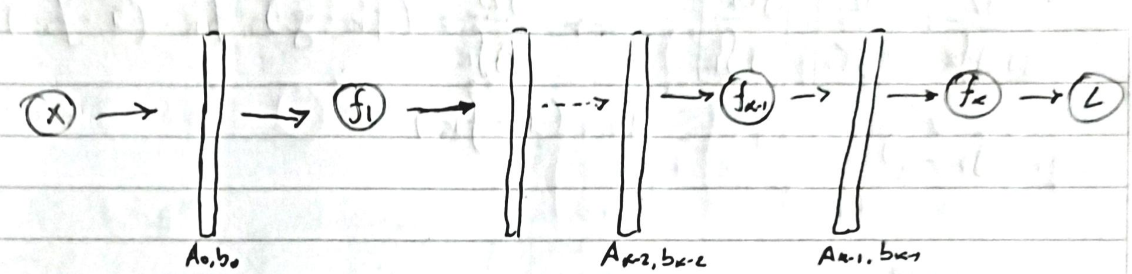
TP FINAL AMIA

EJERCICIO TEÓRICO

BACKGROUND



Partiendo de la siguiente definición:

- $f_0 := \bar{x}$

- $f_i := \sigma_i (A_{i-1} \cdot f_{i-1} + b_{i-1}), \; i = 1, \ldots, k$

- $\dfrac{\partial L}{\partial \theta_i} = \dfrac{\partial L}{\partial f_k} \cdot \dfrac{\partial f_k}{\partial f_{k-1}} \cdots \dfrac{\partial f_{i+2}}{\partial f_{i+1}} \cdot \dfrac{\partial f_{i+1}}{\partial \theta_i}$

Para la siguiente función de activación $\sigma_z = \dfrac{1}{1 + e^{-z}}$
y la siguiente función de costo $L = \frac{1}{2} (f_k(\theta, x) - y)^2$

Se obtienen los siguientes gradientes para cada capa:

CAPA K:

$$\frac{\partial L}{\partial \theta_{k-1}} = \frac{\partial L}{\partial f_k} \cdot \frac{\partial f_k}{\partial \theta_{k-1}}$$ donde:

- $\dfrac{\partial L}{\partial f_k}$ es como varía el error con respecto a la función f_k. Como la función error está compuesta por la función de activación σ_k, entonces hay que aplicar la regla de la cadena:

$$\frac{\partial L}{\partial f_K} = \underbrace{\boxed{\frac{\partial L}{\partial \sigma_K}}}_{(f_K - y)} \cdot \underbrace{\boxed{\frac{\partial \sigma_K}{\partial f_K}}}_{f_K \cdot (1 - f_K)} \Longrightarrow \frac{\partial L}{\partial f_K} = (f_K - y) \cdot f_K \cdot (1 - f_K)$$

y donde:

- $\frac{\partial f_K}{\partial \Theta_{K-1}}$ es como varía ($\Theta = \{A_0, b_0, \ldots, A_{K-1}, b_{K-1}\}$) la función con respecto a los parámetros (pesos). En este caso $\Theta = \{\underset{A_0}{\omega^1}, \underset{b_0}{b^1}, \underset{A_1}{\omega^2}, \underset{b_1}{b^2}\}$

$$\Rightarrow \frac{\partial f_K}{\partial \Theta_{K-1}} = \left[\frac{\partial f_K}{\partial A_{K-1}}, \frac{\partial f_K}{\partial b_{K-1}}\right], \text{ como } \frac{\partial f_K}{\partial A_{K-1}} = f_{K-1} \text{ y}$$

$$\frac{\partial f_K}{\partial b_{K-1}} = 1 \Longrightarrow \frac{\partial f}{\partial \Theta_{K-1}} = \left[f_{K-1}, 1\right]$$

Por lo tanto, si llamamos $\delta = (f_K - y) \cdot f_K \cdot (1 - f_K)$ entonces el gradiente de la capa K es

$$\frac{\partial L}{\partial \Theta_{K-1}} = \boxed{\nabla_K L = \left[\delta^K \cdot f_{K-1}, \delta^K\right]}$$

CAPA $K-1$:

Para esta capa es necesario calcular únicamente como varía la entrada con respecto a la capa anterior. Partiendo de:

$$\frac{\partial L}{\partial \theta_{k-2}} = \frac{\partial L}{\partial f_k} \cdot \frac{\partial f_k}{\partial f_{k-1}} \cdot \frac{\partial f_{k-1}}{\partial \theta_{k-2}}$$, tenemos que calcular $\frac{\partial f_k}{\partial f_{k-1}}$ ya que

$$\frac{\partial L}{\partial f_k} = \delta^k \quad y \quad \frac{\partial f_{k-1}}{\partial \theta_{k-2}} = \left[f_{k-2} , 1 \right]$$

