SCC0276 - Aprendizado de Máquina

Aula Regressão Logistica

Profa. Dra. Roseli Aparecida Francelin Romero
SCC - ICMC - USP

2019

Sumário

CLASSIFICAÇÃO

- Um **hipótese** para aprox. os dados (uma função)
- Uma **função custo** (que determinará o desempenho da hipótese)
- Um **algoritmo de otimização** para os ajustar os coeficientes ou pesos da hipótese.

CLASSIFICAÇÃO

- $h(x) = w_0 + w_1.x_1 + w_2.x_2$; ótimo na regressao linear pois temos que prever um valor continuo,
- Na classificação temos que classificar ou categorizar em SIM=1 ou NÃO=0 (valor binários).

Tabela de Inscrições

exam1_score	exam2_score	admitted
34.62365962451697	78.0246928153624	0
30.28671076822607	43.89499752400101	0
35.84740876993872	72.90219802708364	0
60.18259938620976	86.30855209546826	1
79.0327360507101	75.3443764369103	1
45.08327747668339	56.3163717815305	0
61.10666453684766	96.51142588489624	1
75.02474556738889	46.55401354116538	1
76.09878670226257	87.42056971926803	1

Figura 1: Tabela de notas de candidatos

CLASSIFICAÇÃO

- Mais concretamente, em nosso problema acima, precisamos produzir um "sim", sinalizando que o aluno será ACEITO, ou "não", sinalizando que o aluno NÃO FOI ACEITO. O Não podemos dizer, simplesmente, 45% "sim"!
- Matematicamente, podemos codificar esta informação em nossa equação (de h (x)) aplicando uma função sigmóide.

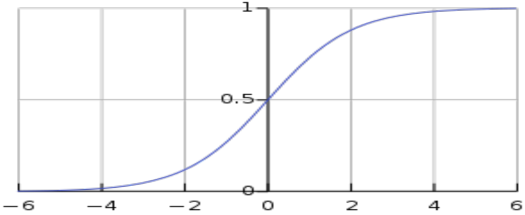
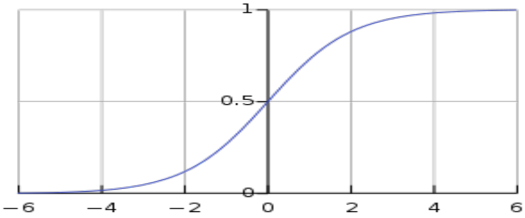


Figura 2: Função Sigmoid

Função Logistica ou Sigmoid

- 

- O alcance da função logística, como podemos ver, está entre 0 e 1. Portanto, se quisermos passar nossa saída de h (x) para a função logística, estaremos limitando o intervalo a [0, 1].

- $$g(h(x)) = 1/(1 + exp(-h(x)))$$

$$g(h(x)) = 1/(1 + \exp(-w_0 - w_1.x_1 - w_2.x_2))$$

Função Logistica

- Não é suficiente limitar o intervalo de h (x) a [0, 1]. Precisamos gerar um SIM (1) ou um NÃO (0). Para conseguir isso, devemos primeiro observar como a saída da função logística é.
- Fica claro que a curva é simétrica em torno de y = 0,5.
- Portanto, podemos afirmar com segurança que, se a saída da função logística for maior que 0,5, podemos tratá-la como um "SIM"e, se for menor ou igual a 0,5, podemos tratá-la como "NÃO". Formalmente falando, envie 'SIM' se: $g(h(x)) > 0,5$

Função Logistica

- $g(h(x)) > 0.5 \qquad h(x) > 0$
- o que significa: $w_0 + w_1x_1 + w_2x_2 > 0$
- $w_0 + w_1x_1 > -w_2x_2$
- $(\frac{-w_0}{w_2}) + (\frac{-w_1}{w_2}).x_1 > x_2$
- $c + mx > y \qquad c = \frac{-w_0}{w_2},$

  $m = \frac{-w_1}{w_2}\ , \qquad x = x_1$ and $y = x_2$

Função Logistica

- Portanto, a equação representa a parte do gráfico que está acima da linha y = mx + c.
- Como estávamos tentando codificar as informações de "SIM"ou "NAO"em g (h (x)) - o que nos levou a um ponto em que podemos dizer que a região acima da reta $w_0 + w_1.x_1 + w_2.x_2$ é onde emitimos " SIM ", e a região abaixo é onde produzimos"NÃO ".

