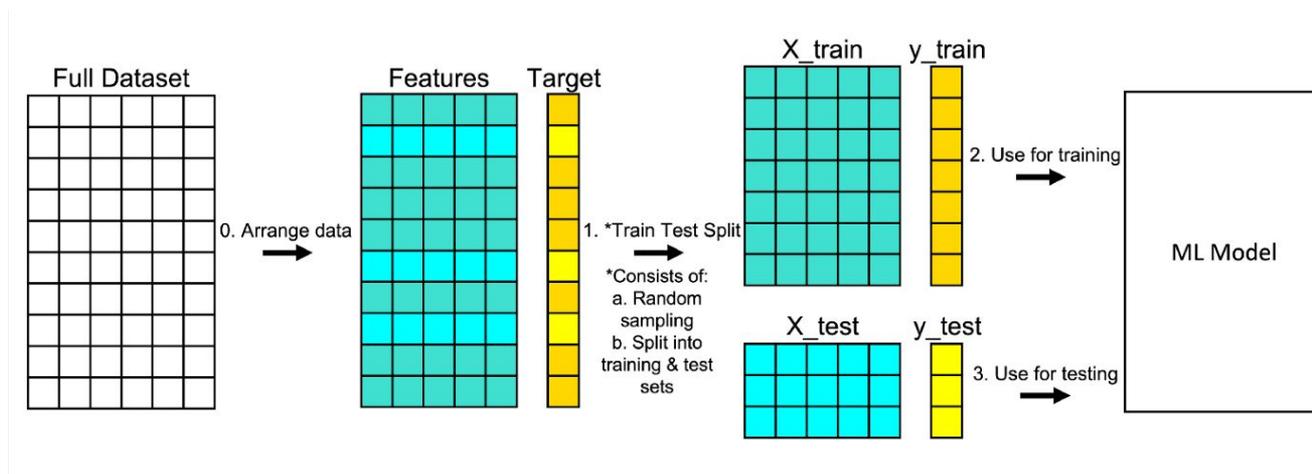


Validación de modelos

Fabián Villena

Separación de subconjuntos de entrenamiento y prueba

Para saber qué tan bien generaliza un modelo hay que dividir el conjunto de datos en dos subconjuntos: Un subconjunto de entrenamiento, al cual ajustaré mi modelo y un subconjunto de prueba, con el cual se evaluará la generalización.



Subconjunto de validación

Si probamos varios modelos con distintos hiperparámetros sobre exactamente el mismo subconjunto de entrenamiento, estaremos sesgando a los modelos a comportarse de manera correcta específicamente en ese subconjunto.

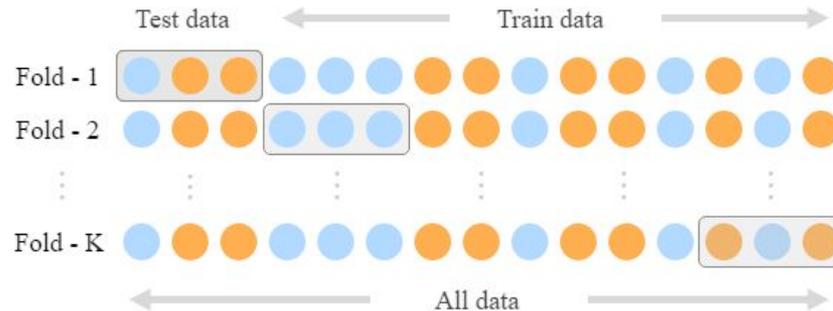
Para eso podemos utilizar una parte del conjunto de entrenamiento sólo para seleccionar los mejores hiperparámetros del modelo y después de eso entrenamos con todo el subconjunto de entrenamiento y probamos en el subconjunto de prueba.



Validación cruzada

Un subconjunto de validación puede ser demasiado pequeño o puede disminuir significativamente los ejemplos del subconjunto de prueba, disminuyendo el rendimiento general del modelo.

