

INF-477 Redes Neuronales Artificiales

Quiz 1. I-2019.

1. (10%) Diseñe una red neuronal con a lo más $3(d-1)$ neuronas que compute la función de paridad de d variables booleanas.
2. (10%) Diseñe una red neuronal de 1 sola capa que compute la función de paridad de d variables booleanas. ¿Es posible construir esta red usando a lo más $O(d)$ neuronas?
3. (10%) Explique qué se entiende por *epoch*, *tamaño de batch* y *tasa de aprendizaje* en redes neuronales. ¿Es mejor usar un valor grande o pequeño de este último hyper-parámetro?
4. (10%) Explique para qué tipo de problemas entrenaría una red con la función de pérdida (loss) *cross-entropy* y porqué. Explique también, para qué tipo de problemas entrenaría una red con capa de salida *soft-max* y porqué.
5. (15%) Considere el digrafo de computación definido en la Figura 1. Suponga que el estado de un nodo v se obtiene multiplicando el estado de los nodos incidentes $\text{Pa}(v)$ por los correspondientes pesos de conexión, y sumando los resultados obtenidos. Encuentre expresiones para $\partial s/\partial x$ y $\partial s/\partial y$ que dependan sólo de x e y . ¿Cuántas sumas y productos se requieren para evaluar estas expresiones? Encuentre ahora ecuaciones que permitan calcular $\partial s/\partial x$ y $\partial s/\partial y$ a partir de $\partial s/\partial t$, $\partial s/\partial u$ y $\partial s/\partial v$. Construya expresiones similares para $\partial s/\partial t$, $\partial s/\partial u$, $\partial s/\partial v$ a partir de $\partial s/\partial p$, $\partial s/\partial q$, $\partial s/\partial r$. ¿Cuántas sumas y productos se requieren para evaluar todas estas expresiones? Cómo cambiaría el resultado si las “capas” del grafo fuesen completamente conectadas.

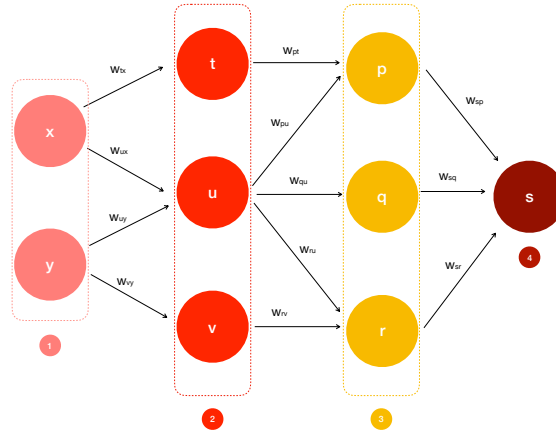


Figure 1: Grafo para la pregunta 5.

6. (10%) Demuestre que la función de pérdida (loss) $L : [0, 1] \times [0, 1] \rightarrow [0, 1]$ denominada “cross-entropy” $L(y, f) = -y \log f - (1 - y) \log(1 - f)$, $y, f \in [0, 1]$ es una función convexa de f . Suponga ahora que el valor de f se obtiene calculando la salida de una red neuronal con parámetros θ sobre un determinado patrón de entrada x . ¿Es L una función convexa de θ ?
7. (10%) Explique brevemente la diferencia entre las técnicas que hemos denominado *Adagrad* y *RMS-Prop* para entrenamiento de redes neuronales artificiales.

8. (10%) Considere una red neuronal feed-forward estándar con L capas lineales (función de activación lineal en cada capa). Sea $E = (f(x) - y)^2$ donde $y \in \mathbb{R}$ y $f(x) \in \mathbb{R}$ es la salida de la red. Demuestre que la magnitud de $\partial E / \partial a^\ell$ “se desvanece” (decrece) exponencialmente rápido en ℓ . Explique la relevancia de este resultado.
9. (15%) Suponga que se entrena una red neuronal $f(x; \theta)$ con cierta función objetivo $J(\theta)$ diferenciable y “suave”, en el sentido de que satisface

$$\|\nabla J(\theta_1) - \nabla J(\theta_2)\| \leq L \|\theta_1 - \theta_2\|. \quad (1)$$

- (1). Demuestre que si se emplea SGD con tasa de aprendizaje dinámica que satisface las denominadas “ecuaciones de Robbins-Monro”, entonces se puede garantizar la convergencia a un punto estacionario “en valor esperado”.

RNA+CVV L^AT_EX