

## Lista Aprendizado de Máquina - 01

### Recomendações:

- Deverá ser enviado ao professor, um relatório texto contendo os gráficos, resultados e comentários requeridos em cada item.
- Deverá ser enviado ao professor, um notebook em .py com os códigos usados.
- Pode ser feito de prova individual ou em dupla.
- Esse trabalho será usado no trabalho final.
- A entrega será no dia 29/11/2022

### 1. Carregue os dados contidos no arquivo ex1data1.txt.

O arquivo contém 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

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

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

Para este algoritmo utilize  $\alpha = 0.001$  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.

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

### 2. Regressão Linear Múltipla

- Carregue os dados contidos no arquivo ex1data2.txt.

O arquivo contém 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 colunas 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  $\alpha = 0.01$  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.

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

- 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.

### 3. Regularização

- Carregue os dados contidos no arquivo ex1data3.txt.

O arquivo contém 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  $\lambda = [0 \ 1 \ 2 \ 3 \ 4 \ 5]$ . 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  $\lambda$ .

Apresentar: Dois gráficos. EQM x  $\lambda$  no conjunto de treinamento e EQM x  $\lambda$  no conjunto de teste

Comentários: Como os valores dos coeficientes variam com  $\lambda$ ? Explique o motivo.

Comente o crescimento/decrescimento dos erros presente nas figuras EQM x  $\lambda$

Os dados apresentados referem-se a um problema de precificação de casas na cidade de Boston (Boston housing dataset). A estimação do preço de cada casa é feita baseada nos outros atributos apresentados. O significado de cada atributo pode ser encontrado no arquivo "housing.names".

- Divida o conjunto de dados entre treino, validação e teste. Para este problema, utilize 306 dados de cada classe para treino, 100 para validação e 100 para teste.

- Implemente o algoritmo do gradiente descendente para encontrar os coeficientes do classificador

Para a utilização da divisão treinamento, validação e teste, os pesos devem ser obtidos somente com base no conjunto de treinamento. O conjunto de validação indicará o melhor momento para interromper o treinamento e o conjunto de testes dará o desempenho final.

Durante o treinamento, para cada época, verifique o erro no conjunto de treinamento e de validação. Espera-se que o erro de treinamento sempre decresça. O erro do conjunto de validação deve decrescer até um determinado ponto e depois ele irá aumentar, indicando que começa a acontecer overfitting. Utilize este ponto como critério de parada do treinamento.

Apresentar: o gráfico épocas x Erro quadrático médio para os conjuntos de treinamento e validação

Comentários: Comente o gráfico obtido