

I. Parte teórica

Considerando la figura 1 realice los siguientes ejercicios de manera ordenada.

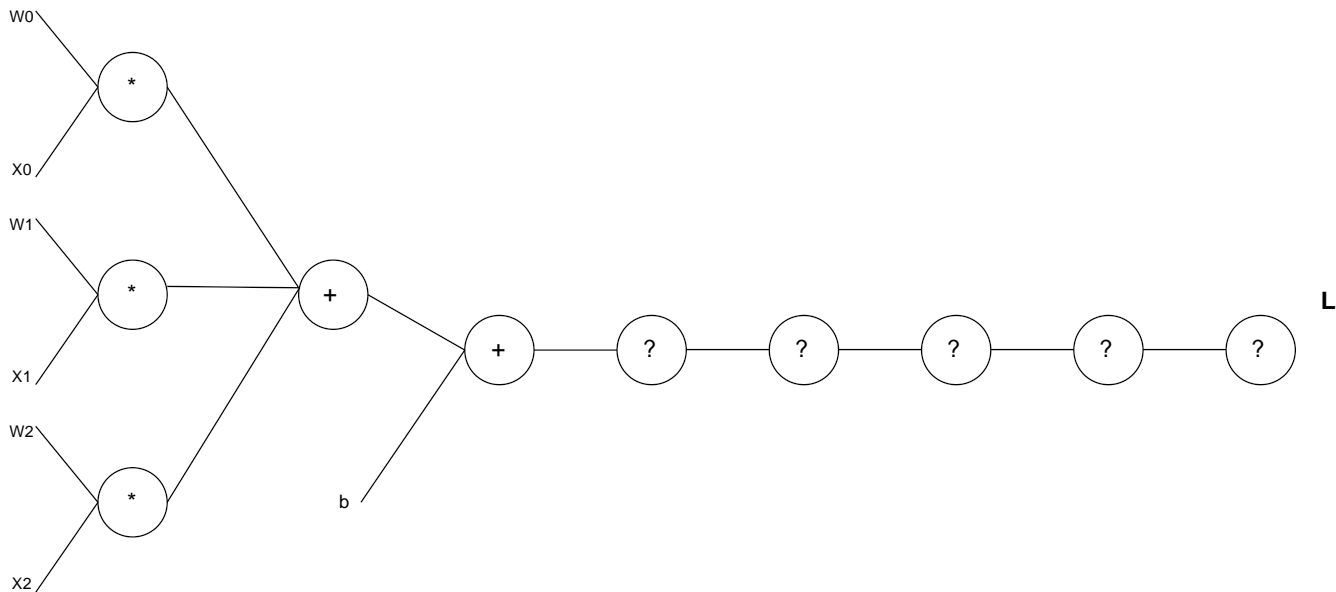
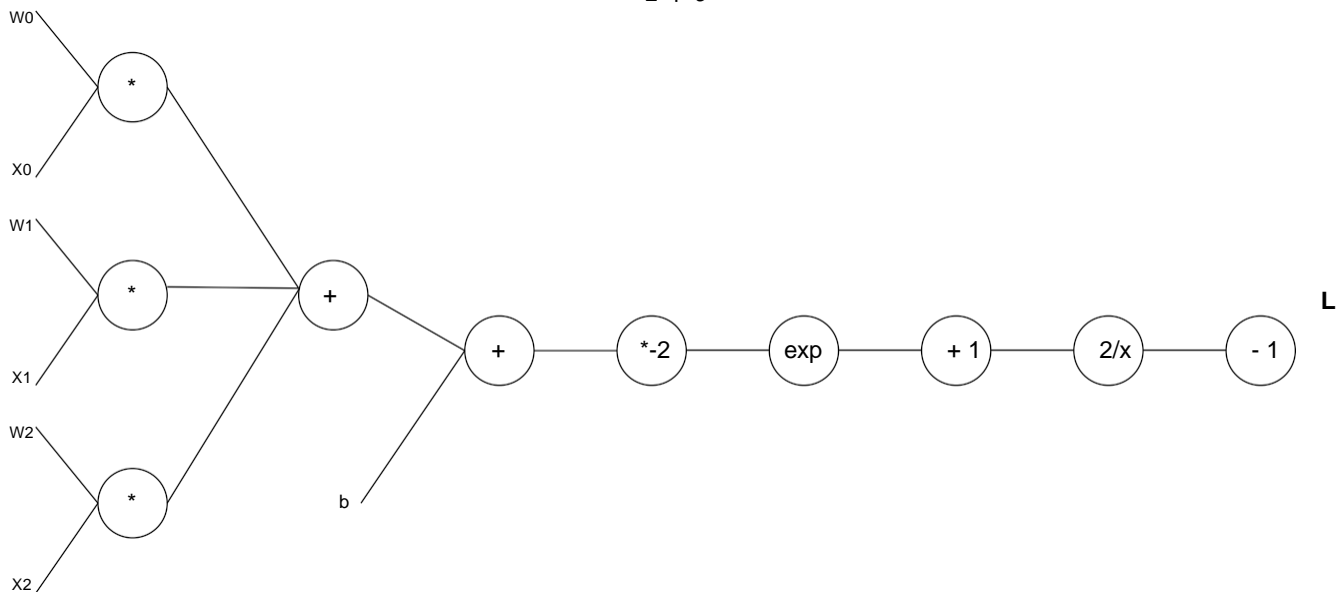


