

Análisis Predictivo y Gestión de Datos

Sesión 7: Comparación de modelos e introducción a redes neuronales

Oscar Leonardo Rincón León

24 de abril de 2025

Comparar modelos supervisados e introducir redes neuronales multicapa usando `MLPClassifier`.

¿Por qué comparar modelos?

- No todos los modelos funcionan igual para todos los datos.
- Comparar modelos permite detectar cuál generaliza mejor.
- Ayuda a justificar decisiones en análisis predictivo.

- Usar la misma métrica de evaluación (ej. F1, AUC).
- Aplicar los modelos sobre los mismos datos.
- Usar validación cruzada para evitar depender de una sola partición.

¿Qué modelos comparamos?

- **KNN**: clasificador basado en vecinos cercanos. Modelo perezoso, sensible a la escala.
- **Regresión logística**: modelo lineal para clasificación. Interpretable y eficiente.
- **Árbol de decisión**: modelo jerárquico. Captura no linealidades pero puede sobreajustar.

¿Qué es la validación cruzada?

- Técnica para estimar el desempeño real de un modelo.
- Divide los datos en k particiones (folds).
- Se entrena sobre $k-1$ partes y se evalúa sobre la restante, repitiendo el proceso.
- Reduce el sesgo de una única división de entrenamiento/prueba.

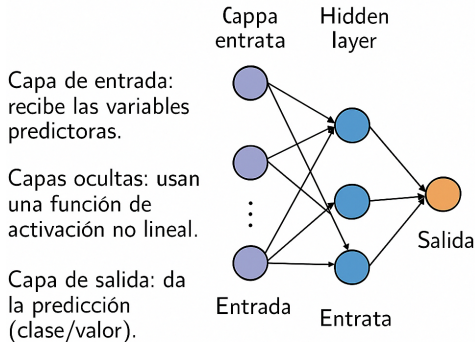
¿Qué es una red neuronal?

- Modelo basado en capas de nodos (neuronas).
- Cada nodo aplica una función no lineal a una combinación de entradas.
- Tiene al menos una capa oculta entre entrada y salida.
- Aprende representaciones complejas de los datos.

Estructura de una red neuronal (MLP)

- Capa de entrada: recibe las variables predictoras.
- Capas ocultas: aplican funciones no lineales.
- Capa de salida: da la predicción (por ejemplo, probabilidad de clase 1).

Estructura de una red neuronal (MLP)



¿Qué es MLPClassifier?

- Implementación de perceptrón multicapa en `scikit-learn`.
- Hiperparámetros:
 - `hidden_layer_sizes`: número y tamaño de capas ocultas.
 - `activation`: función de activación (ReLU, tanh, etc.).
 - `solver`: algoritmo de optimización (ej. Adam).
- Entrena con descenso del gradiente.

Ventajas y desventajas

Ventajas

- Captura relaciones no lineales complejas.
- Puede superar modelos tradicionales si se entrena bien.

Desventajas

- Difícil de interpretar.
- Sensible a la selección de hiperparámetros.

- Se compararon modelos KNN, regresión logística y árboles.
- Se introdujo validación cruzada para evaluación más robusta.
- Se explicó la arquitectura de una red neuronal y el uso de `MLPClassifier`.