Universidad Técnica Federico Santa María

Departamento de Informática

INF-393 Redes Neuronales Artificiales
Control 1

1. (25 pts.) Considere una red con funciones de activación clásicas (*tanh* y salida *softmax*). ¿Por qué los pesos suelen inicializarse con valores pequeños y aleatorios?

   La aleatoriedad permite minimizar el problema de la identificabilidad de las redes, mientras que los valores pequeños son para evitar el desvanecimiento del gradiente, ya que *tanh* y salida *softmax* tienen gradientes cercanos a cero cuando $\mathbf{x}$ tiende a infinito (o menos infinito). Además el gradiente es más grande cuando estas funciones de activaciones están cercanas a cero.

2. (25 pts.) Explique la diferencia y el objetivo de las técnicas que hemos denominado *Momentum* y *Momentum de Nesterov* para entrenamiento de redes neuronales artificiales.

   El método de Momentum estándar primero calcula el gradiente en la ubicación actual y luego toma un gran salto en la dirección del gradiente acumulado actualizado. En cambio, el gradiente Nesterov primero da un gran salto en la dirección del anterior gradiente acumulado y luego mide el gradiente donde uno termina (considerando no solo la ubicación sino que también la velocidad) para hacer la corrección.

3. (50 pts.) Considere el grafo de computación definido en la Figura 1, donde un arco de v_1 a v_2 indican una relación de dependencia directa, es decir, "v_2 requiere v_1". Escriba expresiones recursivas explícitas para las derivadas de s con respecto a las variables de cada capa que utilice las derivadas de s con respecto a las variables de la capa siguiente. Si una derivada es cero, debe omitirla explícitamente, es decir debe tener en cuenta la conectividad del grafo en cuestión.

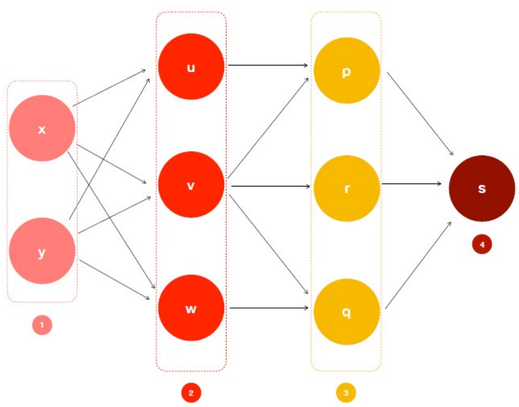

Figure 1: Grafo para la pregunta 3.

   Las derivadas de la capa 3, $\frac{\partial s}{\partial p}$, $\frac{\partial s}{\partial r}$, $\frac{\partial s}{\partial q}$ se obtienen directamente. Las derivadas de la capa 2, se calculan recursivamente como:

$$\frac{\partial s}{\partial u} = \frac{\partial s}{\partial p}\frac{\partial p}{\partial u}$$
$$\frac{\partial s}{\partial v} = \frac{\partial s}{\partial p}\frac{\partial p}{\partial v} + \frac{\partial s}{\partial q}\frac{\partial q}{\partial v} + \frac{\partial s}{\partial r}\frac{\partial r}{\partial v}$$
$$\frac{\partial s}{\partial w} = \frac{\partial s}{\partial q}\frac{\partial q}{\partial w}$$

Las derivadas de la capa 1, se calculan recursivamente como

$$\begin{aligned}\frac{\partial s}{\partial x} &= \frac{\partial s}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial s}{\partial v}\frac{\partial v}{\partial x} + \frac{\partial s}{\partial w}\frac{\partial w}{\partial x} \\ \frac{\partial s}{\partial y} &= \frac{\partial s}{\partial u}\frac{\partial u}{\partial y} + \frac{\partial s}{\partial v}\frac{\partial v}{\partial y} + \frac{\partial s}{\partial w}\frac{\partial w}{\partial y}\end{aligned}$$