
CIF 8458 Ing. Conocimiento
Pauta Prueba Integral II
Lunes 18 de Diciembre de 2023

1 Preguntas

1. Considerando la red de la Figura 1, donde la capa escondida tiene activaciones ReLU y la capa de salida activaciones softmax ($\text{Softmax}(x)_i = \frac{e^{x_i}}{\sum_{j=1}^n e^{x_j}}$).

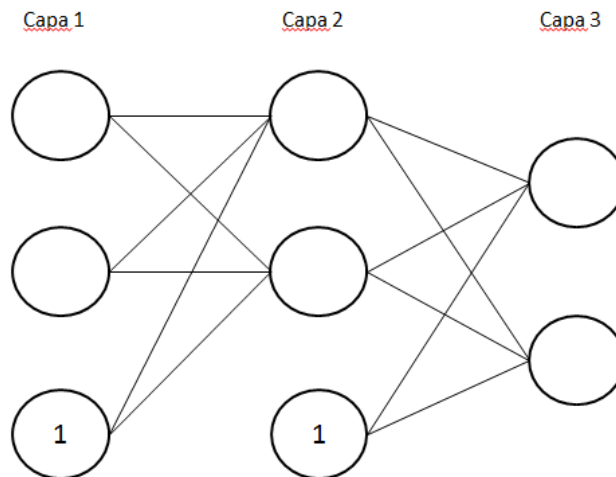


Figura 1: Esquema de la Red

- (a) (0,5 pts) Escriba el código keras para definir esta red.

```
from keras.models import Sequential
from keras.layers.core import Dense, Activation

model = Sequential()
model.add(Dense(2, input_dim=2, init='uniform'))
model.add(Activation('relu'))
model.add(Dense(1))
model.add(Activation('softmax'))
```

R: Si los pesos de la red son:

$$W^{(1)} = \begin{bmatrix} -0.4 & 0.2 & 0.1 \\ -0.2 & 0.4 & -0.1 \end{bmatrix}$$

$$W^{(2)} = \begin{bmatrix} 0.1 & -0.2 & 0.1 \\ 0.4 & -0.1 & 0.1 \end{bmatrix}$$

Considerando lo anterior responda las preguntas b) y c):

- (b) (1 pto) Dada la entrada $x_1 = (1, 0)$. Calcule la salida de la red usando forward pass.

$$a^{(1)} = \text{ReLU}(-0.2, 0.2) = (0, 0.2)$$

Luego:

$$a^{(2)} = \text{softmax}(0.1, 0.4099668) = (1.10517092, 1.50676776)/(1.10517092+1.50676776) = (0.42312284, 0.57687716)$$

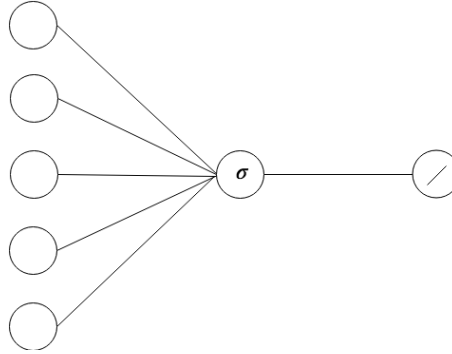
- (c) (0,5 pts) Calcule el error rate si x_1 es de la clase 2.

R: Predicción clase 2, target clase 2, es decir la predicción es correcta, y el error rate es 0.

2. (0,5 pts) ¿Cuál es la diferencia entre parámetros e hiperparámetros?

R: Los valores de los parámetros del modelo se determinan entrenando al modelo, es decir, del error observado en los ejemplos de entrenamiento. Por ejemplo: Pesos y sesgos de la red. Mientras que el valor de los Hiper-parámetros del modelo: no pueden ser determinados a partir de los datos de entrenamiento. Se pueden estimar usando técnicas de validación. Por ejemplo: Número de capas y número de neuronas en cada capa.

3. (0,5 pts) Dibuje la red neuronal equivalente para realizar una regresión logística considerando un vector de entrada de 5 atributos.



sigma es la función de activación sigmoidea.

4. (1 pts) En redes neuronales profundas. ¿El algoritmo de gradiente descendente usualmente converge al óptimo? Fundamente su respuesta positiva o negativa.

R: No, porque la función de costo es no convexa, de hecho suele tener muchos puntos silla, ya que existen muchas combinaciones de pesos que llevan al mismo valor de la función de costo.

5. (0,5 pts) ¿Cuál es la diferencia entre aprendizaje supervisado y aprendizaje no supervisado?

R: En el aprendizaje supervisado se aprende el mapeo entre la entrada x y el target y . En cambio, en el aprendizaje no supervisado, el algoritmo no usa los targets, sino que se centra en usar medidas de distancia para agrupar vectores de entradas "similares".

6. Se desea agrupar 5 esferas, cuyos datos se encuentran en el dataframe df en 2 clases. A cada esfera se le mide su radio en centímetros y su peso en kilogramos:

Radio	Peso
16,3173	0,7201
12,6377	0,7409
12,8786	0,6224
15,0350	0,8657
10,0774	0,8202

- (a) (0,5 pts) Escriba el código python para graficar los datos.

```
import matplotlib.pyplot as plt

plt.scatter(df['Radio'], df['Peso'], color='blue', marker='o')

plt.xlabel('Radio')
plt.ylabel('Peso')

plt.show()
```

- (b) (1 pts) Usando los 2 primeros datos como los centroides iniciales, etiquete cada dato con su clúster asociado usando la distancia euclídea.

dist a C1	dist a C2	C
0	3,679658789	1
3,679658789	0	2
3,440087641	0,268467987	2
1,290539674	2,400546257	1
6,240702847	2,561527782	2

(c) (0,5 pts) Usando los nuevos cluster obtenidos en a) recalcule los centroides.

Centroide 1: $((16.3173+15.0350)/2, (0.7201+0.8657)/2) = (15.67615, 0.7929)$

Centroide 2: $((12.6377 + 12.8786 + 10.0774)/3, (0.7409 + 0.6224 + 0.8202)/3) = (11.8646, 0.7278)$.