Figura 1: Ejercicio para la parte teórica

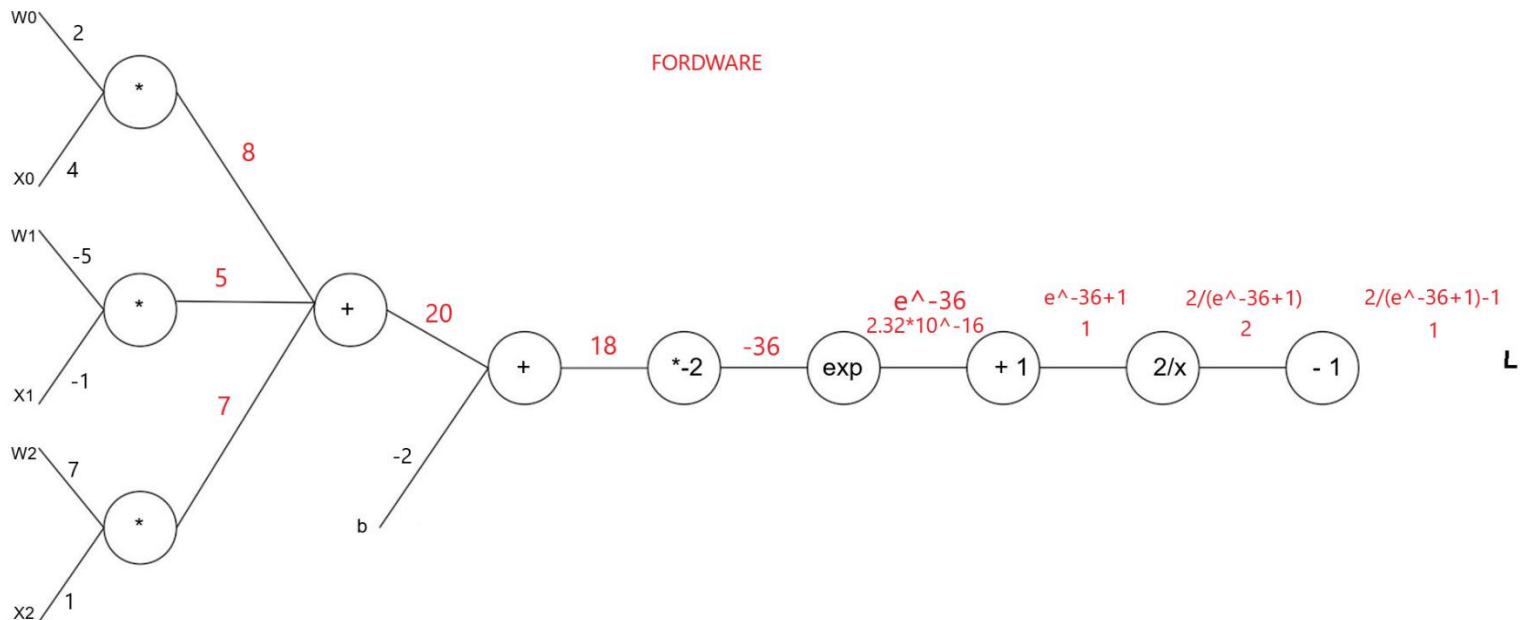
Ejercicio 1

1. Complete el grafo de la figura 1 utilizando la función de activación \tanh . Debe colocar en cada nodo con "?" la operación correspondiente hasta llegar a la función de costo L. Asegúrese de que el grafo sea legible de tal manera que sea similar al que se realizó en clase. (2 puntos)

