

Aprendizagem de Máquina e Mineração de Dados

Trabalho 1

Data de entrega: 27/08/2017

1) Existem várias definições de Aprendizagem de Máquina, pesquise uma e indicando o autor da definição. (1 ponto)

“Machine Learning at its most basic is the practice of using algorithms to parse data, learn from it, and then make a determination or prediction about something in the world.” – Nvidia

Aprendizado de máquina é em seu básico é a prática de usar algoritmos para analisar dados, aprender com eles e com isso fazer alguma previsão sobre alguma coisa no mundo..

2) Descreva as principais etapas do processo de aprendizado. (1 ponto)

Aquisição de dados: Adquire os dados, muitas vezes sem classificação ou rotulação.

Extração de informações: Em meio a toda aquela turbilhão de informações é extraído informações que serão úteis ao sistema.

Classificação: em caso de Supervisionado é classificado ou seja “Rotulado” para que assim a máquina tenha um aprendizado consistente.

Treinamento: A etapa do aprendizado em si por meio de algoritmos a máquina vai “aprender” sobre a base de dados com o intuito de prever novos dados

Aplicação: Etapa da execução em si com dados novos ele consegue prever novas respostas sem ser sido previamente apresentados a eles.

3) Explique e dê exemplos de métodos para os seguintes tipos de aprendizado. a)

Aprendizado supervisionado b) Não supervisionado. (1 ponto)

Supervisionado, como a própria palavra diz, tem alguma supervisão, sendo previa ou durante, geralmente os dados são pré classificados para que a máquina com base nesses “rotulos” consiga “prever” dados ainda não mostrados a ela.

EX: o que é feito nas próximas questões, previsão de preços das casas com base em elementos pré definidos.

Não Supervisionado, dados são dados as máquinas sem classificação previa ou rotulação e a máquina tenta ela mesma descobrir a rotulação e classificar por si só.

Ex: Rede neural da google que é capaz de agrupar animais por espécies sem nunca ter visto, consegue descobrir semelhanças entre gatos e cachorros por exemplo.

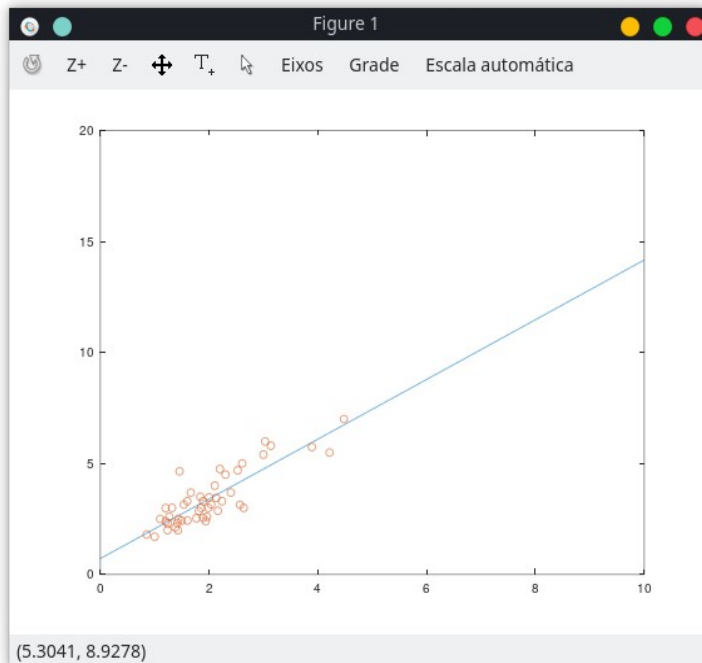
4) Utilizando o Octave, aplique o método dos mínimos quadrados para encontrar a reta de regressão nos dados housing_1.txt (2 colunas de atributos e 1 coluna de rótulo). Mostre o gráfico e a reta de regressão, e os coeficientes obtidos. (2 pontos)

2D:

$b+ax$

$TH = [a \ b]$

$E = \text{Erro Quadratico Medio}$



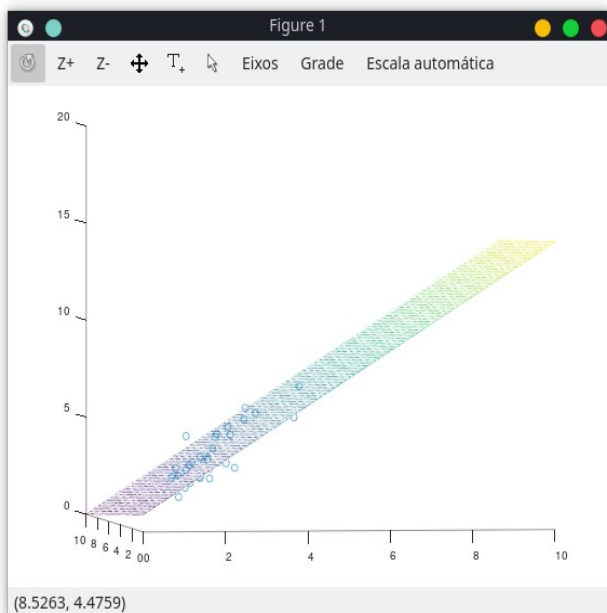
```
th =  
  
    1.34525  
    0.71270  
  
E = 0.41163  
>>  
>> |
```

3D:

$b+ax+cy$

$TH = [a \ c \ b]$

$E = \text{Erro Quadratico Medio}$



```
th =  
  
    1.392107  
   -0.087380  
    0.895979  
  
E = 0.40866  
~~
```

5) Utilizando o octave, aplique novamente o método dos mínimos quadrados para encontrar a reta de regressão porém utilizando a base de dados housing_2.txt (completa). (3 pontos)

a) Escolha duas dimensões (dois atributos da base) de sua preferência e justifique essa escolha, mostre os coeficientes resultantes e o erro quadrático médio.

