MULTIVARIATE LINEAR REGRESSION

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

Regressão Linear Multivariada

	Tamanho (pés^2)	**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
$x^{(n)}$	...	...	...	...	...
	x_1	x_2	x_3	x_4	y

Regressão Linear Multivariada

Temos que adicionar um valor x_0 para conseguir realizar multiplicação de matrizes

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

Regressão Linear multivariada $\hat{f}(\mathbf{x}) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \ldots + \theta_m x_m$

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

Regressão Linear Multivariada

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

Vamos assumir a existência de um novo atributo x_0

		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**	…	…	…	…	…
	$\mathbf{x_0}$	x_1	x_2	x_3	x_4	y

Regressão Linear Multivariada

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

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

$$\Theta^T = \begin{bmatrix} \Theta_0 & \Theta_1 & \Theta_2 & \Theta_3 & \dots & \Theta_m \end{bmatrix}$$

$$\mathbf{X} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_3 \\ \dots \\ x_m \end{bmatrix}$$

Exercícios

Com os dados abaixo realize uma predição

Instância

Tamanho (pés^2)	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: [redacted] **= 455.80**

Regressão Linear Multivariada

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

Alguns parâmetros: $\mathbf{\Theta}$

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

Regressão Linear Multivariada

Atualização dos pesos neste cenário:

$$\textit{repetir até convergir} \ \{$$

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

$$\textit{para} \ (j = 0, 1, \ldots, m)$$

Armazenar os novos valores em variáveis temporárias!

x_0 é **sempre** 1

Regressão Linear Multivariada

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_model.LinearRegression.html

```
In [ ]: 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 $\Theta 0$, $\Theta 1$… $\Theta 13$)
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