La validación cruzada es una técnica para generar múltiples subconjuntos de validación. Cada modelo es evaluado una vez por cada uno de los subconjuntos de validación generados y entrenado en cada subconjunto de entrenamiento generado. Al promediar todos los resultados podemos tener un resultado más preciso



k-fold cross validation

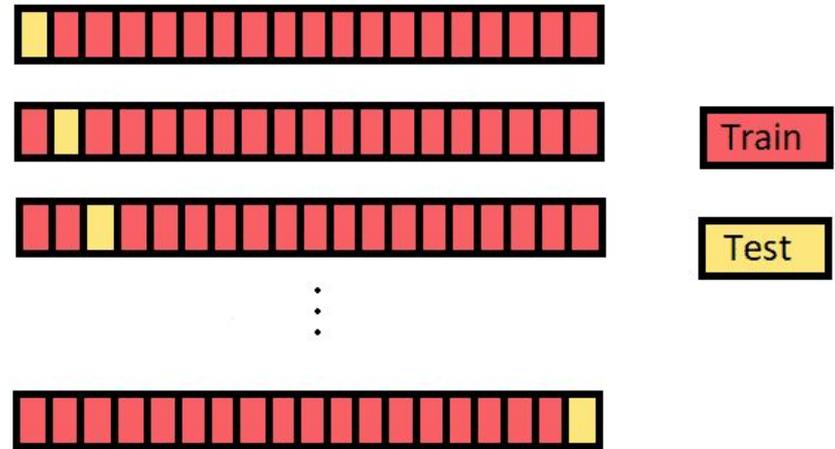
1. Establecemos una cantidad determinada de grupos en los cuales un conjunto de datos se dividirá.
2. Para cada grupo:
 - a. Guardar el grupo para validación.
 - b. Tomar los grupos restantes para entrenamiento.
 - c. Entrenar en los grupos de entrenamiento.
 - d. Evaluar sobre el grupo de validación.
 - e. Guardar el resultado y descartar el modelo.
3. Resumir los resultados de cada uno de los grupos.



Leave-one-out cross validation

Es un caso especial de la validación cruzada en donde sólo guardamos un ejemplo para validación y utilizamos los ejemplos restantes para el entrenamiento.

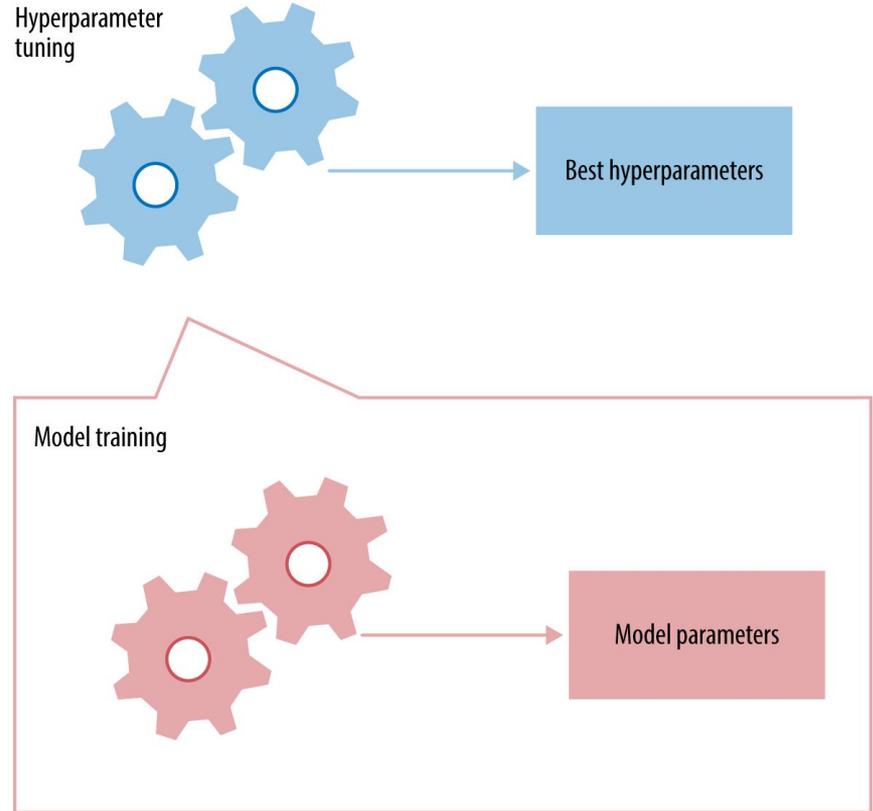
Este tipo de validación cruzada es recomendada para conjuntos de datos pequeños.



Hiperparámetros

Un hiperparámetro es un parámetro cuyo valor es utilizado para controlar el proceso de entrenamiento. Al contrario de los parámetros del modelo que son ajustados a través del entrenamiento.

Debemos seleccionar el conjunto de hiperparámetros que mejor rendimiento genere en el modelo.



Optimización de hiperparámetros

Es el proceso de selección del conjunto de hiperparámetros que optimiza el rendimiento de un modelo específico.

