MULTIVARIATE LINEAR REGRESSION

Prof. Nielsen Rechia nielsen.machado@uniritter.edu.br

	Tamanho (pés²)	Quartos	Pisos	Anos	Preço (\$1000)
X ⁽¹⁾	2104	5	1	45	460
X ⁽²⁾	1416	3	2	40	232
X ⁽³⁾	1534	3	2	30	315
X ⁽⁴⁾	852	2	1	36	178
X ⁽ⁿ⁾	0 0 0	• • •	• • •	0 0 0	
	X ₁	X ₂	X ₃	X ₄	У

Temos que adicionar um valor x₀ para conseguir realizar multiplicação de matrizes

Regressão Linear univariada $\hat{f}(x) = \theta_0 + \theta_1 x$

Regressão Linear multivariada $\hat{f}(\mathbf{x}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + ... + \theta_m x_m$ $\hat{f}(\mathbf{x}) = \theta_0 + \sum_{i=1}^m \theta_i x_i$

$$\hat{f}(\mathbf{x}) = \theta_0 + \sum_{i=1}^m \theta_i x_i$$

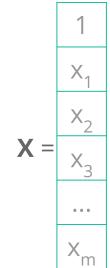
Vamos assumir a existência de um novo atributo x₀

		Tamanho (pés²)	Quartos	Pisos	Anos	Preço (\$1000)
x ⁽¹⁾	1	2104	5	1	45	460
X ⁽²⁾	1	1416	3	2	40	232
X ⁽³⁾	1	1534	3	2	30	315
X ⁽⁴⁾	1	852	2	1	36	178
X ⁽ⁿ⁾	1	•••			• • •	•••
	X ₀	X ₁	X ₂	X ₃	X ₄	У

Então podemos reescrever $\hat{f}(\mathbf{x}) = \theta_0 + \sum_{i=1}^{n} \theta_i x_i$

Como
$$\hat{f}(\mathbf{x}) = \boldsymbol{\Theta}^T \mathbf{x}$$

$$\mathbf{\Theta}^{\mathsf{T}} = \begin{bmatrix} \mathbf{\Theta}_0 & \mathbf{\Theta}_1 & \mathbf{\Theta}_2 & \mathbf{\Theta}_3 & \cdots & \mathbf{\Theta}_{\mathsf{m}} \end{bmatrix}$$



Exercícios

Com os dados abaixo realize uma predição

Instância

Tamanho (pés²)	Quartos	Pisos	Anos	Preço (\$1000)
1416	3	2	40	323

Pesos:
$$\Theta_0 = 1$$
, $\Theta_1 = 0.3$, $\Theta_2 = 0.4$, $\Theta_3 = 0.4$, $\Theta_0 = 0.5$

Resposta: = 455.80

Temos um modelo: $\hat{f}(\mathbf{x}) = \boldsymbol{\Theta}^T \mathbf{x}$

Alguns parâmetros: Θ

Uma função de custo:
$$J(\Theta) = \frac{1}{2N} \sum_{i=1}^{N} \left(\Theta^{T} \mathbf{x}^{(i)} - f(\mathbf{x}^{(i)}) \right)^{2}$$

Um objetivo, minimizar a função de custo

Atualização dos pesos neste cenário:

repetir até convergir {

$$\theta_j := \theta_j - \alpha \frac{1}{N} \sum_{i=1}^N \left[\left(\Theta^T \mathbf{x}^{(i)} - f(\mathbf{x}^{(i)}) \right) x_j^{(i)} \right]$$

$$para \ (j = 0, 1, ..., m)$$

$$para (j = 0,1,...,m)$$

Armazenar os novos valores em variáveis temporárias!

x_o é **sempre** 1

Dicas:

Quando o dataset possui poucos atributos ou quando o conjunto de dados possui dados não lineares podemos utilizar estratégias como padronização ou transformação de dados

Ou podemos utilizar equações diferentes para o modelo

Exercício prático

Realizar a regressão linear com base no dataset Casas/ex1data2.csv utilizando a regressão linear do sklearn.

http://scikit-learn.org/stable/modules/generated/sklearn.linear_mo

del.LinearRegression.html

```
from sklearn import linear_model

clf = linear_model.LinearRegression(normalize=True)
clf.fit(trainset_xs, trainset_fx)
predictions = clf.predict(testset_xs)
score = clf.score(trainset_xs, trainset_fx)
```

Exercício para entregar

Implemente uma regressão linear multivariada para prever o preço de uma casa, baseado em outros atributos (dataset Casas.data.csv

Utilize 1 para todos os parâmetros Θ0, Θ1... Θ13)

batch_size=8

Encontre um valor de α que atinja a convergência em no máximo 10 iterações

Atualize os pesos

Defina % de dados para treino e teste e faça variações

Qual o melhor score? Aconteceu Overfitting ou Underfitting em algum caso?

Exercício para entregar (conti ...)

Basta implementar a hipotese de custo como abaixo

theta.T * array(1+ 13 atributos)

E adicionar novas atualizações de peso para os thethas de todas variáveis de 2 até 13.

Conclusão

Leitura recomendada:

Apêndice D de Introduction to Data Mining

