

## Examen 2

1. Si  $l$  está en el rango  $[1, L]$ , seleccione aquellos que son hiperparametros de una RNP.
  - Ritmo de aprendizaje
  - Valor de  $L$
  - Numero de iteraciones del descenso de gradiente.
  - Valores  $n^{[l]}$
  - Funciones  $g^{[l]}$
2. El valor de la derivada para valores positivos de la funcion de activación ReLu es igual a:
  - 1
3. Sobre la propagacion hacia atras en una red neuronal superficial es cierto que:
  - Permite encontrar los valores de las derivadas para actualizar los valores de los parametros en cada iteracion del descenso de gradiente.
  - Las dimensiones del vector  $dW^{[l]}$  son las mismas que las del vector  $W^{[l]}$
4. Con respecto a la funcion de activacion tangente hiperbolica (tanh), seleccione lo que es cierto:
  - Generalmente es preferible usar la funcion tanh en vez de la funcion sigmoide
  - Aunque no modela una probabilidad, se puede aplicar para hacer clasificacion binaria.
  - Tiene la ventaja de que retorna valores centrados en 0 y no en 0.1.
  - Esta funcion tiene el problema de que para valores muy altos su derivada tiende a cero y esto puede estancar el aprendizaje.
5. Forma correcta de modificar las dimensiones un arreglo "a" de numpy.
  - `a.reshape(3,4)`
6. Con respecto a las dimensiones del conjunto de entrenamiento y el conjunto de prueba, seleccione la opcion u opciones verdaderas.
  - El conjunto de prueba debe tener el mismo numero de atributos que el conjunto de entrenamiento.
  - El conjunto de prueba suele tener menos instancias que el conjunto de entrenamiento.
7. Un modelo de regresion logistica puede ser considerado como una red neuronal con una sola unidad (neurona).
  - Verdadero
8. Si  $l$  esta en el rango  $[1, L]$ , seleccione las ecuaciones correctas para  $dz^{[l]}$  en la propagacion hacia atras, puede guiarse por el siguiente diagrama:
  - $dz^{[l]} = dA^{[l]} * g^{[l]'}(z^{[l]})$
9. sobre el sobreajuste o overfitting es cierto que:
  - Consiste en que el modelo se ajusta demasiado a los datos de entrenamiento y pierde su exactitud en datos diferentes.
  - Puede ocurrir cuando se crea una red neuronal superficial con muchas unidades en la capa oculta.
10. Segun la notacion usada en la clase, en una red neuronal,  $a^{[l]}$  representa lo siguiente:
  - La salida de la unidad 3 de la capa 1 para una instancia.
11. De acuerdo a la notacion usada en clase, ¿para que se utiliza el superindice entre corchetes:  $[i]$ ?
  - Ninguna de las anteriores
12. Si tenemos un conjunto de entrenamiento con 30 instancias y 120 atributos, en una red neuronal superficial con 10 unidades ocultas y una unidad en la capa de salida, de acuerdo a la convencion usada en clase ¿Cuales deberian ser las dimensiones de  $A^{[2]}$ ?
  - 1,30
13. Sobre los datos con que se cuenta para entrenar un modelo es cierto lo siguiente:
  - Se suelen dividir en dos partes: conjunto de entrenamiento y conjunto de prueba.
  - Los datos del conjunto de entrenamiento suelen tener mas instancias que los del conjunto de prueba.
14. Si tenemos un conjunto de entrenamiento con 30 instancias y 10 atributos, en una red neuronal superficial con  $h$  unidades ocultas, de acuerdo a la convencion usada en clase ¿Cuales deberian ser las dimensiones de  $W^{[l]}$ ?
  - $h \times 10$
15. Si  $l$  esta en el rango  $[1, L]$ , dada la ecuacion de  $Z$  en la propagacion hacia adelante, seleccione para cada matriz su tamaño correcto  $Z^{[l]} = W^{[l]}A^{[l-1]} + b^{[l]}$ 
  - $A^{[l-1]}$ :  $n^{[l-1]} \times m$
  - $W^{[l]}$ :  $n^{[l]} \times n^{[l-1]}$
  - $b^{[l]}$ (antes de hacer broadcasting):  $n^{[l]} \times 1$
  - $z^{[l]}$ :  $n^{[l]} \times m$
16. En una RNP con  $L$  capas, donde se el numero de iteraciones esta definido en 1500 ¿cuantas veces se realizara la inicializacion de los parametros  $W$  y  $b$ ?
  - 1
17. Con respecto a la funcion de activacion sigmoide es cierto que:
  - Su derivada se aproxima a cero para valores muy grandes de  $z$  y esto causa que el aprendizaje sea lento en estos casos.
  - Puede ser utilizada en la capa de salida en problemas de clasificacion binaria
  - es muchos casos se sustituye por la funcion de activacion tanh( $z$ ) en la capa intermedia ya que esta ultima retorna resultados centrados en cero.
18. De acuerdo con la implementacion de las RNP realizada en el laboratorio ¿Que es cierto sobre la lista denominada <cache>?