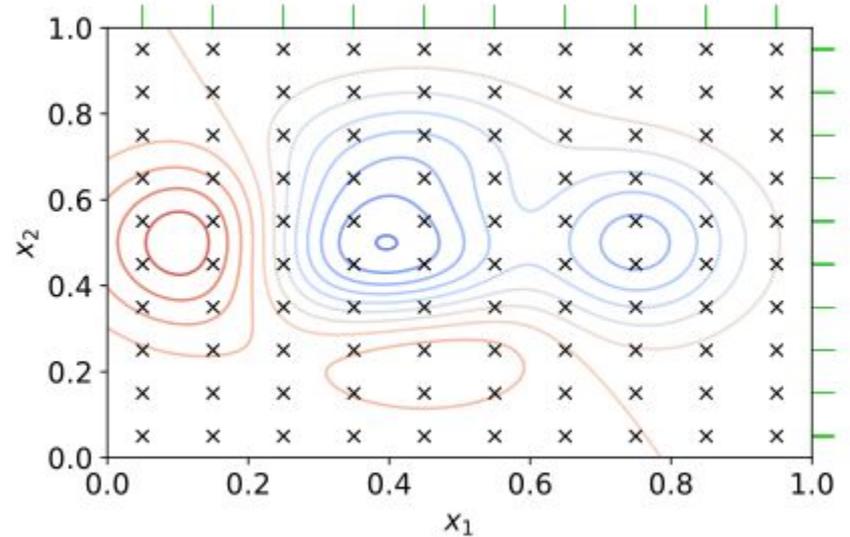
Cada algoritmo para el entrenamiento de modelos puede necesitar distintos hiperparámetros para generalizar distintos patrones de los datos.

A través del uso de validación cruzada podemos tener un rendimiento promedio de cada combinación de hiperparámetros.

Grid-search

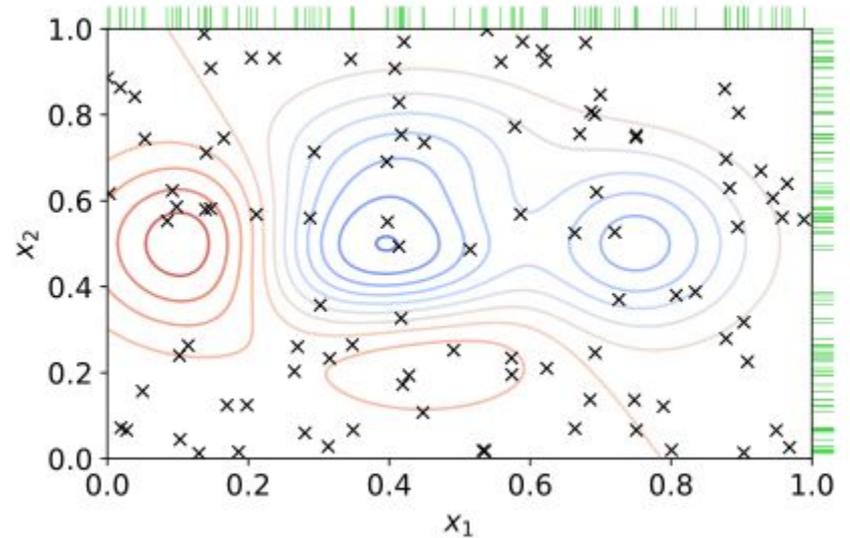
Es la forma tradicional de optimizar los hiperparámetros de un modelo.

Es una búsqueda exhaustiva del mejor conjunto de hiperparámetros sobre un espacio de hiperparámetros manualmente definido.



