



Casa abierta al tiempo

UNIVERSIDAD AUTÓNOMA METROPOLITANA

Unidad Iztapalapa

Actividades 28: Validación cruzada de 10 bloques

Autor

Martínez Buenrostro Jorge Rafael

Profesor

Eduardo Filemón Vázquez Santacruz

27 de enero de 2024

La validación cruzada de 10 bloques, también conocida como validación cruzada de k bloques o validación cruzada k -fold, es una técnica comúnmente utilizada en el campo del aprendizaje automático y la estadística para evaluar el rendimiento de un modelo predictivo. La idea principal es dividir el conjunto de datos en k bloques o partes, y luego realizar el entrenamiento y la evaluación del modelo k veces, utilizando diferentes combinaciones de bloques para cada iteración.

En el caso específico de la validación cruzada de 10 bloques:

1. División en Bloques:

- El conjunto de datos se divide en 10 bloques o partes iguales.

2. Iteraciones:

- El modelo se entrena y evalúa 10 veces. En cada iteración, se utiliza un bloque diferente como conjunto de prueba, mientras que los otros 9 bloques se utilizan como conjunto de entrenamiento.

3. Promedio de Resultados:

- Después de las 10 iteraciones, se promedian los resultados de rendimiento del modelo (por ejemplo, precisión, error, etc.) para obtener una medida general de su desempeño.

La validación cruzada de k bloques es útil porque permite utilizar todo el conjunto de datos tanto para entrenamiento como para prueba a lo largo de las k iteraciones, lo que puede proporcionar una evaluación más robusta del rendimiento del modelo en comparación con una simple división en entrenamiento/prueba.

La elección de k (el número de bloques) puede variar según el tamaño del conjunto de datos y la naturaleza del problema. Comúnmente, se utiliza $k = 10$, pero otras opciones comunes incluyen $k = 5$ o $k = 20$.

Esta técnica es especialmente valiosa cuando el conjunto de datos es limitado y se desea maximizar el uso de los datos disponibles para la evaluación del modelo. La validación cruzada ayuda a reducir el riesgo de que los resultados estén sesgados por una partición específica del conjunto de datos.