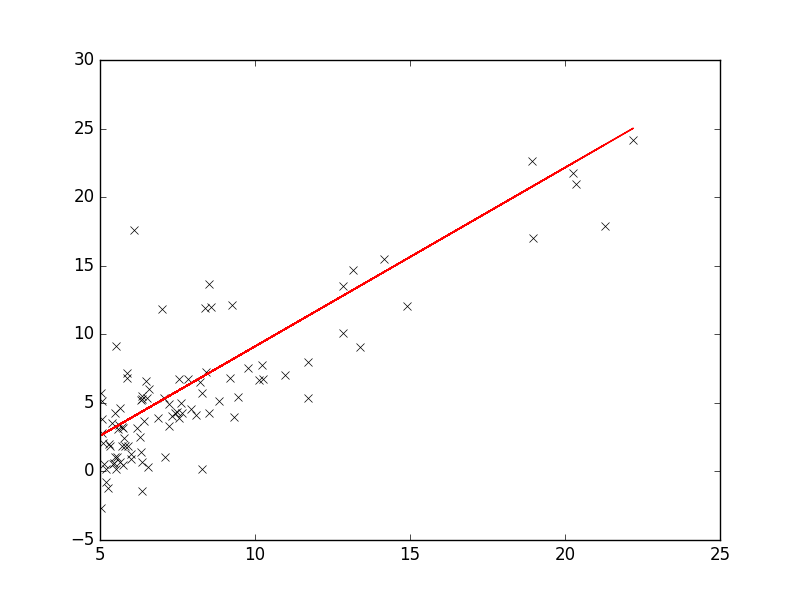
Para este algoritmo utilize α = 0.001 e utilize 1000 épocas de treinamento. Para cada época de treinamento, calcule o erro quadrático médio (EQM) . Plote o gráfico “épocas x EQM”

**Apresentar**: Valor final dos coeficientes e gráfico épocas x EQM.

*W0 = -3.917517058471562, W1 = 1.303534229171897*

**

*EQM x Épocas*



*Dataset1 com a reta gerada pela regressão linear*

**Comentários**: Através do gráfico “épocas x EQM” é possível verificar que o algoritmo está “aprendendo” ? Comente.

*Sim, através do gráfico epoch x mse podemos observar que o algoritmo está aprendendo, pois há uma queda considerável no erro ao longo das épocas. No entanto, também é possível observar que grande parte do aprendizado ocorre antes da época 200, melhorando pouquíssimo após isso.*

1. **Regressão Linear Múltipla**

- Carregue os dados contidos no arquivo ex1data2.txt.

O arquivo contem 47 linhas e 3 colunas de dados. Cada coluna se refere a uma variável. Neste problema, deve-se desenvolver um modelo onde a variável da coluna 3 será estimada a partir das variáveis das coluna 1 e 2. Os dados apresentados referem-se a um problema de estimação do preço de casas. As variáveis 1 e 2 são a área da casa e o número de quartos, respectivamente. A variável 3 é o preço do imóvel.

- Implemente o algoritmo do gradiente descendente estocástico para encontrar os coeficientes da regressão.

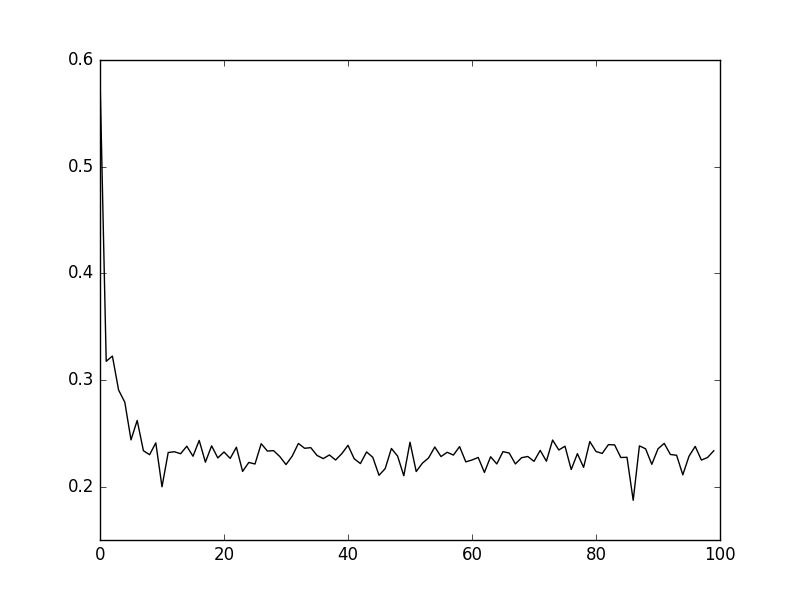
Para este algoritmo utilize α = 0.01 e utilize 100 épocas de treinamento. Para cada época de treinamento, calcule o erro quadrático médio (EQM) . Plote o gráfico “épocas x EQM”

**Apresentar**: Valor final dos coeficientes e gráfico épocas x EQM.

W0 = 0.86476024993741019

W1 = 1.3966730190128032

W2 = -0.10865948958961531



*EQM x Épocas*

**Comentários**: Através do gráfico “épocas x EQM” é possível verificar que o algoritmo está “aprendendo” ? Comente.

*Sim, novamente podemos perceber o aprendizado através do gráfico pela diminuição do erro quadrático médio ao longo das épocas. A oscilação que ocorre no gráfico, isso é, alguns aumentos do erro ao longo do gráfico, ocorre por conta da permutação dos dados a cada época.*

- Encontre os coeficientes da regressão utilizando o método dos mínimos quadrados.

**Apresentar**: Valor final dos coeficientes

**Comentários**: Os valores obtidos pelos dois métodos são iguais? Comente.

1. **Regularização**

- Carregue os dados contidos no arquivo ex1data3.txt.

O arquivo contem 47 linhas e 6 colunas de dados. Cada coluna se refere a uma variável. Neste problema, deve-se desenvolver um modelo onde a variável da coluna 6 será estimada a partir das demais variáveis. Os dados apresentados referem-se a um problema de estimação do preço de casas. As variáveis 1, 2, 3, 4 e 5 são características dos imóveis e serão utilizadas como entrada do problema de regressão. A variável 6 é o preço do imóvel.

- Divida o conjunto de dados entre treino e teste. Para este problema, os primeiros 30 dados serão utilizados para treino e o restante será usado para teste.

- Encontre os coeficientes da regressão utilizando o método dos mínimos quadrados regularizado para os seguintes valores de λ = [0 1 2 3 4 5]. Utilize o conjunto de treinamento.

**Apresentar**: Valores finais dos coeficientes

**Comentários:** Quais variáveis parecem ser menos relevantes para a regressão?

- Encontrar o valor do EQM para os dados de treinamento e de teste para cada um dos valores de λ.

**Apresentar**: Dois gráficos. EQM x λ no conjunto de treinamento e EQM x λ no conjunto de teste

**Comentários**: Como os valores dos coeficientes variam com λ ? Explique o motivo.

Comente o crescimento/decrescimento dos erros presente nas figuras EQM x λ