Modelos avanzados de análisis de datos II

Germán Andrés Sanchez Código 201514733
Héctor Fabián Rodríguez Código 201921382
Ingrid Yulieth Hernández Código 202028041

February 10, 2022

1 Introduction

Comprender la arquitectura y el funcionamiento de las redes neuronales a distintos niveles requiere conceptualizar con rigor matemático el funcionamiento de las neuronas que componen la red. En este ejercicio entendemos cuáles son los elementos que nos permiten calcular el gradiente de la función de pérdida, considerando que la función de entropía cruzada categórica es la capa de salida dada por:

$$-\sum_{j=1}^{N} n_j log(p_j)$$

Y la función de activación es la función softmax dada por:

$$\frac{e^{x_j}}{\sum_{j=1}^{J} e^{x_j}}, j = 1, ..., J$$

Para obtener el gradiente se tiene en cuenta la formula general:

$$\frac{\partial C_0}{\partial w^{(L)}} = \frac{\partial Z_{(L)}}{\partial w^{(L)}} \frac{\partial a_{(L)}}{\partial Z^{(L)}} \frac{\partial C_0}{\partial a^{(L)}}$$

2 Desarrollo de las derivadas parciales

Las consideraciones generales vistas en clase establecen que:

$$a^{(L)} = \sigma(w^{(L)} a^{(L-1)} + b^{(L)}))$$

$$C_0 = (a^{(L)} - y)^2$$

$$Z(L) = w^{(L)} a^{L-1} + b^{(L)}$$

2.1 Derivada de la función de activación

Se considera para este caso particular:

$$a^L = \frac{e^{Z(L)}}{\sum_{j=1}^{J} e^{Z(L)}}$$

$$\frac{\partial a^{(L)}}{\partial Z^{(L)}} = \frac{e^{Z(L)} \sum_{j=1}^{J} e^{Z(L)} - e^{Z(L)} e^{Z(L)}}{(\sum_{j=1}^{J} e^{Z(L)})^2}$$

$$= \frac{e^{Z(L)}((\sum_{j=1}^{J} e^{Z(L)}) - e^{Z(L)})}{(\sum_{j=1}^{J} e^{Z(L)})^2}$$

$$= \frac{e^{Z(L)}}{\sum_{j=1}^{J} e^{Z(L)}}(1 - \frac{e^{Z(L)}}{\sum_{j=1}^{J} e^{Z(L)}})$$

$$a^{(L)}(1 - a^{(L)})$$

2.2 Derivada de la función de costo o pérdida asumiendo una entropía cruzada, respecto a la función de activacion

$$C_0 = -\sum_{j=1}^{J} n_j log(a^L)$$

Se desarrolla la derivada así:

$$\frac{\partial C_0}{\partial a^{(L)}} = \sum_{j=1}^{J} \frac{\partial(n_j log(a^{(L)}))}{\partial a^{(L)}}$$

$$= \sum_{j=1}^{J} n_j \frac{\partial(log(a^{(L)}))}{\partial a^{(L)}}$$

$$= -\sum_{j=1}^{J} n_j \frac{1}{a^{(L)}} \frac{\partial a^{(L)}}{\partial Z^{(L)}}$$

$$= -\frac{n_i}{a_i^L} \frac{\partial a(L)}{\partial Z_i^L} - \sum_{j \neq i}^{J} \frac{n_j}{a^{(L)}} \frac{\partial a^{(L)}}{\partial Z_j^{(L)}}$$

$$= -n_i(1 - a_i^L) + \sum_{j \neq i}^{J} n_j a_i^L$$

$$= -n_i + n_i a_i^{(L)} + \sum_{j \neq i}^{J} n_j a_i^L$$

$$= -n_i + a_i^L$$

Obteniendo la siguiente expresión:

$$= a_i^L - n_i, \forall i \epsilon C$$

Se aclara que

$$\sum_{j=1}^{J} n_j = 1$$

porque i pertenece a la muestra C

2.3 Derivada de la función de costo

Se tiene en cuenta que la definicion de w super l está en función de w y que su derivada corresponde a w super l. Por lo anterior se tiene que, teniendo en cuenta la regla de la cadena, se tiene que:

$$\frac{\partial C_0}{\partial w^{(L)}} = (a_i^L - n_i)a^L(1 - a^{(L)})w^{(L-1)}$$

$$= (a_i^L - n_i)(a^L - (a^{(L)})^2)w^{(L-1)}$$

$$= w^{(L-1))}((-a^{(L)})^3 + (a^{(L)})^2 + n(a^{(L)})^2 - na^{(L)})$$

$$= w^{(L-1))}((-a^{(L)})^3 + (n+1)(a^{(L)})^2 - na^{(L)})$$

Finalmente, se puede reemplazar a super L con la definición dada en la sección 2.1. para obtener el resultado final.