Função Custo para Regressão

Continuando a disciplina...

Continuaremos agora com a parte teórica da disciplina, onde aprenderemos o conceito de **função custo**.

Definição informal de função custo:

A função custo visa quantificar o quão bem um modelo está se saindo ao tentar aproximar os dados.

Terminologia

função custo = função objetivo

Tabela com os dados

Área da casa [m^2]	Custo em R$
32	51.000
149	265.000
78	110.000
...	...
220	315.000

Considere que os dados da tabela acima serão modelados por meio de uma função linear

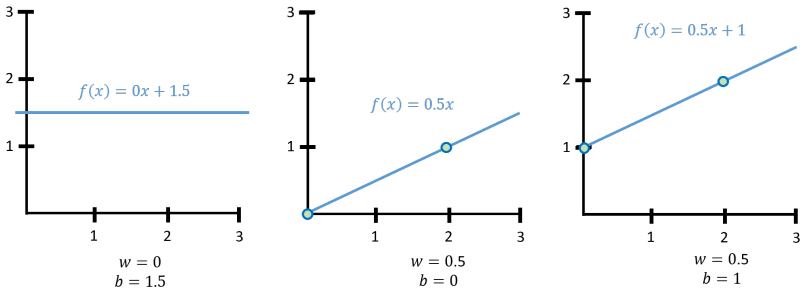
$$f(x) = wx + b$$

- w, b: parâmetros do modelo = coeficientes do modelo = pesos do modelo

"Em ML, parâmetros do modelo são as variáveis que podem ser ajustadas, durante o treinamento, com objetivo de melhorar o desempenho do modelo"

O que os parâmetros w, b definem?

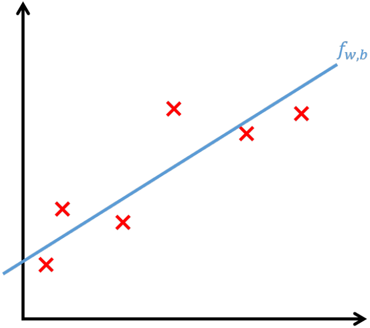
$$f(x) = wx + b$$



O que os parâmetros w, b definem?

Pergunta:

A reta abaixo aproxima bem os dados? Por que?

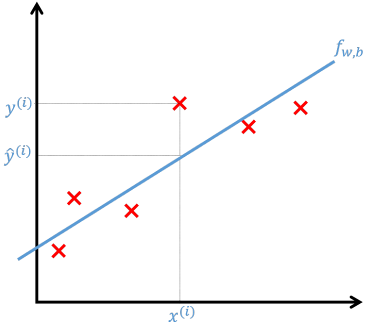


$$f_{w,b}(x) = wx + b$$

$$\hat{y} = f_{w,b}(x)$$

O que os parâmetros w, b definem?

Para uma amostra de treinamento específica, temos:



$$f_{w,b}(x^{(i)}) = wx^{(i)} + b$$
$$\hat{y}^{(i)} = f_{w,b}(x^{(i)})$$

Definindo o problema

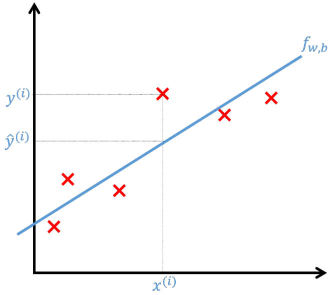
Encontrar w, b que faz com que $\hat{y}^{(i)}$ seja próximo de $y^{(i)}$ para todas as amostras $\left(x^{(i)}, y^{(i)}\right)$

Opção 1

Encontrar w, b que faz com que

$$J(w,b) = \left(\hat{y}^{(1)} - y^{(1)}\right) + \left(\hat{y}^{(2)} - y^{(2)}\right) + \cdots + \left(\hat{y}^{(6)} - y^{(6)}\right)$$

assuma o menor valor possível. Lembrar que $y^{(i)} = wx^{(i)} + b$



Pergunta:

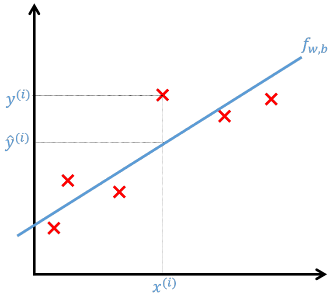
É uma boa ideia definir J como um somatório simples de erros?