Sumário

Função Custo Não-Convexa

- Não podemos usar a mesma função de custo que a regressão linear:
- $$J(\theta) = \frac{1}{m}\sum_{1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$
- Isto porque, na equação acima, se h (x) é a função logística, então J () será não-convexo (devido à presença de termo exponencial na função logística):

Nova Função de Custo

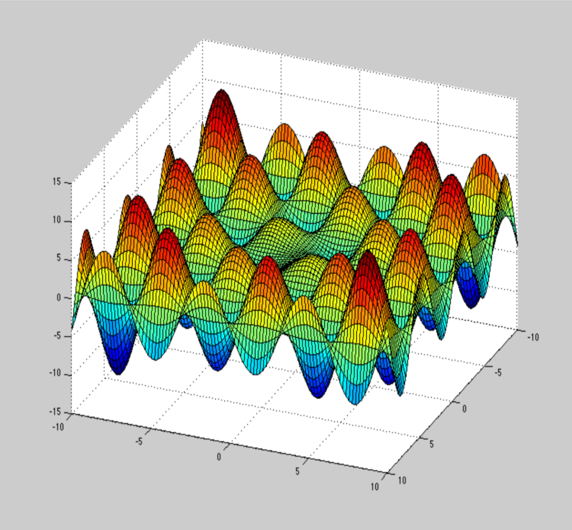


Figura 3: Função nao-convexa com varios pts de máximo

Nova Função de Custo

- Precisamos criar uma nova função de custo para problemas baseados em classificação
- Usando probabilidade condicional:
  h(x) = P(y = 1 , x)
  1 - h(x) = P(y = 0 , x) (pela regra de probabilidade complementar)
- Combinando as 2 equações, obtemos:
  P(y,x) = $(h(x)^y) * ((1 - h(x))^{(1-y)})$
- Como nossa tarefa é classificação, deseja-se maximizar a probabilidade de produzir o valor correto - independentemente do valor de 'y', podemos afirmar com segurança que a equação acima é uma boa representação de como nosso modelo está se desempenhando em um dado exemplo.

Nova Função de Custo

- Para determinar nosso modelo sobre todos os exemplos, nós devemos tomar o produto da eq. para todos os exemplos: $P(y,x) = [(h(x)^{y(1)}) * ((1 - h(x))^{(1-y(1))})] * [(h(x)^{y(2)}) * ((1 - h(x))^{(1-y(2))})] * [(h(x)^{y(3)}) * ((1 - h(x))^{(1-y(3))})] \ldots * [(h(x)^{y(N)}) * ((1 - h(x))^{(1-y(N))})]$
  onde y(1), y(2), ..., y(N) são N examples do nosso dataset.

Nova Função de Custo

- Para se livrar do produto e exponenciais, aplicando uma transformação de log para simplificar a equação:

$$log(P(y,x)) = \sum(log((h(x)^{y(i)}) * ((1-h(x))^{(1-y(i))})))$$

$$log(P(y,x)) = \sum(log(h(x)^{y(i)}) + log((1-h(x))^{(1-y(i))}))$$

$$log(P(y/x)) = \sum(y(i)log(h(x)) + (1-y(i))log((1-h(x))))$$

Esta equação é nova função custo e é conhecida com **cross-entropy cost function. FUNÇÃO CONVEXA**

Nova Função de Custo

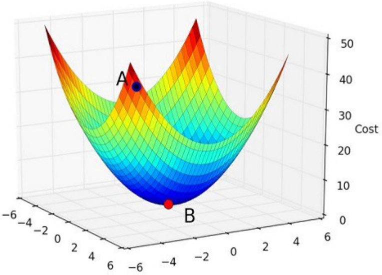


Figura 4: Função convexa

Sumário

Otimização

- Semelhante à Regressão Linear, podemos usar o Gradient Descent para otimizar nossa função de custo de entropia cruzada, sem nos preocuparmos com o fato de nosso modelo ficar preso a um ótimo local.