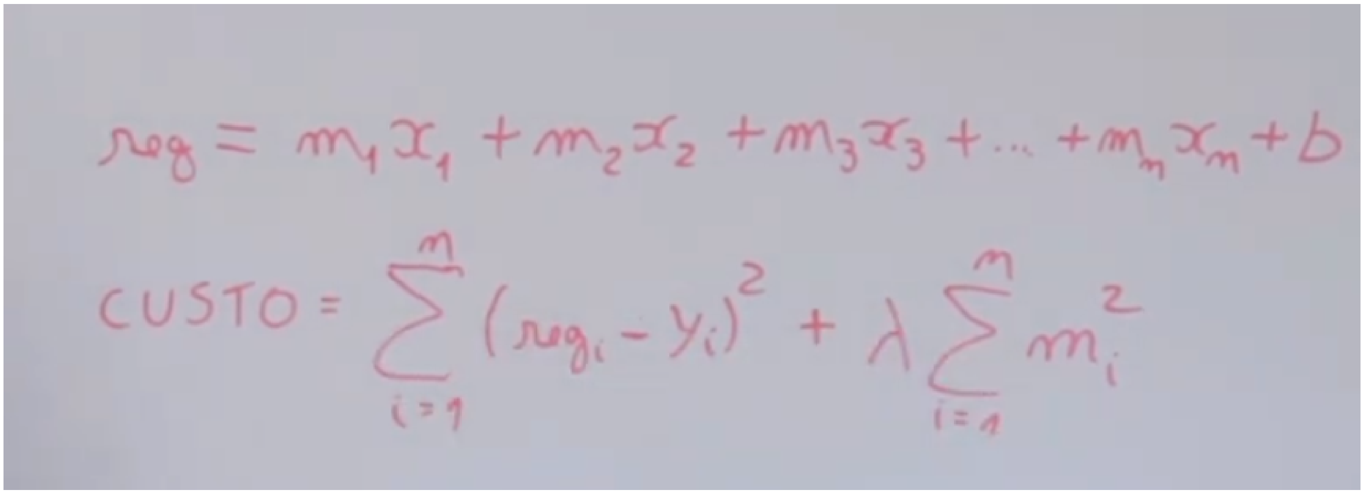
Ridge e Lasso Regression

A função de custo é diferente de uma regressão linear simples ou múltipla. Ela possuí uma constante normalizadora que é multiplicada pelo somatório de todos os pesos elevado ao quadrado.



