Concentración IA Avanzada para la Ciencia de Datos

Tarea I: Vectorización

Fecha límite de entrega: miércoles 5 de octubre de 2022.

Resuelve los siguientes problemas. La tarea deberá ser entregada en hojas blancas (digitalizada). No se aceptarán tareas en hojas de libreta, o de algún otro tipo de cuaderno. Trabajen con limpieza y hagan procedimientos legibles y claros, argumentando cada paso en su solución. No entreguen la tarea con portada, pero especifiquen bien su nombre, matrícula, grupo, y el número de la tarea que están entregando, escriban estos datos en la parte superior de la primera hoja.

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

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

   Demuestra que la siguiente expresión es equivalente:

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

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

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

Tarea 1: Vectorización

Ileana del Carmen Parra Enríquez A00827284

Nicolás Cárdenas Valdéz A01114959
José Pablo Cruz Ramos A01138740
Andres Piñones Besnier A01570150
Russel Rosique Rodríguez A01283727

October 2022

1 Función de costo

Expansión de $h(x_i)$, recordando que h tiene como parametros w, donde $w \in \mathbb{R}^{k+1}$

$$h\left(x_i\right) = w_0 + w_1 x_1 + w_2 x_2 + \ldots + w_k x_n \tag{1}$$

En su forma matricial:

$$h\left(x_i\right) = \left[\begin{smallmatrix} x_1 \\ x_2 \\ \cdots \\ x_n \end{smallmatrix}\right]\left[\begin{smallmatrix} w_1 \\ w_2 \\ \cdots \\ w_k \end{smallmatrix}\right] = Xw \tag{2}$$

y_i en su forma matricial:

$$y_i = \left[\begin{smallmatrix} y_1 \\ y_2 \\ \cdots \\ y_n \end{smallmatrix}\right] \tag{3}$$

Recordamos que:

$$u \cdot v = u^T v \tag{4}$$

Sustituyendo:

$$C(w) = -\frac{1}{n}(y^T log(h(Xw)) + (1-y)^T log(1-h(Xw))) \tag{5}$$

2 Gradiente de la función de costo

$$\nabla C(w) = \frac{\partial}{\partial w}(C) = \frac{\partial}{\partial w}[-\frac{1}{n}(y^T log(h(Xw)) + (1-y^T) log(1-h(Xw)))] \tag{6}$$

Recordando cálculo matricial:

$$\frac{\partial}{\partial w}(Xw) = X^T \tag{7}$$

Debido a que es una suma se puede derivar por partes:

$$\frac{\partial}{\partial w}[(y^T log(h(Xw))] = y\frac{1}{h}h(1-h)X^T = y(1-h)X^T \tag{8}$$

$$\frac{\partial}{\partial w}[(1-y^T)log(1-h(Xw))] = \frac{1-y}{1-h}(-h(1-h))X^T = -(1-y)hX^T \tag{9}$$

Ahora podemos juntar los resultados:

$$\frac{\partial}{\partial w}(C) = -\frac{1}{n}[y(1-h)X^T - (1-y)hX^T] \tag{10}$$

Simplificando:

$$\nabla C(w) = \frac{1}{n}X^T(h-y) \tag{11}$$

recordando que $h = h(w)$