Regressão Logística

10 de julho de 2023

1 Exercícios

1. Considere um único ponto de dados com um único recurso. Seu objetivo é prever se um email é spam (1) ou não (0) com base na frequência de uma palavra. Se a frequência da palavra for 0.3 e o email não for spam, qual é o valor da função log-verossimilhança se $\beta = [0, 0]$? Use a fórmula para a função log-verossimilhança dada anteriormente.

2. Usando o mesmo ponto de dados do Exercício 1, qual é o gradiente da função log-verossimilhança se $\beta = [0, 0]$? Use as fórmulas para as derivadas parciais dadas anteriormente.

3. Ainda referindo-se ao ponto de dados do Exercício 1, se você usar a subida do gradiente com uma taxa de aprendizado de 0.1 para atualizar β, qual será o novo valor de β?

4. Agora considere um segundo ponto de dados com frequência de palavra 0.7 e o email é spam. Calcule o valor da função log-verossimilhança e seu gradiente para esses dois pontos de dados se $\beta = [0, 0]$.

5. Ainda usando os dois pontos de dados do Exercício 4, se você usar a subida do gradiente com uma taxa de aprendizado de 0.1 para atualizar β, qual será o novo valor de β?

6. Considere agora três pontos de dados: (0.3, 0), (0.7, 1), (0.1, 0), onde o primeiro número de cada par é a frequência da palavra e o segundo número indica se o email é spam. Calcule o valor da função log-verossimilhança e seu gradiente para esses três pontos de dados se $\beta = [0, 0]$. Em seguida, use a subida do gradiente com uma taxa de aprendizado de 0.1 para atualizar β. Qual é o novo valor de β?

2 Soluções

1. **Exercício 1:**

   Para calcular a função log-verossimilhança, primeiro precisamos calcular a probabilidade prevista $p = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}} = \frac{1}{1+e^{-(0+0*0.3)}} = 0.5$.

   Então a log-verossimilhança é $y\log(p)+(1-y)\log(1-p) = 0\log(0.5)+(1-0)\log(1-0.5) = -\log(2)$.

$$p = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}} = \frac{1}{1+e^{-(0+0\cdot 0.3)}} = 0.5$$
$$L(\beta) = y\log(p) + (1-y)\log(1-p) = 0\log(0.5) + (1-0)\log(1-0.5) = -\log(2)$$

2. **Exercício 2:**

   Para calcular o gradiente da função log-verossimilhança, precisamos calcular suas derivadas parciais em relação a β_0 e β_1.

Usando as fórmulas fornecidas, temos $\frac{\partial \log L(\beta)}{\partial \beta_0} = y - p = 0 - 0.5 = -0.5$ e $\frac{\partial \log L(\beta)}{\partial \beta_1} = (y-p)x = (0-0.5)0.3 = -0.15$. Portanto, o gradiente é $[-0.5, -0.15]$.

$$\frac{\partial \log L(\beta)}{\partial \beta_0} = y - p = 0 - 0.5 = -0.5$$
$$\frac{\partial \log L(\beta)}{\partial \beta_1} = (y-p)x = (0-0.5)0.3 = -0.15$$

3. **Exercício 3:**

   Para atualizar β, subtraímos a taxa de aprendizado vezes o gradiente de β. Portanto, $\beta \leftarrow \beta - \eta \nabla \log L(\beta) = [0, 0] - 0.1 * [-0.5, -0.15] = [0.05, 0.015]$.

   $$\beta \leftarrow \beta - \eta \nabla \log L(\beta) = [0, 0] - 0.1 \cdot [-0.5, -0.15] = [0.05, 0.015]$$

4. **Exercício 4:**

   Para dois pontos de dados, calculamos a probabilidade prevista e a log-verossimilhança para cada ponto e somamos os resultados.

   Para o primeiro ponto de dados (0.3, 0), temos $p = 0.5$ e a log-verossimilhança é $-\log(2)$, como calculamos no Exercício 1.

   Para o segundo ponto de dados (0.7, 1), também temos $p = 0.5$ e a log-verossimilhança é $\log(0.5) = -\log(2)$.

   Portanto, a log-verossimilhança para os dois pontos de dados é $-\log(2) - \log(2) = -2\log(2)$.

   O gradiente também é a soma dos gradientes para cada ponto de dados. Portanto, o gradiente para β_0 é $-0.5 - 0.5 = -1$ e o gradiente para β_1 é $-0.15 - 0.35 = -0.5$. O gradiente total é $[-1, -0.5]$.

   $$L(\beta) = -\log(2) - \log(2) = -2\log(2)$$
   $$\frac{\partial \log L(\beta)}{\partial \beta_0} = -0.5 - 0.5 = -1$$
   $$\frac{\partial \log L(\beta)}{\partial \beta_1} = -0.15 - 0.35 = -0.5$$

5. **Exercício 5:**

   Para atualizar β, subtraímos a taxa de aprendizado vezes o gradiente de β. Portanto, $\beta \leftarrow \beta - \eta \nabla \log L(\beta) = [0, 0] - 0.1 * [-1, -0.5] = [0.1, 0.05]$.

   $$\beta \leftarrow \beta - \eta \nabla \log L(\beta) = [0, 0] - 0.1 \cdot [-1, -0.5] = [0.1, 0.05]$$

6. **Exercício 6:**

   Para três pontos de dados, procedemos da mesma forma que no Exercício 4, somando as log-verossimilhanças e gradientes para cada ponto de dados.

   A log-verossimilhança para os três pontos de dados é $-\log(2) - \log(2) - \log(2) = -3\log(2)$.

O gradiente para β_0 é $-0.5-0.5-0.5 = -1.5$ e o gradiente para β_1 é $-0.15-0.35-0.05 = -0.55$. O gradiente total é $[-1.5, -0.55]$.

Para atualizar β, temos $\beta \leftarrow \beta - \eta\nabla \log L(\beta) = [0,0] - 0.1 * [-1.5, -0.55] = [0.15, 0.055]$.

$$\begin{aligned}
L(\beta) &= -\log(2) - \log(2) - \log(2) = -3\log(2) \\
\frac{\partial \log L(\beta)}{\partial \beta_0} &= -0.5 - 0.5 - 0.5 = -1.5 \\
\frac{\partial \log L(\beta)}{\partial \beta_1} &= -0.15 - 0.35 - 0.05 = -0.55 \\
\beta \leftarrow \beta - \eta\nabla \log L(\beta) &= [0,0] - 0.1 \cdot [-1.5, -0.55] = [0.15, 0.055]
\end{aligned}$$