Randomized search

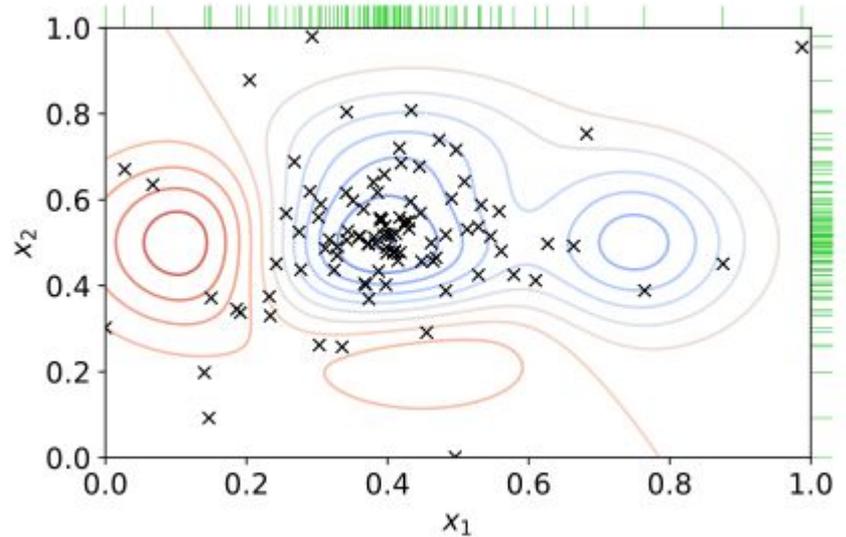
Es un reemplazo a la búsqueda exhaustiva en donde se seleccionan al azar un subconjunto del espacio de hiperparámetros para medir el rendimiento del modelo.



Optimización Bayesiana

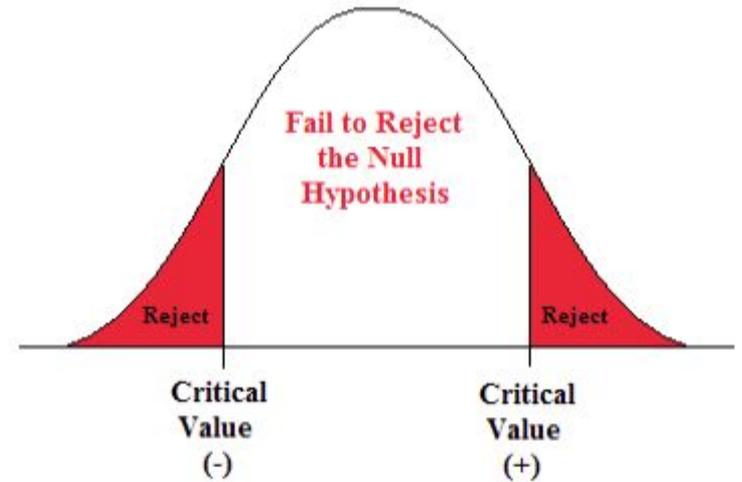
Es un método de optimización global en donde se construye un modelo probabilístico entre los valores de los hiperparámetros y la función objetivo.

Se ha demostrado que este tipo de optimización obtiene mejores resultados con menores evaluaciones comparado a una búsqueda exhaustiva.



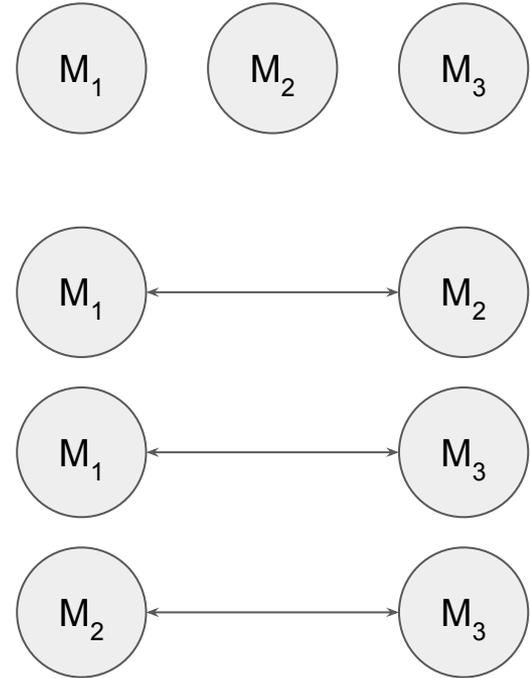
Contraste de hipótesis

Para poder detectar si existen diferencias estadísticamente significativas entre los resultados de los modelos a comparar, podemos realizar pruebas de contraste de medias para comparar los rendimientos promedio de cada modelo después de realizar una validación cruzada.



Prueba de t de student

ra detectar si existen diferencias estadísticamente significativas entre los resultados promedio de cada uno de los modelos (que se distribuyan de manera normal) podemos realizar pruebas entre todas las combinaciones de pares de modelo que queremos comparar.



Corrección de bonferroni

Para poder corregir el problema de comparaciones múltiples, debemos corregir los p-valores de los resultados de los contrastes de hipótesis.

Selección del mejor modelo

Al final debemos seleccionar el modelo que mejor se comporte en promedio, si no existen diferencias estadísticamente significativas entre modelos, podemos seleccionar el modelo que tenga la menor complejidad computacional.