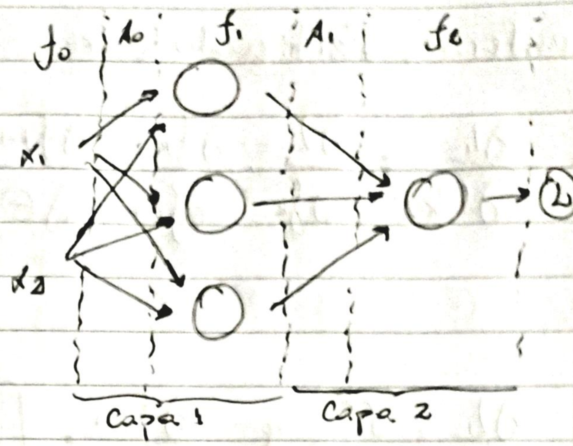
Entonces, aplicando la regla de la cadena (ya que f es una composición de la función de activación)

$$\Rightarrow \frac{\partial f_k}{\partial f_{k-1}} = \underbrace{\boxed{\frac{\partial f_k}{\partial \sigma_{k-1}}}}_{A_{k-1}} \cdot \underbrace{\boxed{\frac{\partial \sigma_{k-1}}{\partial f_{k-1}}}}_{f_{k-1} \cdot (1 - f_{k-1})}$$

Si llamamos δ^{k-1} a $\delta^k \cdot f_{k-1} \cdot (1 - f_{k-1}) \cdot A_{k-1}$ entonces tenemos que el gradiente de la capa k-1 es

$$\frac{\partial C}{\partial \theta_{k-2}} = \left\{ \nabla_{k-1} C = \left[\delta^{k-1} \cdot f_{k-2} , \delta^{k-1} \right] \right\}$$

EJERCICIO



- $f_0 = \begin{pmatrix} 1.8 \\ -3.4 \end{pmatrix}$

- $f_1 = \sigma(w^1 \cdot f_0 + b^1)$

$$\Rightarrow f_1 = \begin{pmatrix} 0.1 & -0.5 \\ -0.3 & -0.9 \\ 0.8 & 0.02 \end{pmatrix} \begin{pmatrix} 1.8 \\ -3.4 \end{pmatrix} + \begin{pmatrix} 0.1 \\ 0.5 \\ 0.8 \end{pmatrix}$$

$$\Rightarrow f_1 = \sigma \begin{pmatrix} 1.98 \\ 3.02 \\ 2.17 \end{pmatrix} \Rightarrow \boxed{f_1 = \begin{pmatrix} 0.88 \\ 0.95 \\ 0.90 \end{pmatrix}}$$

- $f_2 = \sigma(w^2 \cdot f_1 + b^2) \Rightarrow f_2 = \begin{pmatrix} -0.4 \\ 0.2 \\ -0.5 \end{pmatrix}^T \begin{pmatrix} 0.88 \\ 0.95 \\ 0.90 \end{pmatrix} + 0.7 \Rightarrow f_2 = \sigma(0.09)$

$\Rightarrow \boxed{f_2 = 0.52}$ Error: $\frac{1}{2}(0.52 - 5)^2 = 10.03$

CAPA 2:

$$\nabla_2 L = (\delta^2 \cdot f_1, \delta^2) \text{ donde } \delta^2 = (\overset{0.52}{f_2} - \overset{5}{y}) \cdot \overset{0.52}{f_2} \cdot (1 - \overset{0.52}{f_2}) \Rightarrow \delta^2 = -1.11$$

$$\therefore \nabla_2 L = \left[(-1.11) \begin{pmatrix} 0.88 \\ 0.95 \\ 0.90 \end{pmatrix}, -1.11 \right] \Rightarrow \boxed{\nabla_2 L = \left[\begin{pmatrix} -0.98 \\ -1.05 \\ -1 \end{pmatrix}, -1.11 \right]}$$

CAPA 1:

$$\nabla_1 L = (\delta^1 \cdot f_0, \delta^1) \text{ donde } \delta^1 = \delta^2 \cdot f_1 \cdot (1 - f_1) \cdot A_1 \;/\; A_1 = w^2$$

$$\text{Como } \delta^1 = (-1.11) \begin{pmatrix} 0.88 \\ 0.95 \\ 0.90 \end{pmatrix} \left[1 - \begin{pmatrix} 0.88 \\ 0.95 \\ 0.90 \end{pmatrix} \right] \cdot \begin{pmatrix} -0.4 \\ 0.2 \\ -0.5 \end{pmatrix} = \begin{pmatrix} 0.11 \\ -0.05 \\ 0.13 \end{pmatrix}$$

$$\therefore \nabla_1 L = \left[\begin{pmatrix} 0.11 \\ -0.05 \\ 0.13 \end{pmatrix} \begin{pmatrix} 1.8 \\ -3.4 \end{pmatrix}^T, \begin{pmatrix} 0.11 \\ -0.05 \\ 0.13 \end{pmatrix} \right] \Rightarrow \boxed{\nabla_1 L = \left[\begin{pmatrix} 0.20 & -0.37 \\ -0.09 & 0.17 \\ 0.23 & 0.44 \end{pmatrix}, \begin{pmatrix} 0.11 \\ -0.05 \\ 0.13 \end{pmatrix} \right]}$$