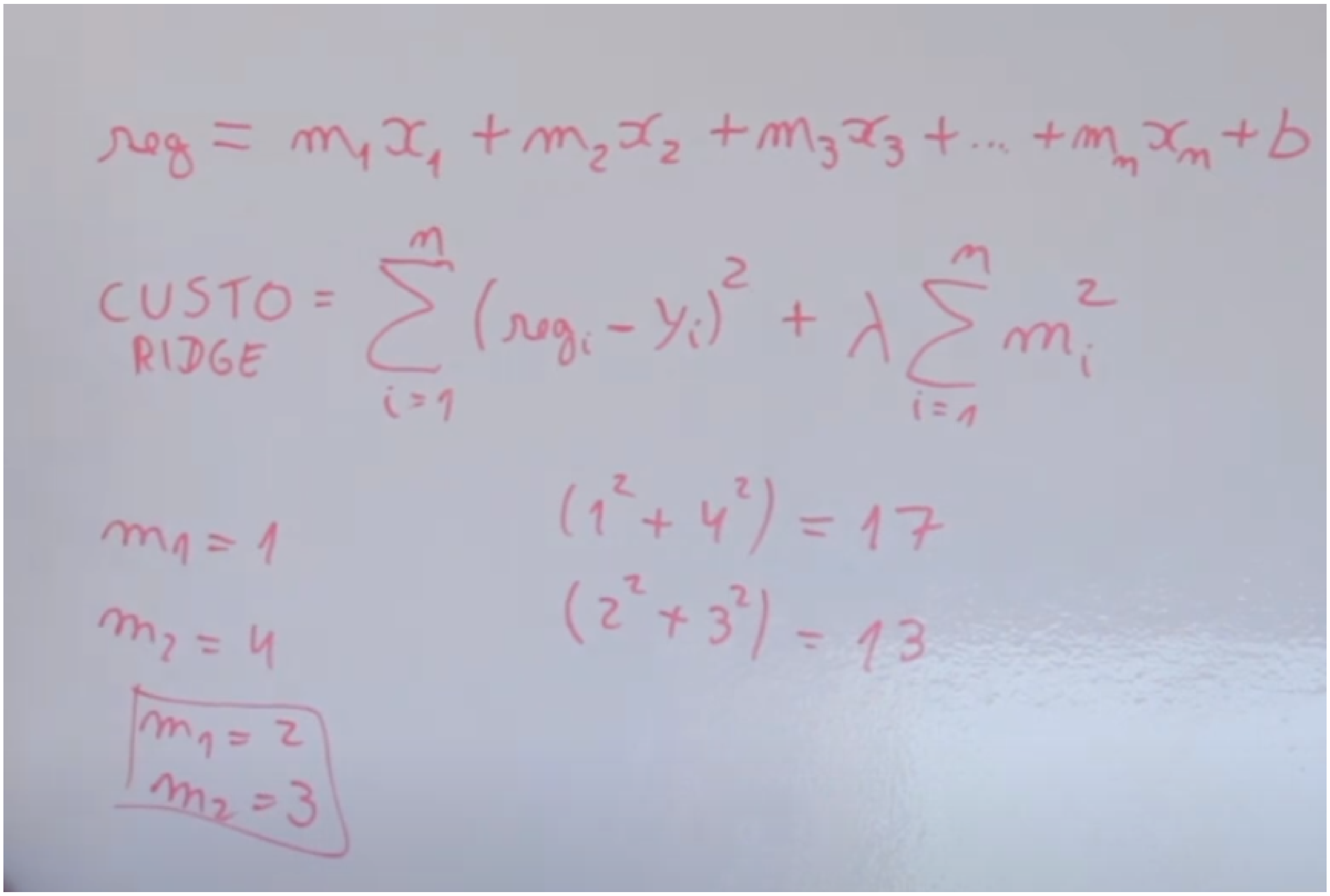
Queremos encontrar o menor erro desde que os valores de cada coeficiente sejam menores. Ou seja, se tivermos um coeficiente muito grande, ao elevar ao quadrado ele receberá muito peso, portanto o algoritmo adapta os valores dos coeficientes para ficarem mais próximos e mantendo o erro menor possível.

Impedir que uma variável recebe muito mais peso que as outras, ou seja "REGULARIZAÇÃO"

Mas para quê usar?

