



TAREA – INFORME – PREGUNTAS REFLEXIVAS

CLUSTERING - BREAST CANCER WISCONSIN
DIAGNOSTIC DATASET

OBJETIVO

Aplicar técnicas de regresión, clasificación y redes neuronales para explorar y analizar patrones en los datos del cáncer de mama.

REALIZADO POR

LUIS MIGUEL RAMÍREZ / AVILES PAUTE
JOSÉ / ESPINOZA BONE JOSÉ (GRUPO 7)

MATERIA

INTELIGENCIA ARTIFICIAL: DATA MINING I

1. ¿Qué método te pareció más adecuado para este conjunto de datos? ¿Por qué?

El modelo de regresión logística fue el más adecuado para este conjunto de datos, ya que la clasificación entre tumores benignos y malignos presenta una separación clara en el espacio de características, lo cual favorece a los modelos lineales. La regresión logística alcanzó una exactitud del 98%, comparable con la red neuronal, pero con un modelo mucho más simple, interpretable y rápido de entrenar. Pero sin subestimar el potencial de las redes neuronales.

2. ¿Qué criterios utilizaste para seleccionar las variables predictoras?

Utilizando el análisis del trabajo anterior se tomaron todas las variables pero escalando sus datos usando un escalador StandardScaler. Se tomó todas las variables para evitar tener que eliminar información sensible, dado el poco conocimiento en lo que respecta a análisis de muestras de tejido mamario, se requiere la asistencia de varios científicos y expertos en lo que respecta a cáncer mamario.

3. ¿Cómo influyó el número de capas y neuronas en la red neuronal en el desempeño del modelo?

El número de capas y neuronas tuvo poca influencia en el desempeño final, ya que el dataset es de tamaño moderado y el problema no es altamente no lineal.

Con una arquitectura simple de 1 capas oculta (3 neuronas), el modelo logró una exactitud muy similar a la regresión logística (~98%), aumentar el número de capas o neuronas podría generar sobreajuste (overfitting) si no se regulariza adecuadamente. Se debería hacer pruebas con mayor cantidad de datos.

4. ¿Qué ventajas observaste al usar redes neuronales frente a regresión logística?

Las ventajas observadas al usar redes neuronales fueron:

Capacidad para modelar relaciones no lineales, en caso de que existan interacciones complejas entre variables.

Flexibilidad: pueden ampliarse para incluir más capas, técnicas de regularización (dropout, batch normalization) o incluso extenderse a arquitecturas más profundas.

Potencial de generalización: con datos más grandes o con ruido, las redes pueden adaptarse mejor que un modelo lineal.

5. ¿Cómo podrían usarse los resultados de los modelos en decisiones médicas o investigación clínica?

En práctica clínica:

Un modelo de regresión logística podría usarse para estimar la probabilidad de malignidad y apoyar decisiones de biopsia o seguimiento.

Una red neuronal podría integrarse en sistemas más grandes de diagnóstico asistido por imagen, combinando datos clínicos, histológicos o radiológicos.

En investigación:

Permiten identificar variables más influyentes en la detección temprana de cáncer.

Facilitan estudios de correlación entre características celulares y evolución clínica, contribuyendo a la medicina personalizada.