Concentración IA Avanzada para la Ciencia de Datos

Tarea I: Vectorización

Fecha de entrega: 07 de octubre de 2022.

Integrantes:

A01652277 | Samuel Méndez Villegas

A01653115 | Iker Ledesma Durán

A01731813 | Mariana Pérez Carmona

A01734337 | Nancy Lesly Segura Cuanalo

A01750164 | Paul Martín García Morfín

1. **Función de costo:** Como vimos en clase, la función de costo del modelo de regresión logística es la siguiente:

$$C(w) = -\frac{1}{n}\sum_{i=1}^{n} y_i log\big(h(x_i)\big) + (1 - y_i)log\big(1 - h(x_1)\big).$$

Demuestra que la siguiente expresión es equivalente.

$$C(w) = -\frac{1}{n}\Big(y^T log\big(h(Xw)\big) + (1 - y)^T log\big(1 - h(Xw)\big)\Big).$$

Ecuación 1

Se usa la convención en la que todos los vectores son vectores columna. X es la matriz de datos cuyas filas son los datos x_i^T. Partiendo de que una función escalar que se aplica a un vector se aplica a la entrada, tenemos:

$$\begin{aligned} C(w) &= -\frac{1}{n}\sum_{i=1}^{n} y_i log\big(h_w(x_i)\big) + (1 - y_i)log\big(1 - h_w(x_1)\big) \\ &= -\frac{1}{n}\Big(y^T log\big(h(Xw)\big) + (1^T - y^T)log\big(1 - h(Xw)\big)\Big). \end{aligned}$$

Finalmente:

$$C(w) = -\frac{1}{n}\Big(y^T log\big(h(Xw)\big) + (1 - y)^T log\big(1 - h(Xw)\big)\Big).$$

2. **Gradiente de la función de costo:** tomando como punto de partida la ecuación (1), demuestra que el gradiente de dicha función está dado por:

$$\nabla C(w) = \frac{1}{n} X^T (h(Xw) - y).$$

La derivada, que es el Jacobiano, de C con respecto a w se obtiene utilizando la regla de la cadena. Se tiene que:

$$\mathrm{D_v f(Mv) = diag\big(f'(Mv)\big)M}$$

donde,

diag se refiere a la matriz diagonal

M es una matriz

v es un vector columna

Calculando:

$$nD_w C = -y^T \big[diag\big((1-h)(Xw)\big)\big] X - (1-y)^T \big[diag\big(-h(Xw)\big)\big] X.$$

$$= -y^T X + 1^T \big[diag\big(h(Xw)\big)\big] X.$$

$$= -y^T X + \big(h(Xw)\big)^T X.$$

$$(D_w C)^T = \frac{1}{n} X^T (h(Xw) - y).$$

Finalmente:

$$\nabla C(w) = \frac{1}{n} X^T (h(Xw) - y).$$

Si quisiéramos obtener el gradiente de la función de costo expresada de forma tradicional, se calculan las derivadas parciales:

$$\frac{\partial C(w)}{\partial w_j} = \frac{\partial}{\partial w_j} - \frac{1}{n} \sum_{i=1}^{n} \big[y_i log\big(h(x_i)\big) + (1 - y_i) log\big(1 - h(x_1)\big)\big].$$

$$= -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \frac{\partial}{\partial w_j} log\big(h(x_i)\big) + (1 - y_i) \frac{\partial}{\partial w_j} log\big(1 - h(x_1)\big)\right].$$

Se aplica regla de la cadena:

$$= -\frac{1}{n} \sum_{i=1}^{n} \left[y_i \frac{\frac{\partial}{\partial w_j} h(x_i)}{h(x_i)} + (1 - y_i) \frac{\frac{\partial}{\partial w_j} \big(1 - h(x_i)\big)}{1 - h(x_i)}\right].$$

Sustituyendo $\mathrm{h}(\mathrm{x}) = \sigma(\mathrm{wx})$:

$$= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \frac{\frac{\partial}{\partial w_j}\sigma(\mathrm{wx_i})}{h(x_i)} + (1-y_i)\frac{\frac{\partial}{\partial w_j}\left(1-\sigma(\mathrm{wx_i})\right)}{1-h(x_i)}\right].$$

Derivada de la función sigma:

$$\frac{d}{dx}\sigma(x) = \frac{d}{dx}\left(\frac{1}{1+e^{-x}}\right) = \sigma(x)\left(1-\sigma(x)\right).$$

Entonces:

$$= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \frac{\sigma(wx_i)\left(1-\sigma(wx_i)\right)\frac{\partial}{\partial w_j}\sigma(\mathrm{wx_i})}{h(x_i)}\right.$$
$$\left. - -(1-y_i)\frac{\sigma(wx_i)\left(1-\sigma(wx_i)\right)\frac{\partial}{\partial w_j}\sigma(\mathrm{wx_i})}{1-h(x_i)}\right].$$

Reemplazando $\sigma(\mathrm{wx}) = \mathrm{h}(\mathrm{x})$:

$$= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i \frac{\mathrm{h}(\mathrm{x_i})\left(1-\mathrm{h}(\mathrm{x_i})\right)\frac{\partial}{\partial w_j}\sigma(\mathrm{wx_i})}{h(x_i)} - (1-y_i)\frac{\mathrm{h}(\mathrm{x_i})\left(1-\mathrm{h}(\mathrm{x_i})\right)\frac{\partial}{\partial w_j}\sigma(\mathrm{wx_i})}{1-h(x_i)}\right].$$

Simplificando:

$$= -\frac{1}{n}\sum_{i=1}^{n}\left[y_i\left(1-\mathrm{h}(\mathrm{x_i})\right)x_i - (1-y_i)\mathrm{h}(\mathrm{x_i})\mathrm{x_i}\right].$$

$$= -\frac{1}{n}\sum_{i=1}^{n}[y_i - y_i h(x_i) - h(x_i) + y_i h(x_i)]x_i.$$

$$= -\frac{1}{n}\sum_{i=1}^{n}[y_i - h(x_i)]x_i.$$

Finalmente:

$$\nabla C(w) = \frac{1}{n}\sum_{i=1}^{n}[h(x_i) - y\_i]x_i.$$

Como se puede observar, se llegó a la misma expresión que se tiene en la forma matricial.

$$\nabla C(w) = \frac{1}{n}\sum_{i=1}^{n}[h(x_i) - y_i]x_i = \frac{1}{n}X^T(h(Xw) - y).$$

De esta manera, se comprueba que los cálculos realizados son correctos y las expresiones son equivalentes.

Referencias

Otero, F. (2022). Inteligencia Artificial para la Ciencia de Datos. Regresión Logística. [diapositivas]. Facultad de Ingeniería y Ciencias, Tecnológico de Monterrey.

Matrix Calculus [PDF].

Función de costo de regresión logística y su fórmula de descenso de gradiente. (2020). Programador Clic. https://programmerclick.com/article/61431910230/

Cálculo matricial para el aprendizaje profundo. (2020). Programador Clic. https://programmerclick.com/article/6587987603/

Matrix notation for logistic regression [foro]. (2019). Stack Exchange. https://stats.stackexchange.com/questions/229014/matrix-notation-for-logistic-regression

Gradient descent for logistic regression partial derivative doubt [foro]. (2019). Stack Exchange. https://stats.stackexchange.com/questions/261692/gradient-descent-for-logistic-regression-partial-derivative-doubt