Opção 2

Encontrar w, b que faz com que

$$J(w,b) = \left(\hat{y}^{(1)} - y^{(1)}\right)^2 + \left(\hat{y}^{(2)} - y^{(2)}\right)^2 + \cdots + \left(\hat{y}^{(6)} - y^{(6)}\right)^2$$

assuma o menor valor possível. Lembrar que $y^{(i)} = wx^{(i)} + b$



Pergunta:

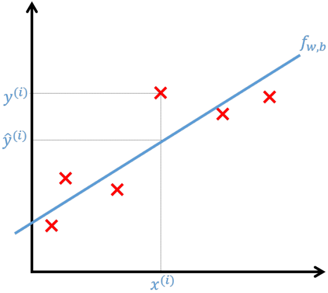
É uma boa ideia definir J como um somatório quadrático de erros?

Opção 3

Encontrar w, b que faz com que

$$J(w,b) = \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

assuma o menor valor possível. Lembrar que $y^{(i)} = wx^{(i)} + b$



Pergunta:

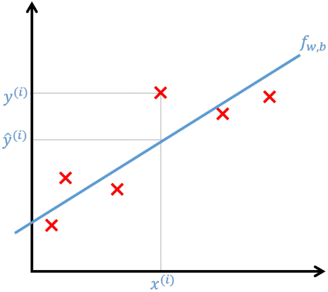
O que foi feito aqui?

Opção 4

Encontrar w, b que faz com que

$$J(w,b) = \frac{1}{m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

assuma o menor valor possível. Lembrar que $y^{(i)} = wx^{(i)} + b$



Observação

Para que J não tenda a aumentar à medida com que m aumenta, podemos fazer o chamado **Erro Quadrático Médio**.

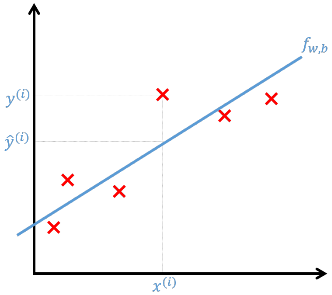
Definindo a função objetivo

Opção 5

Encontrar w, b que faz com que

$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}\left(\hat{y}^{(i)} - y^{(i)}\right)^2$$

assuma o menor valor possível. Lembrar que $y^{(i)} = wx^{(i)} + b$



Observação

Uma divisão por "$2m$" ao invés de "m" também é comum em problemas de ML.

Observação 1

Em problemas de regressão, pessoas irão definir diferentes tipos de função objetivo dependendo da aplicação. Entretanto, cumpre destacar que funções do tipo **mínimos quadrados** são vastamente utilizadas.

Observação 2

Nós seguiremos com a definição

$$J(w, b) = \frac{1}{2m} \sum_{i=1}^{m} \left(\hat{y}^{(i)} - y^{(i)} \right)^2$$

que também pode ser reescrita conforme abaixo

$$J(w, b) = \frac{1}{2m} \sum_{i=1}^{m} \left(f_{w,b} \left(x^{(i)} \right) - y^{(i)} \right)^2$$

Buscaremos encontrar valores para w, b que fazem com que $J(w, b)$ seja o menor possível.

The cost function used for linear regression is

$$J(w,b) = \frac{1}{2m}\sum_{i=1}^{m}(f_{w,b}(x^{(i)}) - y^{(i)})^2$$

Which of these are the parameters of the model that can be adjusted?

- w and b
- $f_{w,b}(x^{(i)})$
- w only, because we should choose b=0
- $\hat{y}$

Fonte: **Machine Learning Specialization**, *deeplearning.ai*, Stanford Online, Coursera.org.

Um resumo

Modelo:

$$f_{w,b}(x) = wx + b$$

Parâmetros:

$$w, b$$

Função custo:

$$J(w, b) = \frac{1}{2m} \sum_{i=1}^{m} \left(f_{w,b}\left(x^{(i)}\right) - y^{(i)} \right)^2$$

Objetivo:

$$\min_{w,b} J(w, b)$$

O que a função objetivo de fato calcula?