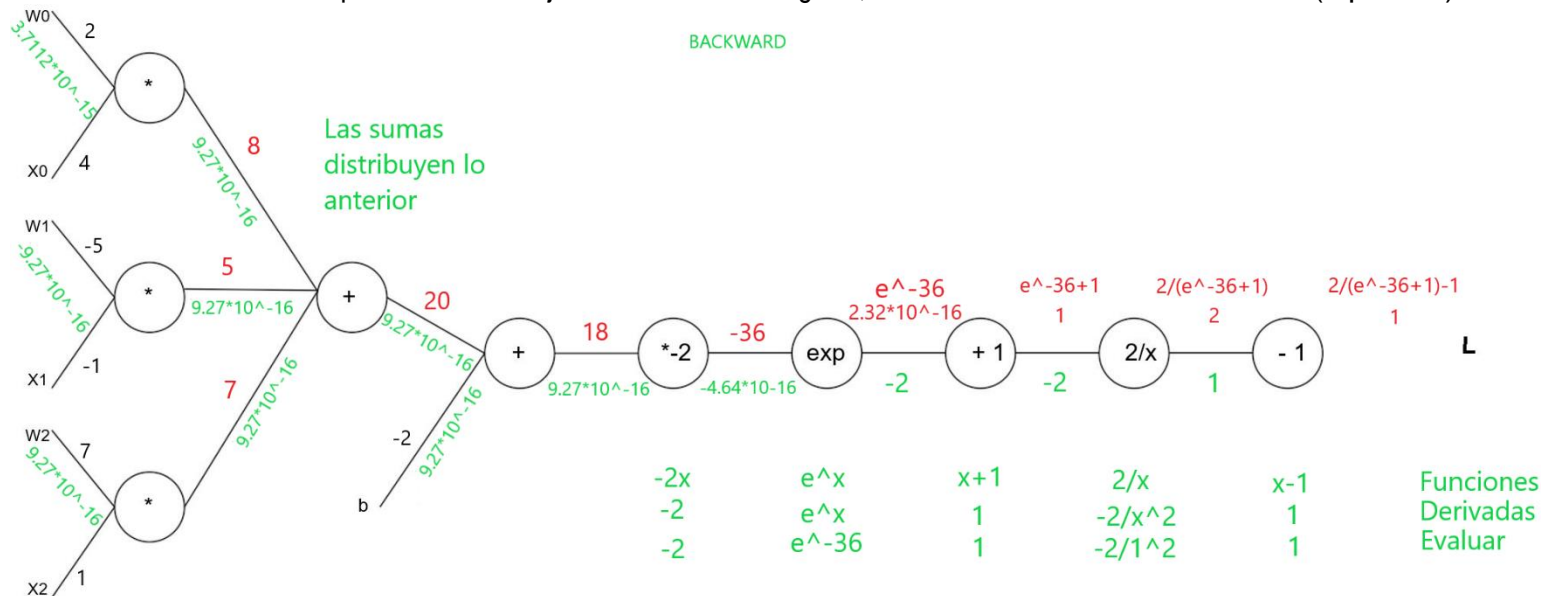
$$\tanh = \frac{2}{1 + e^{-2x}} - 1$$



2. Realice el forward (pasada hacia adelante), colocando los valores correspondientes encima de cada línea del grafo. Para este apartado utilice los siguientes valores para los pesos, entrada y el bias: $\mathbf{W} = [2, -5, 7]$, $\mathbf{X} = [4, -1, 1]$, $b = -2$ (5 puntos)



3. Basado en el resultado del punto anterior, realice el backward (pasada hacia atrás), anotando los valores correspondientes debajo de cada línea del grafo, similar a como se realizó en clase. (5 puntos)



4. Indique de manera clara, los nuevos valores de \mathbf{W} y b . (2 puntos)

$$\mathbf{W} = [3.7112 \times 10^{-15}, -9.27 \times 10^{-16}, 9.27 \times 10^{-16}], b = 9.27 \times 10^{-16}$$

Ejercicio 2

Explique por qué una red fully connected puede no converger si se aumenta la profundidad en las capas ocultas. (3 puntos)

Desvanecimiento de gradientes: A medida que se propaga el error hacia atrás a través de las capas durante el entrenamiento, los gradientes pueden disminuir o aumentar exponencialmente en magnitud. Si los gradientes se vuelven muy pequeños, pueden hacer que los pesos en las capas anteriores no se actualicen significativamente.

Sobreajuste (overfitting): Una red neuronal más profunda tiene una mayor capacidad para memorizar los datos de entrenamiento. Si la red es demasiado profunda en relación con el tamaño del conjunto de datos de entrenamiento, existe un mayor riesgo de sobreajuste. El modelo puede aprender patrones y ruido específicos de los datos de entrenamiento que no se generalizan bien a nuevos datos, lo que resulta en un rendimiento deficiente en la fase de prueba.