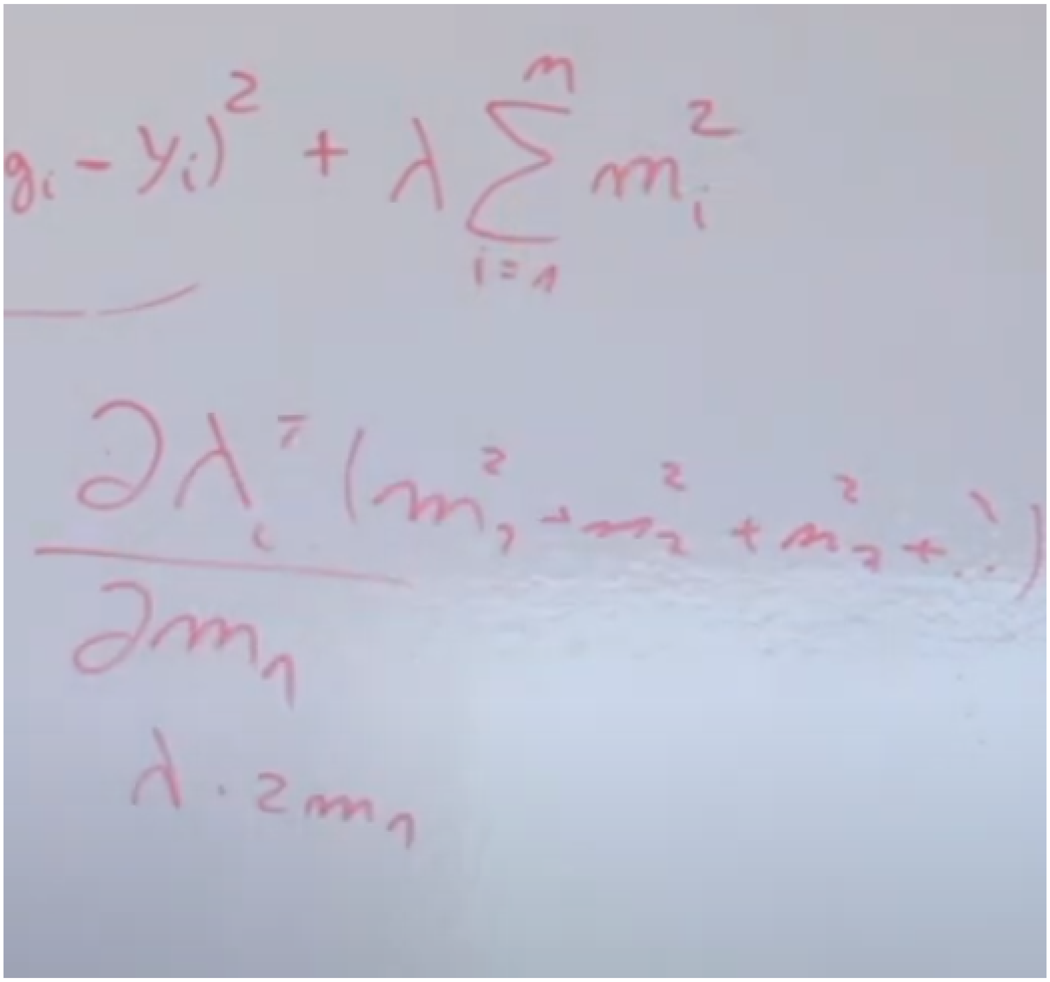
A função de custo da regressão linear é feita na base de treino, ou seja minimizamos o erro da reta baseado nos dados de treino, porém é necessário deixar o modelo mais generalizável para não causar overfitting, por isso é interessante regularizar os termos para que variáveis não recebam um peso maior do que deveriam durante o treinamento do modelo. O modelo desconfia dos pesos das variáveis.

Exemplo:



Veja que ao aproximar m1 de m2, o custo pode ter ficado um pouco pior, porém mais regularizado com os pesos dos coeficientes de 17 para 13.

Derivada



Ao realizar a derivada do termo em relação a m1 por exemplo, ficaria constante *2m1, pois todo o somatório ficaria constante, restando apenas m1 elevado ao quadrado, e a derivada disso é 2* m1.

Colabora com a colinearidade, mantendo os pesos na mesma ordem de grandeza. Ou seja Correlação de Pearson

A FUNÇÃO RIDGE UTILIZA A REGULARIZAÇÃO **L2** NA REGRESSÃO LINEAR. O TERMO L2 SE DEVE AO ELEVAR TODOS OS COEFICIENTES AO QUADRADO FAZENDO COM QUE HAJA A REGULARIZAÇÃO DOS PESOS DE CADA VARIÁVEL.

JÁ A FUNÇÃO L1 É O MÓDULO DOS COEFICIENTES, UTILIZADA NA FUNÇÃO LASSO. ELA TEM O MESMO OBJETIVO DE MINIMIZAR OS TERMOS, PORÉM MENOS RÍGIDA QUE A L2, TENDO UM PESO MENOR.

REGRESSÃO LASSO:

$$\text{CUSTO} = \sum_{i=1}^{m} \left(ng_i - y_i \right)^2 + \sum_{i=1}^{m} |m_i|$$