Para ganharmos um pouco mais de intuição acerca disso, vamos por enquanto considerar o **modelo simplificado**:

$$f_w(x) = wx \quad \rightarrow \quad \text{é como se estivéssemos considerando } b = 0$$

Ou seja, agora o problema é conforme abaixo

Modelo:

$$f_w(x) = wx$$

Parâmetro:

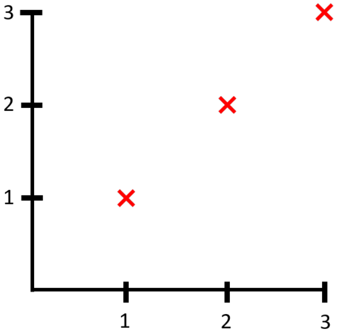
$$w$$

Função custo:

$$J(w) = \frac{1}{2m} \sum_{i=1}^{m} \left(f_w \left(x^{(i)} \right) - y^{(i)} \right)^2$$

Objetivo:

$$\min_{w} J(w)$$



Pergunta:

Para os dados acima, qual é o valor de w que resulta no melhor modelo $f_w(x) = wx$?

A) $w = 1$

B) $w = 0$

C) $w = 0.5$

Exemplo

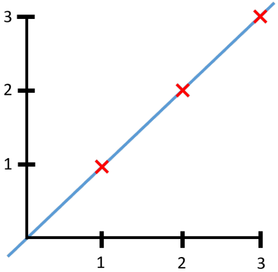
Para $w = 1$, temos

Modelo:

$$f_w(x) = wx \quad \rightarrow \quad f_1(x) = x \quad \rightarrow \quad \text{Esse modelo encontra-se representado na figura abaixo}$$

Função custo:

$$J(w) = \frac{1}{2m} \sum_{i=1}^{m} \left(f_w\left(x^{(i)}\right) - y^{(i)} \right)^2 \quad \rightarrow \quad J(1) = \frac{1}{2 \cdot 3} \sum_{i=1}^{3} \left(x^{(i)} - y^{(i)} \right)^2 = 0$$



Pergunta: É um bom valor para w?

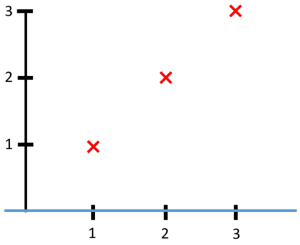
Para $w = 0$, temos

Modelo:

$$f_w(x) = wx \quad \rightarrow \quad f_0(x) = 0 \quad \rightarrow$$ Esse modelo encontra-se representado na figura abaixo

Função custo:

$$J(w) = \frac{1}{2m} \sum_{i=1}^{m} \left(f_w\left(x^{(i)}\right) - y^{(i)} \right)^2 \quad \rightarrow \quad J(0) = \frac{1}{2 \cdot 3} \sum_{i=1}^{3} \left(0 - y^{(i)} \right)^2 \approx 2.33$$



Pergunta: É um bom valor para w?

Exemplo

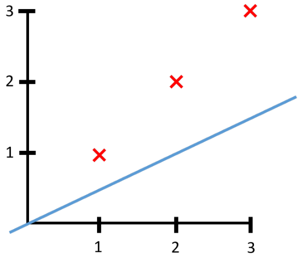
Para $w = 0.5$, temos

Modelo:

$$f_w(x) = wx \quad \rightarrow \quad f_{0.5}(x) = 0.5x \quad \rightarrow$$ Esse modelo encontra-se representado na figura abaixo

Função custo:

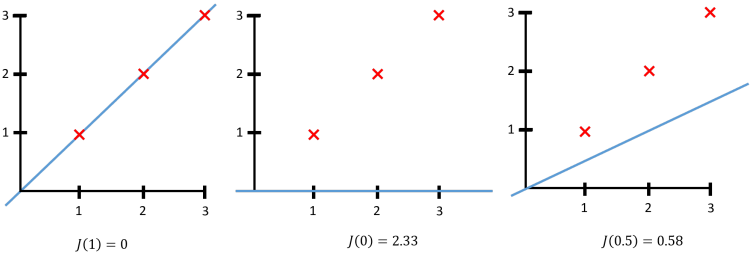
$$J(w) = \frac{1}{2m}\sum_{i=1}^{m}\left(f_w\left(x^{(i)}\right) - y^{(i)}\right)^2 \quad \rightarrow \quad J(0.5) = \frac{1}{2 \cdot 3}\sum_{i=1}^{3}\left(0.5x^{(i)} - y^{(i)}\right)^2 \approx 0.58$$



Pergunta: É um bom valor para w?

Exemplo

Testamos 3 valores para w, conforme resumido abaixo



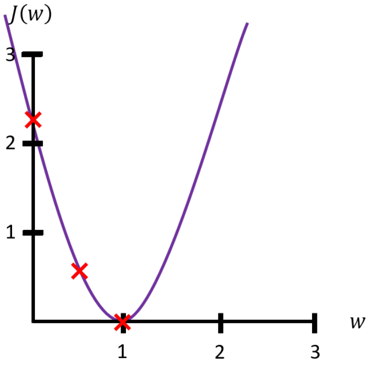
Pergunta 1:

Qual valor para w produz o menor "custo"?

