

Universidad de Buenos Aires
Facultad de Ingeniería

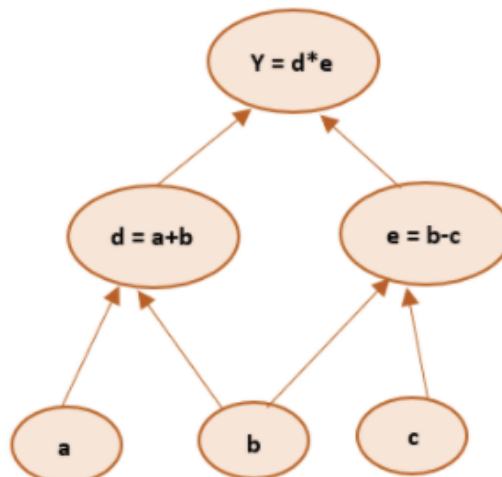
Postgrado de Inteligencia Artificial
Deep Learning - Jun 2021
Ejercicios de mitad de curso

1. Stochastic Gradient Descent

- a. Implementar en Python un algoritmo basado en Stochastic Gradient Descent para encontrar numéricamente el mínimo de $y = (x - 2)^2 + 3$. ¿Cuál es el mínimo valor que encontró el algoritmo y para qué valor de x ? ¿En qué momento detuviste la búsqueda y por qué? ¿Cómo inicializamos los valores random del algoritmo y cómo afectan la convergencia? ¿Cuáles son los hiper parámetros del algoritmo?
- b. Agregar al algoritmo del punto (a) un momento de primer orden. ¿Cuáles son los hiper parámetros del algoritmo?
- c. Agregar al algoritmo del punto (b) un momento de segundo orden. ¿Cuáles son los hiper parámetros del algoritmo?
- d. Definir los mismos valores random iniciales para los tres optimizadores (a), (b) y (c). Para todos los casos graficar los gradientes en función de las iteraciones. ¿Qué algoritmo converge más rápido?

2. Backpropagation y regla de la cadena

- a. Automatic Differentiation: para el grafo de cómputo que se muestra en la siguiente imagen, con $a=1$, $b=2$ y $c=3$, calcular el valor final de Y y la derivada de Y respecto a cada una de las entradas (a , b y c). Para hacer el cálculo, primero realizar el paso forward y luego utilizar backpropagation con la regla de la cadena.



- b. Numerical Differentiation: para el grafo de cómputo que se muestra en la siguiente imagen, con $a=1$, $b=2$ y $c=3$, utilizar la siguiente expresión y un valor de h pequeño para calcular las derivadas respecto de cada entrada. Comparar los resultados con los obtenidos en el punto (a)

$$\frac{\partial Y}{\partial a} = \lim_{h \rightarrow 0} \frac{Y(a+h, b, c) - Y(a, b, c)}{h}$$

3. Deep Learning y Regularización

- a. Crear un dataset sintético con 10,000 muestras a partir de la siguiente expresión, donde N es una normal de media cero y desvío 0.2. Utilizar 80% de los datos para entrenamiento y 20% de los datos para testing.

$$y = (x - 2)^2 + 3 + N$$

- b. Suponer que se tiene un modelo, tal que la relación entre la salida y la entrada está dada por la ecuación presentada a continuación. Suponiendo que se va a utilizar mini-batch SGD para encontrar los valores óptimos de los pesos w_1 , w_2 y w_3 tal que minimizan el ECM (Error Cuadrático Medio) (el ECM es la función de costo para este problema), ¿cuál sería la regla de actualización de cada parámetro?. Implementar el algoritmo en Python (usar mini-batch), encontrar los valores óptimos para w_1 , w_2 y w_3 y reportar ECM sobre dataset de entrenamiento y dataset de testing.

$$\hat{y} = w_1 x^2 + w_2 x + w_3$$

- c. Ahora se le agrega regularización L_2 a la función de costo, ¿cómo cambia la regla de actualización de cada parámetro?. Implementar el algoritmo en Python (usar mini-batch), encontrar los valores óptimos para w_1 , w_2 y w_3 y reportar ECM sobre dataset de entrenamiento y dataset de testing. Seleccionar manualmente los parámetros para el learning rate y para el coeficiente de regularización.
- d. En lugar de seleccionar manualmente el learning rate y el coeficiente de regularización, utiliza el 10% del dataset de entrenamiento como dataset de validación. Hacer una búsqueda (grid search) de los hiper parámetros que mejor resultado te dan para el ECM sobre el dataset de validación. Luego, utiliza los mejores hiper parámetros para reportar las métricas sobre el dataset de testing.
- e. En los puntos (b), (c) y (d) utilizaste Python para hacer la implementación y el modelo ya estaba establecido. En este punto se pide hacer una implementación con Deep Learning en PyTorch. Las redes neuronales, si tienen la suficiente cantidad de layers y se están utilizando las funciones de activación adecuadas, son capaces de representar cualquier función para mapear la entrada a la salida. Proponer una arquitectura, agregar

regularización mediante la técnica dropout, entrenar la red neuronal, buscar los mejores hiper parámetros y presentar el ECM sobre el dataset de testing. Comparar los resultados obtenidos en (e) con los puntos anteriores.

- f. Para el punto (e), ¿qué función de activación seleccionaste y por qué? Para la función que hayas seleccionado, explica sus características, su forward y su backward.