HOJA N°

FECHA

Parcial 2022

1) a) Rango Inter Quartil $= Q_3 - Q_1 = 141 - 97 = 44$

$LS = Q_3 + 1{,}5 . RIQ = 207$

$LI = Q_1 - 1{,}5 . RIQ = 31$

Es atípico si está por fuera del rango $[31, 207]$

→ leve si está en el intervalo $[Q_1 - 3 . RIQ, LI]$ o $[LS, Q_3 + 3RIQ]$

→ extremo si es mayor a $Q_3 + 3RIQ$ o menor a $Q_1 - 3RIQ$

, $Q_1 - 3RIQ = -35$ $\quad Q_3 + 3 . RIQ = 273$

Normalización lineal uniforme: $X_1' = \dfrac{X - X_{min}}{X_{max} - X_{min}}$

Normalización Estandard: $X_2' = \dfrac{X - media(X)}{desviacion(X)}$

300

atípico extremo

$X_1' = 0{,}9$

$X_2' = 4{,}16$

30

atípico leve

$X_1' - 0{,}12$

$X_2' = -2{,}33$

250

atípico leve

$X_1' = 0{,}71$

$X_2' = 2{,}95$

200

no es atípico

$X_1' = 0{,}52$

$X_2' = 1{,}75$

NOTA

b) Los valores normalizados linealmente representan una transformación de los datos originales en una escala del 0 al 1, representando sus relaciones proporcionales, es muy sensible a valores atípicos (outliers) y si se recortan los extremos se obtienen valores negativos y/o mayores a 1 (valores inferiores al mínimo darán números negativos y mayores al máximo, mayores a 1)

c) Los valores normalizados con media y desvío representan una transformación de los datos para que tengan media 0 y desviación estandard 1. Se representa cuantas desviaciones estandar esta un dato por encima o por debajo de la media, se suele usar si los datos tienen una distribución normal

d) la normalización con media y desvío se relaciona más con la detección de valores atípicos, ya que si un valor está entre 2 a 3 desviaciones estandar por encima o debajo de la media se consideran atípicos leves y si está a mas de 3 es un atípico extremo. En cambio en la normalización lineal uniforme no tenemos forma de saber si un valor es atípico.

2) $$y = \sum_{i=1}^{n} x_i w_i + \text{bias}$$

$$y = 10 \cdot 0{,}392 + 150 \cdot 0{,}001 + 2 \cdot (-2{,}441) + 0{,}4022$$
$$y = -0{,}4098$$

Función de activación umbral binaria: $\text{salida} = \begin{cases} 1 & \text{si } y \geq 0 \\ 0 & \text{si } y < 0 \end{cases}$

$\Rightarrow$ salida $= 0$, por lo que suponiendo que no $= 0$ y sí $= 1$, se puede ver que es no seleccionado

NOTA

HOJA N°

FECHA

3)	Perceptrón	Combinador Lineal	Neurona no lineal
Salida neta	$\sum_i X_i W_i + b$	$\sum_i X_i W_i + b$	$\sum_i X_i W_i + b$
Función de activación	Umbral binario {0 o 1}	Función lineal $(-\infty, +\infty)$	Sigmoide $[0,1]$, tanh $[-1,1]$, relu $[0,+\infty)$, etc
Salida	$y = \begin{cases} 1 & neta \geq 0 \\ 0 & neta < 0 \end{cases}$	$y = \sum_{i=1}^{n} x_i W_i + W_0$	$y = \frac{1}{1+e^{-x}}$ $y = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$ $y = \max(0, x)$
tipo de problema que resuelve	clasificación binaria en problemas linealmente separables	regresión lineal	clasificación binaria y regresión, incluso en problemas linealmente no separables
aprendizaje	actualiza vector proyección $W_{nuevo} = W + \alpha(t-y)X$ t: valor esperado y: valor obtenido	descenso de gradiente $W_{nuevo_i} = W_i + \alpha \frac{\partial(ECM)}{\partial W_i}$	descenso de gradiente $W_{nuevo_i} = W_i + \alpha \frac{\partial(C)}{\partial W_i}$, C: función de costo, puede ser ECM, entropía cruzada binaria, etc

NOTA

4) a)