Pergunta 2:

Seria possível plotar um gráfico de $J(w)$ *versus* w? Qual seria sua forma?

Exemplo



Observação:

Note que agora temos um gráfico $J(w)$ *versus* w, e não y *versus* x

Quiz

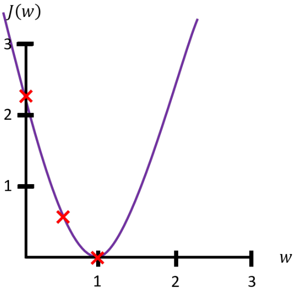
When does the model fit the data relatively well, compared to other choices for parameter w?

- When $f_w(x)$ is at or near a minimum for all the values of x in the training set.
- When w is close to zero.
- When the cost J is at or near a minimum.
- When x is at or near a minimum.

Fonte: **Machine Learning Specialization**, *deeplearning.ai*, Stanford Online, Coursera.org.

Observação importante

Quando a reta passa próximo dos dados, $J(\omega)$ é pequeno.



Escolher um valor de w que minimiza $J(w)$ parece uma boa opção! Ou seja

$$\min_{w} J(w)$$

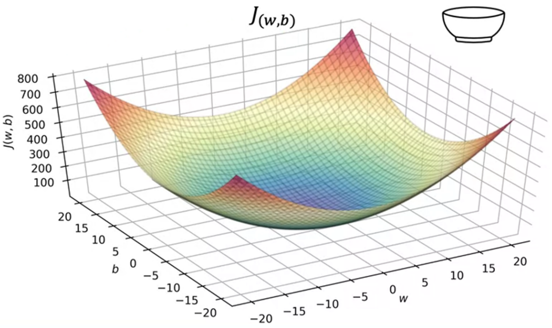
Essa ideia também é válida mesmo quando temos um modelo mais genérico:

$$\min_{w,b} J(w,b)$$

Pergunta:

Mas qual seria a forma que $J(w,b)$ tem quando temos 2 parâmetros ao invés de 1? Alguém sabe?

Como seria $J(w,b)$?



Fonte: **Machine Learning Specialization**, *deeplearning.ai*, Stanford Online, Coursera.org

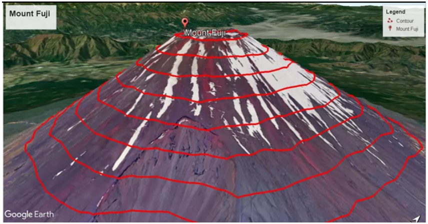
Treinando a interpretação...

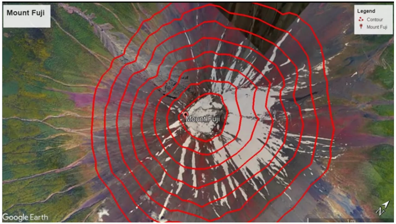
Indique no gráfico acima $J(-10, -15)$.

Pergunta:

Seria possível visualizar $J(w, b)$ como um **gráfico de contorno**?

Gráfico de contorno



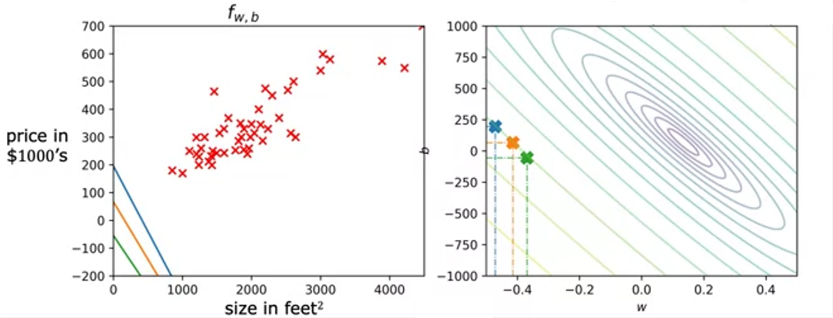


Fonte: **Machine Learning Specialization**, *deeplearning.ai*, Stanford Online, Coursera.org

O que são gráficos de contorno?

Gráficos de contorno são formados por cortes horizontais que indicam valores (“alturas”) semelhantes para uma função (“montanha”).

