

1 Regressão Logística

1. Comente sobre algumas desvantagens causadas pelo uso de regressão linear para resolver um problema de classificação binária.
2. Apresente a função de hipótese da regressão logística, definindo cada elemento envolvido.
3. O que é fronteira de decisão? Como ela é obtida no caso da regressão logística?
4. Descreva matematicamente a função de custo da regressão logística, definindo cada elemento envolvido.
5. Descreva matematicamente o algoritmo do gradiente descendente no caso específico da regressão logística, definindo cada elemento envolvido.
6. Deduza matematicamente a expressão do gradiente descendente na regressão logística.
7. Descreva a vetorização da regressão logística.
8. Considere os seguintes dados tabelados:

x_1	x_2	y
0.1	0.3	1
0.6	-0.2	1
-1.2	-3.1	0

Obtenha θ_0 , θ_1 e θ_2 de um modelo de regressão logística $y \approx g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$ fazendo duas iterações do gradiente descendente e iniciando com $\theta_0 = \theta_1 = \theta_2 = 0$.

9. Descreva a abordagem “one-vs-all” usada em problemas com múltiplas classes.

2 *Overfitting*

1. Descreva os conceitos de *overfitting* (variância alta) e *underfitting* (viés alto) e a principal causa para cada um destes fenômenos.
2. Como poderia identificar estes problemas visualmente a partir da função de hipótese de uma regressão logística?
3. Defina a função de custo regularizada, tanto na regressão linear quanto na logística.
4. Em que sentido que a regularização pode melhorar um modelo como a regressão linear ou logística?

5. Por que o parâmetro constante θ_0 não é regularizado?
6. Deduza o gradiente descendente da regressão logística regularizada.
7. Mostre o sistema de equações normais no caso da regressão linear regularizada.