

Apuntes Semana 8 - 08/04/2025

Dillan Almendares Barrantes - c.2020033336

I. RESPUESTAS QUIZ #4

A. Describa normalización y estandarización

Normalización: Ajustar los features para que tengan una escala única, normalmente de 0 a 1.

Estandarización: Que los nuevos valores tengan un promedio de 0 y una desviación estándar de 1.

B. Anote la derivada de la función sigmoide

$$\sigma'(x) = \sigma(x)(1 - \sigma(x))$$

C. Describa las fórmulas para calcular el accuracy, recall, precision y F1-Score

- **Accuracy:** $\frac{TP+TN}{TP+TN+FP+FN}$
- **Recall:** $\frac{TP}{TP+FN}$
- **Precision:** $\frac{TP}{TP+FP}$
- **F1-Score:** $\frac{2 \cdot \text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}$

II. OBSERVACIONES TAREA#1

- Falta agregar conclusiones de las gráficas.
- Los boxplots deben ir junto la explicación de los datos que representan.
- Para el proyecto, se deben agregar las observaciones útiles para el informe.
- Asegurarse de registrar en el notebook las selecciones de hiperparámetros.
- Evitar ejecutar el validation set hasta el final del training.

III. LECTURAS

Dos nuevas lecturas para el próximo quiz (lunes 29 de abril):

- Capítulo 6
- ImageNet Classification with Deep Convolutional Neural Networks

IV. FUNCIONES DE ACTIVACIÓN

Permiten resolver problemas más complejos al permitirle a las redes neuronales no estar limitadas a problemas lineales.

A. Función lineal

- La derivada no depende de la entrada y es constante, por lo que no funciona para el descenso del gradiente y no aprende de los samples.
- $g(x) = \alpha x$
- $\frac{\partial(g(x))}{\partial(x)} = \alpha$

B. Función Sigmoide

- En sus extremos la derivada es 0.
- Gradiente pequeño al final de la función.
- **Vanishing gradient:** Los gradientes son tan pequeños que el entrenamiento se estanca o converge muy lentamente.

C. Tangente Hiperbólica

- Activación entre -1 y 1.
- Positivo y negativo.
- Centrado en el origen.
- Estrictamente creciente.
- Limitada.
- Utilizada en modelos de lenguaje (LSTM).
- Diferenciable en cualquier punto.
- Mantiene los mismos problemas de las anteriores, pero al estar en el centro funciona mejor que la sigmoide.

D. Rectifier Linear Unit (ReLU)

- Acotada debajo del cero.
- No está acotada en los positivos.
- Estrictamente creciente.
- Eficiente para Deep Learning.
- Mata las activaciones.
- Cualquier valor en el negativo de la función es información muerta, todo dará 0.
- No es derivable en 0, para solucionarlo se considera como 1 o 0 cuando el X es exactamente 0.
- $g(x) = \max(x, 0)$

E. Leaky (ReLU)

- Asigna una pequeña constante al mínimo permitido.
- Resuelve el problema de las neuronas muertas.
- Hay que seleccionar el valor que multiplica a la x.
- $g(x) = \begin{cases} 0.01x, & \text{si } x < 0 \\ x, & \text{si } x \geq 0 \end{cases}$
- $\frac{\partial(g(x))}{\partial(x)} = \begin{cases} 0.01, & x < 0, \\ 1, & x \geq 0 \end{cases}$

F. Parametric (ReLU)

- Permite aprender un parámetro para dejar que la señal continúe.
- El parámetro es aprendido, al igual que el resto de la red.
- $g(x) = \begin{cases} w, & x < 0, \\ x, & x \geq 0 \end{cases}$
- $\frac{\partial(g(x))}{\partial(x)} = \begin{cases} w, & x < 0, \\ 1, & x \geq 0 \end{cases}$

G. Softmax

- Multiclase.
- Distribución de probabilidad.
- Suma 1.
- Cross Entropy Loss.
- Usada como salida de clasificación.
- El vector de entrada se llama logits.

H. Cross Entropy Loss

- Log-Loss o Logistic Loss.
- Representa probabilidades en un espacio logarítmico $[0,1]$.
- Numéricamente estable.
- $L = -\log(P(Y = y, X = x))$
- $L = -\log\left(\frac{e^{s_k}}{\sum e^{s_j}}\right)$, con j desde 1,...,C.