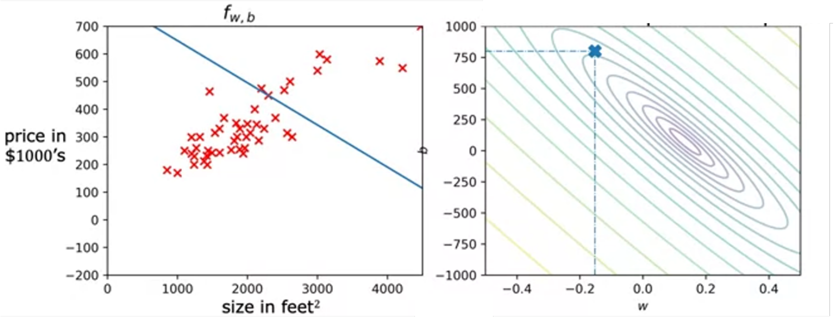
Considere o exemplo abaixo onde busca-se estimar o preço de casas (preço *versus* área).



Perguntas:

- No gráfico de contorno, o que representam os 3 pontos destacados? Quais são os seus valores aproximados de b e w? Que retas esses valores definem? Tais retas são bons modelos para os dados?
- Onde se encontra o mínimo da função $J(w, b)$? Quais são os valores aproximados para b e w nesse ponto? Que reta esses valores definem? Tal reta seria um bom modelo para os dados?

Visualizando o ponto $(-0.15; 800)$ no gráfico de contorno...

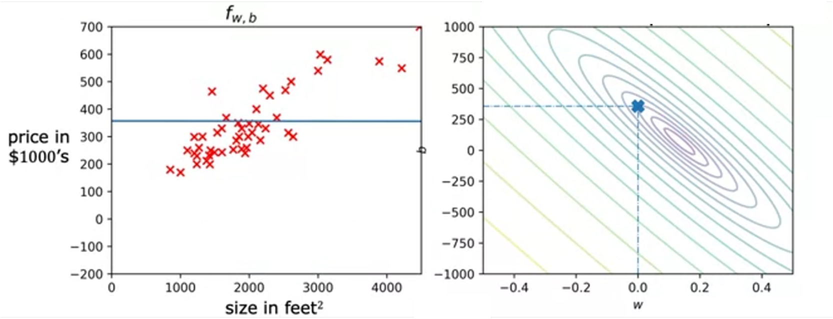


Observações

- Note que esse ponto leva a um valor de $J(w, b)$ que ainda está longe do mínimo global. Ou seja, não gera uma boa reta para representar os dados.

Exemplo

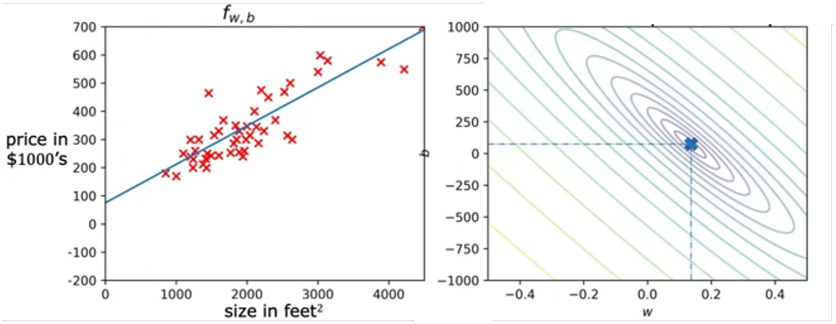
Visualizando o ponto $(0; 360)$ no gráfico de contorno...



Observações

- Estamos mais próximos do mínimo de $J(w, b)$, mas ainda um pouco longe do mínimo global.

Visualizando o ponto $(0.13; 71)$ no gráfico de contorno...



Observações

- Note que agora estamos praticamente no mínimo de $J(w, b)$
- Isso significa que as distâncias verticais existentes entre $f(x)$ e y foram minimizadas

Vamos agora ver como representar em código a nossa função custo.

Nome do arquivo que trabalharemos agora

código - Função custo.ipynb

Atividade de aula

Parte 1

Rode todo o "código - Função custo.ipynb" sem fazer qualquer tipo de alteração. Certifique-se de que você o compreendeu.

Parte 2

Insira no código da Parte 1 o conjunto de medições que você já criou anteriormente para um resistor de 50 Ω, faça as adaptações necessárias e verifique os resultados.