



Instituto Tecnológico
de Buenos Aires

Validation Set vs. Cross Validation

y

Random Search vs. Grid Search

—



Instituto Tecnológico
de Buenos Aires

Elección del Enfoque Correcto

Validation set

- Subconjunto de datos utilizado para evaluar el rendimiento del modelo durante el entrenamiento.
- Se utiliza para ajustar los hiperparámetros y evitar el sobreajuste del modelo.
- Permite medir cómo se generaliza el modelo a datos no vistos y tomar decisiones de ajuste.

Ventajas:

- Implementación más simple y rápida.
- Adecuado cuando se dispone de suficientes datos y no hay limitaciones computacionales.
 - Permite ajustar los hiperparámetros de forma más específica al modelo

Desventajas:

- El rendimiento puede variar según la división específica de los datos.
- No proporciona una estimación tan precisa del rendimiento.

Cross Validation

- Técnica que divide los datos en múltiples conjuntos de entrenamiento y validación (*folds*).
- Proporciona una estimación más precisa del rendimiento del modelo al promediar los resultados de varias divisiones.
- Ayuda a evaluar el rendimiento del modelo de manera más precisa, reduciendo la dependencia de una única división de datos.

Ventajas:

- Proporciona una estimación más precisa del rendimiento del modelo al promediar múltiples divisiones de datos.
- Útil cuando hay escasez de datos y se desea evaluar el rendimiento sin depender de una sola división.

Desventajas:

- Mayor costo computacional debido a los múltiples entrenamientos del modelo.
 - Puede ser más lento en comparación con el uso de un validation set único.

Comparación de Técnicas de Optimización

Grid Search

- Técnica que realiza una búsqueda exhaustiva evaluando todas las combinaciones posibles de hiperparámetros en una cuadrícula definida.
- Garantiza encontrar la mejor combinación dentro de la cuadrícula, sin omitir ninguna opción.
- Adecuado cuando se tiene un conocimiento claro del espacio de hiperparámetros y se desea una búsqueda sistemática.

Ventajas:

- Realiza una búsqueda exhaustiva, garantizando los mejores hiperparámetros dentro de la cuadrícula definida.

Desventajas:

- Costo computacional alto cuando el espacio de hiperparámetros es grande.
- Puede ser impráctico cuando hay muchos hiperparámetros a considerar.

Random Search

- Técnica que selecciona de manera aleatoria combinaciones de hiperparámetros para evaluar el rendimiento del modelo.
- Explora un espacio amplio de hiperparámetros de manera eficiente, permitiendo encontrar combinaciones óptimas.
- Útil cuando se desconoce la forma exacta del espacio de hiperparámetros o cuáles son los valores óptimos.

Ventajas:

- Explora un espacio de hiperparámetros más amplio de manera eficiente.
- Es especialmente útil cuando los hiperparámetros óptimos son inciertos, y hay una gran cantidad

Desventajas:

- Existe la posibilidad de perder los mejores valores de los hiperparámetros.



Instituto Tecnológico
de Buenos Aires

¡Gracias!

—