R: Escolhidos foram : crim dis

Criminalidade e distancia do trabalho , seria uma das coisas que eu olharia para compra e/ou alugar uma casa.

```
th =  
-0.36657  
0.52310  
21.87224  
  
E = 70.654  
.,
```

b) Em seguida escolha uma terceira dimensão de sua preferência, mostre também o erro quadrático médio e os coeficientes resultantes.

R: Escolhido RAD, Pois para viajar é melhor ter uma boa saída as rodovias...

```
th =  
-0.25550  
0.25662  
-0.21452  
24.53077  
  
E = 68.804
```

c) Compare o erro quadrático médio do item a) e do item b), verifique se o erro diminuiu ou aumentou, e justifique sua resposta.

O erro melhorou ,mas foi muito pouco, refletindo cheguei a conclusão que as distancias ao trabalho e o acesso as rodovias provavelmente tem alguma correlação, então resolvi fazer mais algumas.

crim dis rad age

```
th =  
-0.24334  
-0.83455  
-0.17779  
-0.11854  
36.40624  
  
E = 64.022
```

crim dis rad age tax

```
th =  
-0.238703  
-0.994890  
0.385925  
-0.102695  
-0.034573  
44.642090  
  
E = 58.617
```

all

```
th =  
-0.10801136  
0.04642046  
0.02055863  
2.68673382  
-17.76661123  
3.80986521  
0.00069222  
-1.47556685  
0.30604948  
-0.01233459  
-0.95274723  
0.00931168  
-0.52475838  
36.45948839  
  
E = 21.895  
.,
```

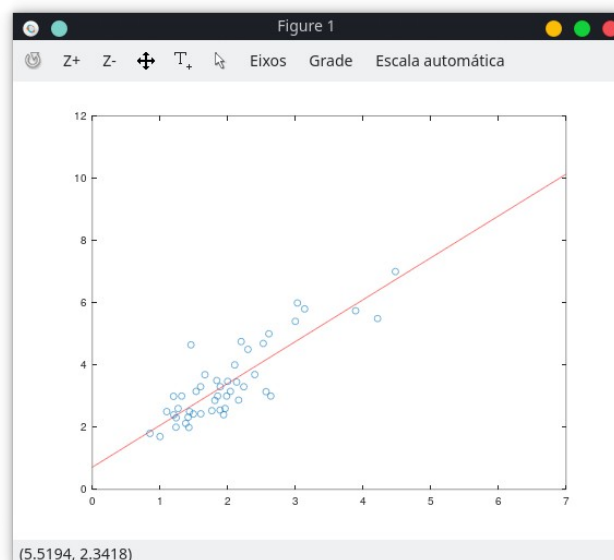
Conversando com colegas achei uma boa solução para somente duas, poderia ter melhores ,mas não testei todas: **RM dis**

```
th =  
    8.80141  
    0.48885  
   -34.63605  
  
E = 42.588
```

6) Implemente o algoritmo de regressão utilizando o algoritmo do gradiente descendente unidimensional aplicando sobre primeira coluna dos dados de housing_1.txt. Mostre o gráfico dos dados com a reta de regressão e os coeficientes obtidos. (2 pontos)

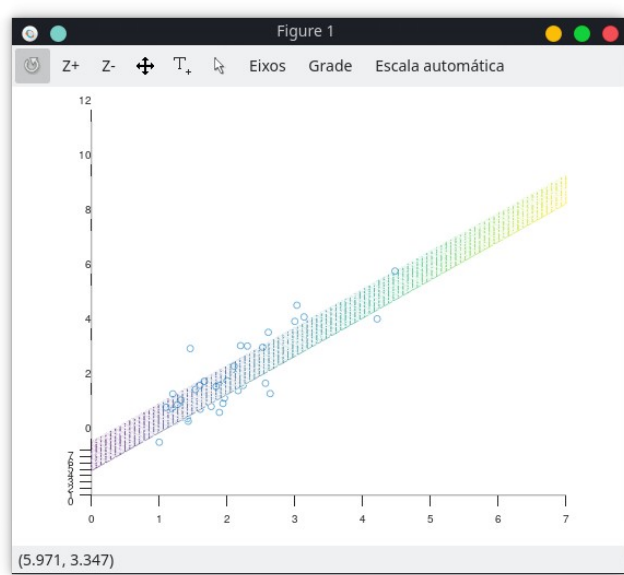
(O algoritmo do gradiente descendente é iterativo, possuindo uma etapa para cálculo do erro além de uma etapa para atualização dos coeficientes. Possui um parâmetro de ajuste da taxa de aprendizado e a cada atualização dos coeficientes espera-se que o erro diminua convergindo para um ponto de mínimo)

2D:



```
th =  
    1.34525  
    0.71270  
  
E = 0.41163
```

3D:



th =

1.392107
-0.087380
0.895979

E = 0.40866