

Actividad Práctica: Clasificación con K-Nearest Neighbors (KNN) en R usando el dataset Iris

Objetivo:

En esta actividad, aprenderás a:

1. Cargar y explorar el dataset ``iris``.
2. Dividir el dataset en conjunto de entrenamiento (80%) y prueba (20%).
3. Aplicar el algoritmo K-Nearest Neighbors para distintos valores de ``k``.
4. Evaluar el modelo usando una matriz de confusión.

Instrucciones:

Sigue los siguientes pasos en R:

Paso 1: Cargar librerías y el dataset

- Cargar las librerías necesarias (``class``, ``caret``, ``dplyr``).
- Cargar el dataset ``iris``.
- Visualizar las primeras filas del dataset.

Pregunta 1:

¿Cuántas clases tiene el dataset? ¿Cuántas muestras hay por clase?

Paso 2: División del dataset en entrenamiento y prueba (80% - 20%)

- Dividir el dataset en un conjunto de entrenamiento (80%) y uno de prueba (20%).
- Separar las características (X) y las etiquetas (Y).

Pregunta 2:

¿Cuántas muestras quedaron en el conjunto de entrenamiento? ¿Y en el de prueba?

Paso 3: Aplicar KNN con distintos valores de ``k``

- Probar el algoritmo KNN con valores de ``k`` diferentes: ``[1, 3, 5, 7, 9]``.
- Registrar la precisión de cada modelo.

Pregunta 3:

¿Cuál es el valor de ``k`` que proporciona la mayor precisión? ¿Por qué?

Paso 4: Evaluación con Matriz de Confusión

- Evaluar el modelo con el mejor valor de ``k`` usando una matriz de confusión.

- Mostrar las métricas de evaluación.

Pregunta 4:

¿Cuál fue la precisión final? ¿Hubo errores de clasificación? ¿Qué clases se confundieron más?

Tarea Extra (Opcional):

1. ¿Cómo cambiarían los resultados si aplicas una normalización previa a los datos?
2. Modifica el código para realizar validación cruzada de 10 folds y comparar los resultados.

Conclusión:

- Aplicaste el algoritmo K-Nearest Neighbors en un dataset real.
- Experimentaste con distintos valores de `k` y evaluaste el desempeño del modelo.
- Utilizaste una matriz de confusión para la evaluación final.

¡Buen trabajo! 🚀