59	X	X
0	34	34
0	X	X

$$\text{recall}_2 = 0{,}5 = \frac{TP}{TP+FN} = \frac{34}{34+FN} \Rightarrow FN = 34$$

(clase, predicción)

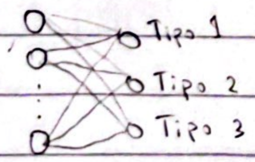
$$\text{matriz} = \begin{matrix} (1,1) & (1,2) & (1,3) \\ (2,1) & (2,2) & (2,3) \\ (3,1) & (3,2) & (3,3) \end{matrix}$$

$$\text{precisión}_{\text{clase}} = \frac{\text{bien predecidos}}{\text{bien predecidos + intrusos}}$$

$$\text{recall}_{\text{clase}} = \frac{\text{bien predecidos}}{\text{bien predecidos + los que fueron a parar a otro lado}}$$

$$\text{precisión clasificador} = \frac{\text{correctas}}{\text{totales}}$$

$$\text{F1-score} = \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

5) 

Sigmoide: $y_j = \frac{1}{1+e^{-x_j}}$ softmax: $y = \frac{e^{neta_j}}{\sum_k e^{neta_k}}$

a) sigmoide: $y_1 = 0{,}18$
$y_2 = 0{,}98$
$y_3 = 0{,}99$

b) softmax: $y_1 = 0{,}0005$
$y_2 = 0{,}11$
$y_3 = 0{,}88$

NOTA => el vino será de tipo 1 => el vino será de tipo 1

HOJA N°

FECHA

6) a) Una función de costo sirve para ver que tan alejada está la predicción, debe devolver valores positivos, ser derivable y representar el error, donde tiende a 0, es por eso que sería ideal que todas las funciones de costo sean convexas. Si el costo es 0 significa que el modelo hizo una predicción perfecta, tiende a 0 a medida que la neurona aprende la salida deseada

b) la entropía binaria cruzada resuelve en algunas situaciones el problema del desvanecimiento de gradiente con neuronas con función de activación sigmoide. Esto se debe a que el gradiente del ECM es $-(t_K - y_K)f'(neta)x_j$ y el de la entropía cruzada Binaria es $-(t_K - y_K)x_j$, por lo que lo único que cambia es que en el ECM aparece la derivada de la función sigmoide y lo único que hace esto es multiplicarlo por un factor de escala que va desde 0 a 0,25, entonces lo que termina haciendo es que el paso sea mucho mas pequeño y tarde mas el entrenamiento, es por eso que usamos la ECB, ya que produce gradientes mas grandes y es más adecuado para clasificación binaria, ya que aprovecha la naturaleza probabilística de la sigmoide

c) no podría usarse dado que el perceptrón no utiliza descenso de gradiente, actualiza el vector proyección. Además se suele usar con la función sigmoide y el perceptrón utiliza la función umbral

7) al utilizar lotes de 150 estaríamos usando descenso de gradiente estocástico con mini lote, lo que hacemos es procesar 150 ejemplos, actualizar los pesos y así hasta haber completado todos los ejemplos, esto lo hacemos usando el

NOTA

algoritmo de backpropagation. si no usamos lotes puede ser que estemos usando descenso de gradiente completo que es como el caso anterior pero como si tuvieramos lotes del tamaño de la cantidad de ejemplos, o descenso de gradiente estocastico que por cada ejemplo que entra se actualizan los pesos

b) i) $3000/150 = 20$

20 iteraciones por época y 200 epocas
=> 4000 iteraciones

ii) los pesos se actualizarán 4000 veces

iii) con descenso de gradiente completo 200 veces
y con descenso de gradiente estocástico $3000 \cdot 200 = 600000$ veces

8) a) $(\overset{\text{tamaño}}{32} + 2 \cdot \underset{\text{padding}}{0} - \underset{\text{Kernel size}}{4}) / \underset{\text{stride}}{2} + 1 = 15$

output_shape $= 15 \times 15 \times 8$

b) $\underset{\text{filtros}}{8} \cdot \underset{\text{Kernel}}{4} \cdot \underset{\text{Kernel}}{4} \cdot \underset{\text{canales}}{3} + \underset{\text{filtros x el bias}}{8} = 392$

NOTA