- En la propagacion hacia atras la lista se recorre al revés
  - Se utiliza para guardar resultados de la propagacion hacia adelante que son utiles para realizar calculos en la propagacion hacia atras.
19. Antes de iniciar el entrenamiento de un modelo de regresion logistica con n atributos y m instancias es necesario inicializar los parametros w y b de la siguiente forma:  
 $w = 0.0$   
 $b = 0.0$
- Falso
20. De acuerdo a la notacion usada en la clase, ¿como se representaria el vector que contiene las activaciones de la penultima capa de una RNP considerando todas las instancias de entrenamiento?
- $A[L-1]$
21. En la formula de la proopagacion hacia adelante de las redes neuronales profundas, ¿ $A^{[0]}$  es igual a x?  
 $Z^{[l]} = W^{[l]}A^{[l-1]} + b^{[l]}$
- Verdadero
22. Si se entrena una RNS con m instancias, n atributos y t iteraciones, ¿Cuantas veces se modificara el vector  $W^{[2]}$ , asumiendo que en cada modificacion se actualizan todos los elementos del mismo?
- (t) veces
23. si tenemos los siguientes arreglos de numpy:
- $dz[z \leq 0] = 0.0$
24. Indique las dimensiones de los vectores que se le solicitan, considerando las formulas de la regresion logistica vectorizada, utilizando la notacion axb,
- $Z.shape == nxm$
  - $A.shape == nxm$  ?
  - $J.shape == 1x1$
  - $dw.shape == nxn$  ?
  - $db.shape == nx1$  ?
25. Sobre los datos con que se cuenta para entrenar un modelo es cierto lo siguiente:
- Los datos del conjunto de entrenamiento suelen tener mas instancias que los del conjunto de prueba.
  - Se suelen dividir en dos partes: conjunto de entrenamiento y conjunto de prueba.
26. Sobre la vectorizacion es cierto que:
- PErmite en muchos casos reducir el tiempo de entrenamiento de una red.
  - Intenta aprovechar las capacidades de ejecucion paralela del procesador.
  - Se trata de eliminar ciclos de repeticion.
27. Con respecto a la funcion de activacion ReLu:
- Tiene problemas cuando se presentan valores de entrada negativos pues su derivada se hace cero.
  - No estanca el aprendizaje para valores altos de entrada pues su derivada es constante positiva.
28. Indique cual seria la forma correcta de obtener el vector Z en la regresion logistica:
- $Z = w.T.dot(X) + b$
29. Si se entrena una red neuronal superficial con m instancias de 6 atributos y n <sup>[l]</sup> unidades en la capa oculta:
- Durante el entrenamiento se produciran m vectores a <sup>[l]</sup> con las activaciones de la capa oculta
30. IMAGEN
- A la izquierda actualizar parametros a la derecha calcula el costo.
31. Con respecto a inicializar  $W[1]$  igual a una matriz de ceros en una red neuronal superficial, seleccione:
- Lo mejor es inicializar a valores aleatorios cercanos a cero.
  - No se debe de hacer porque sino todos los modelos serian y se busca diversidad entre estos.
32. imagen flor:
- Linealmente no separable
  - la regresion logistica no tiene buen performance