Derivada de la función de costo para la regresión logística

Hassler Castro

2018-12-21

Recordemos que la función hipotesis o función de activación para nuestra regresión logística es la función sigmoid que luce así:

$$h_\theta(x) = \frac{1}{1+e^{-x}}$$

Y hemos definido que el costo para la regresión logística se obtiene a través de la siguiente expresión:

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}[-y^{(i)}log(h_\theta(x)) - (1-y^{(i)})log(1-h_\theta(x))]$$

derivar la función de costo nos permitirá encontrar valores de θ tales que reduzcan el costo, esto es, valores θ que serán coeficientes para cada una de las características de un ejemplo dentro del conjunto de entrenamiento/prueba y que permitirán estar más cerca del valor real. Considerando un ejemplo, lo que deseamos es:

$$\hat{z} = theta_1 * x_1^{(i)} + theta_2 * x_2^{(i)} + .. + theta_j * x_j^{(i)}$$

$$Objetivo : h_\theta(\hat{z}) \approx y_{real}$$

Asi que empiece la diversión:

$$\frac{\partial J}{\partial \theta}\frac{1}{m}\sum_{i=1}^{m}[-y^{(i)}log(h_\theta(x)) - (1-y^{(i)})log(1-h_\theta(x))]$$

Derivada de una sumatoria, suma de las derivadas.

$$\frac{1}{m}\sum_{i=1}^{m}\frac{\partial J}{\partial \theta}[-y^{(i)}log(h_\theta(x)) - (1-y^{(i)})log(1-h_\theta(x))]$$

Vamos hacer esto facil de masticar tenemos dos terminos restando, derivemos el de la izquierda, luego el de la derecha y al final juntamos los resultados, de esta forma será más facil seguir el paso a paso.

Empecemos entonces a derivar el término del lado izquierdo, nótese que la sumatoria fue removida significa que estamos trabajando con un solo ejemplo, sin embargo, después debemos agregar cada uno de los errores.

$$\frac{\partial J}{\partial \theta}[-y^{(i)}log(h_\theta(x))]$$

** Derivada de la multiplicacion en palabras sabias *la derivada del primero por el segundo normal, mas (+), el primero normal por la derivada del segundo* pero si a usted le gusta más formal

$$(f.g) = f' * g + f * g'$$

** Derivada del logaritmo es *(1 / el argumento multiplicado) por(*) la derivada del argumento* pero si a usted le gusta la cosa formal

$$(log(f)) = (1/f) * f'$$

** Derivada de la división con el argot del autor, *la derivada del numerador por el denominador, menos, la derivada del denominador por el numerador, todo lo anterior sobre, el denominador al cuadrado* una vez más, si usted es formal

$$(f/g) = \frac{g * f' - f * g'}{(g^2)}$$

Así reemplazando la funcion de hipotesis dentro del logaritmo y derivando obtenemos, debe tenerse en cuenta que estamos derivando respecto a theta por lo tanto y se trata como una constante.

$$\frac{-y^i}{\frac{1}{1+e^{-\theta x^{(i)}}}} * \frac{e^{-\theta x^{(i)}}}{(1+e^{-\theta x^{(i)}})^2} x^{(i)}$$

Una cosa cuanto menos horrorosa, vamos a simplificar y obtenemos

$$-y^{(i)}e^{-\theta x^{(i)}}h_\theta(x)x^{(i)} \tag{1}$$

Con lo anterior listo,y las definiciones establecidas es tiempo de derivar el segundo termino, y se deja como ejercicio para el lector pues es analogo y trivial, solo fue un chiste, acompañame a la siguiente página.

El segundo termino es el siguiente:

$$\frac{\partial J}{\partial \theta}[(1 - y^{(i)})log(1 - h_\theta(x)]$$

Al derivar se obtiene,

$$\frac{1 - y^i}{1 - \frac{1}{1+e^{-\theta x^{(i)}}}} * \frac{-e^{-\theta x^{(i)}}}{(1 + e^{-\theta x^{(i)}})^2} x^{(i)}$$

Vamos a simplificar este galimatias, cancelamos un par de terminos, para obtener:

$$-[(1 - y^{(i)})h_\theta(x)x^{(i)}] \tag{2}$$

Es hora de restar la ecuacion(1) y la ecuacion(2) ya derivadas, notese que la ecuacion dos la precede un signo menos, entonces obtenemos,

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}[-y^{(i)}e^{-\theta x^{(i)}}h_\theta(x)x^{(i)}] + [(1 - y^{(i)})h_\theta(x)x^{(i)}] \tag{3}$$

Empieza el algebra de octavo grado y la carpinteria matemática para reducir este expresión

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}[-y^{(i)}e^{-\theta x^{(i)}}h_\theta(x)x^{(i)}] + h_\theta(x)x^{(i)} - y^{(i)}h_\theta(x)x^{(i)}$$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}h_\theta(x)x^{(i)} - y^{(i)}[e^{-\theta x^{(i)}}h_\theta(x)x^{(i)} + h_\theta(x)x^{(i)}]$$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}h_\theta(x)x^{(i)} - y^{(i)}[h_\theta(x)x^{(i)}(e^{-\theta x^{(i)}} + 1)] \tag{4}$$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}h_\theta(x)x^{(i)} - y^{(i)}x^{(i)}$$

$$J(\theta) = \frac{1}{m}\sum_{i=1}^{m}(h_\theta(x) - y^{(i)})x^{(i)}$$

En la ecuación numero 4 se reemplaza la ecuacion de la hipotesis y se cancela el denominador, para resultar finalmente solo con x.

Y así es como se deriva la función de costo de manera simple y sin atajos extraños.