Trabajo practico Nro. 2

INTELIGENCIA ARTIFICIAL

Gonzalo Bengoechea | 38254089 | 14-10-2023

A continuación, se adjuntan las imágenes del desarrollo de los distintos puntos solicitados en el trabajo practico, los cuales pueden ser buscados de forma más específica en el siguiente índice:

Índice

1.

A.

1)

a) Para obtener la actualización por descenso para $W^1_{1,2}$, primero es necesario calcular la derivada de la función de costo con respecto a dicho peso

$$L = \frac{1}{m}\sum_{i=1}^{m}\left(o^{(i)} - y^{(i)}\right)$$

Derivando, nos queda:

$$\frac{dL}{dw^1_{1,2}} = \frac{dL}{do(i)} \cdot \frac{do(i)}{dz^{(2)}} \cdot \frac{dz^{(2)}}{d\alpha^{(1)}} \cdot \frac{d\alpha^{(1)}}{dz^{(1)}} \cdot \frac{dz^{(1)}}{dw^1_{1,2}}$$

Reemplazamos y Desarrollamos

$$\frac{dL}{dw^1_{1,2}} = \frac{2}{m}\sum_{i=1}^{m}\left(o^{(i)} - y^{(i)}\right) \cdot \sigma'(z^{(2)}) \cdot W^2 \cdot \sigma'(z^{(1)}) \cdot x_i$$

$$\frac{dL}{dw^1_{1,2}} = \frac{2}{m}\sum_{i=1}^{m}\left(o^{(i)} - y^{(i)}\right) \cdot \sigma(z^{(2)}) \cdot \left(1 - \sigma(z^{(2)})\right) \cdot W^2 \cdot \sigma(z^{(1)}) \cdot \left(1 - \sigma(z^{(1)})\right) \cdot x_i$$

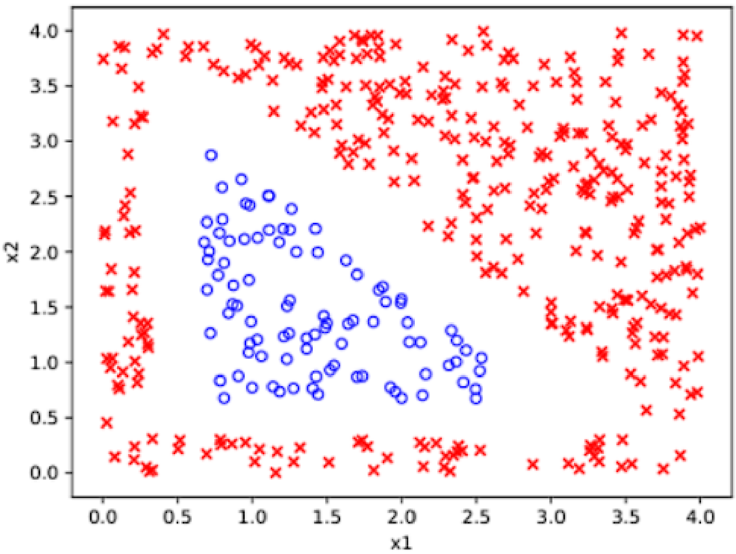
$$\frac{dL}{dw^1_{1,2}} = \frac{2}{m}\sum_{i=1}^{m}\left(o^{(i)} - y^{(i)}\right) \cdot \sigma(w^2_j \cdot \alpha^{(1)} + W^2_0) \cdot \left(1 - \sigma(w^2_j \cdot \alpha^{(1)} + W^2_0)\right) \cdot W^2 \cdot \sigma(w^1_{ij} \cdot x_i + w^1_{0j}) \cdot$$
$$\left(1 - \sigma(w^1_{ij} \cdot x_i + w^1_{0j})\right) x_i$$

$$\frac{dL}{dw^1_{1,2}} = \frac{2}{m}\sum_{i=1}^{m}\left(o^{(i)} - y^{(i)}\right) \cdot \sigma\left(w^2_j \cdot (w^1_{ij} x_i + W^1_{0j}) + w^2_0\right) \cdot \left(1 - \sigma\left(w^2_j \cdot (w^1_{ij} \cdot x_i + W^1_{0j}) + w^2_0\right)\right)$$
$$w^2 \cdot \sigma(w^1_{ij} x_i + w^1_{0j})\left(1 - \sigma(w^1_{ij} x_i + w^1_{0j})\right) x_i$$

Entonces la ecuación de la act por descenso por gradiente es

$$\boxed{W^1_{12\,\text{nuevo}} = W^1_{12\,\text{viejo}} - \alpha \frac{dL}{dW^1_{1,2}}}$$

B.
Teniendo en cuenta el grafico de los datos dado por la consigna:



Es lógico suponer que, al emplear tres neuronas con activación lineal, es factible alcanzar un 100% de precisión en la clasificación de los datos. A modo de ilustración, consideremos estas tres funciones lineales:

- 0y = -1x + 0.5 → 0 = -x + 0.5
- 1y = 0x + 0.5 → 0 = -y + 0.5
- 1y = -1x + 4 → 0 = -x -y + 4

(b) Podemos determinar que un ejemplo de pesos posibles para este problema.

Neurona h_1: $w_1^1 = (-1,0)$ $w_{0,1}^1 = 0.5$

Neurona h_2: $w_2^1 (0,-1)$ $w_{0,2}^1 = 0.5$

Neurona h_3 = $w_3^1 (1,1)$ $w_{0,3}^1 = -4$

Para elegir los pesos y sesgos de la neurona de salida, se utilizó el código del punto 1d, en este código se encuentra una red neuronal simple a la cual se le implementa la función escalonada que se encuentra en la consigna del punto. Dado que los pesos y sesgos de la capa oculta son correctos, se seleccionan los siguientes valores que funcionan para los datos.

$$w^2 = (1,1,4) \qquad w_0^2 = -1$$

C. Cuando la función de activación de las tres neuronas ocultas es lineal, se vuelve improbable alcanzar un 100% de precisión con estos datos. En esta situación, todo se reduce esencialmente a una regresión lineal, resultando únicamente en un límite de decisión lineal.

Justificación de la limitación:
La suma de las funciones lineales proporcionadas, al integrarse en el modelo neuronal, resulta en un sistema que es inherentemente lineal. Dado que las funciones lineales son incapaces de capturar la complejidad de ciertos patrones no lineales en los datos, la capacidad del modelo para realizar clasificaciones precisas se ve limitada. En consecuencia, la obtención de un 100% de precisión se vuelve poco realista, ya que las limitaciones lineales del modelo pueden no ser suficientes para abordar la diversidad y complejidad de los datos.

2. Se adjunta el código de todos los ítems del punto.