COLOMBIA POTENCIA DE LA VIDA

TIC

Introducción a Evaluación de Modelos

TALENTO TECH | AZ PROYECTOS EDUCATIVOS | UTP Universidad Tecnológica de Pereira

Introducción a Evaluación de Modelos

La evaluación de modelos es una parte fundamental en el proceso de desarrollo de modelos de aprendizaje automático. Permite determinar qué tan bien se desempeña un modelo en la tarea para la que fue diseñado y proporciona información valiosa sobre su capacidad predictiva y su generalización a datos no vistos. En este sentido, la evaluación de modelos se lleva a cabo de manera específica según el tipo de problema que se está abordando, ya sea clasificación, regresión o agrupación.

Evaluación en Modelos de Clasificación

En el contexto de modelos de clasificación, se evalúa la capacidad de un modelo para asignar correctamente etiquetas de clase a instancias de datos. Algunas de las métricas de evaluación comunes incluyen:

- **Exactitud (Accuracy):**

Es la proporción de predicciones correctas sobre el total de predicciones realizadas por el modelo. Sin embargo, la exactitud puede ser engañosa en conjuntos de datos desbalanceados.

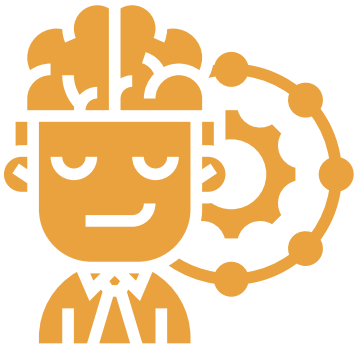
- **Precisión (Precision) y Recall (Recuperación)**

Precisión mide la proporción de instancias positivas correctamente identificadas entre todas las instancias identificadas como positivas, mientras que **Recall** mide la proporción de instancias positivas correctamente identificadas entre todas las instancias realmente positivas.

- **F1-Score:**

Es la media armónica de Precisión y Recall. Es útil cuando hay un desequilibrio entre las clases.

- **Curva ROC (Receiver Operating Characteristic) y Área bajo la curva (AUC):**



Son útiles para evaluar modelos de clasificación binaria y miden el rendimiento del modelo en términos de la tasa de verdaderos positivos frente a la tasa de falsos positivos.

Evaluación en Modelos de Regresión

En el caso de modelos de regresión, la evaluación se centra en la capacidad del modelo para predecir valores numéricos. Algunas métricas de evaluación comunes incluyen:

Error cuadrático medio (Mean Squared Error, MSE)	Error absoluto medio (Mean Absolute Error, MAE)
Es la media de los errores al cuadrado entre las predicciones del modelo y los valores reales.	Es la media de las diferencias absolutas entre las predicciones del modelo y los valores reales.

Coeficiente de determinación (R^2)
Indica la proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes. Un valor de R^2 más cercano a 1 indica un mejor ajuste del modelo

Evaluación en Modelos de Agrupación

En modelos de agrupación, la evaluación se enfoca en la capacidad del modelo para identificar grupos coherentes y significativos en los datos. Sin embargo, evaluar modelos de agrupación es más complejo que en clasificación o regresión, ya que no hay etiquetas de clase para comparar directamente con los clusters encontrados.

Algunas métricas de evaluación comunes incluyen:

Índice de Silueta (Silhouette Score)

Mide la coherencia y separabilidad de los clusters. Un valor más cercano a 1 indica que las instancias están bien agrupadas, mientras que un valor negativo indica que pueden haberse agrupado incorrectamente.

Inercia

Mide la suma de las distancias cuadradas de cada punto a su centroide más cercano. Una menor inercia indica clusters más compactos.

Índice de Rand Ajustado (Adjusted Rand Index, ARI) y Mutual Information (MI)

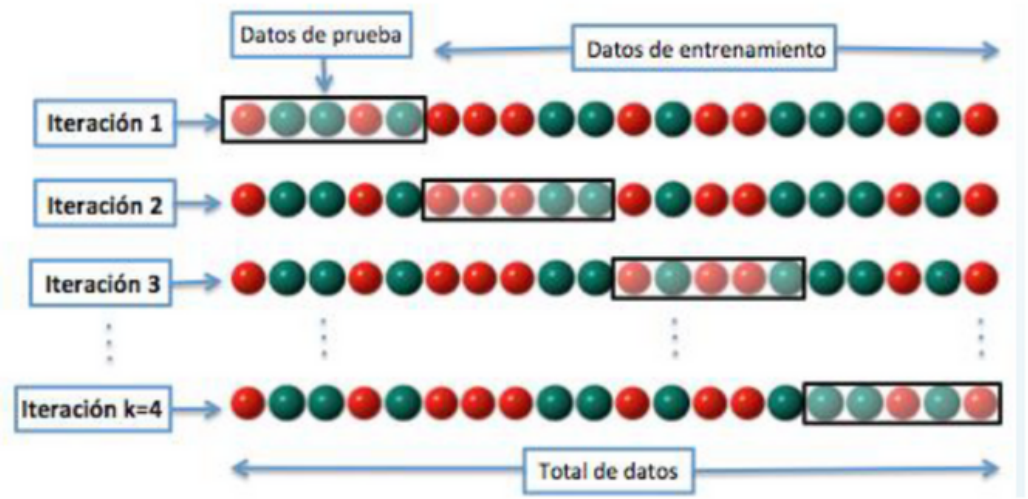
Son métricas que comparan los clusters obtenidos por el modelo con las etiquetas de clase verdaderas, aunque estas últimas no estén disponibles en muchos casos.

La evaluación de modelos en aprendizaje automático es esencial para comprender su rendimiento y generalización a nuevos datos. La selección de métricas de evaluación adecuadas depende del tipo de problema que se esté abordando y las características específicas del conjunto de datos.

1. Capacidad del Modelo

La validación cruzada es una técnica fundamental en el aprendizaje automático que se utiliza para evaluar el rendimiento de un modelo de manera robusta y evitar el sobreajuste. En lugar de depender de un solo conjunto de datos de entrenamiento y prueba, la validación cruzada divide los datos en múltiples conjuntos, permitiendo una evaluación más precisa del modelo.

La validación cruzada implica dividir el conjunto de datos en k pliegues (o "folds"). Luego, el modelo se entrena k veces, utilizando k-1 pliegues para entrenamiento y el pliegue restante para prueba en cada iteración. Esto se repite k veces, y los resultados se promedian para obtener una medida más robusta del rendimiento del modelo.



- **Pasos para Implementar la Validación Cruzada**

1. División de Datos: Divide el conjunto de datos en k pliegues.

2. Iteración del Modelo: Inicia un bucle que va desde 1 hasta k.
En cada iteración, selecciona k-1 pliegues para entrenar el modelo.

3. Evaluación del Modelo: Utiliza el pliegue restante para evaluar el rendimiento del modelo.

4. Promedio de Resultados: Repite el proceso k veces, registrando el rendimiento en cada iteración. Por último, calcula el promedio de los resultados para obtener una métrica global de rendimiento.

- **Ventajas de la Validación Cruzada**

Mejora de la generalización

Al evaluar el modelo en diferentes subconjuntos de datos, la validación cruzada proporciona una evaluación más precisa del rendimiento general del modelo.

Reducción del Sobreajuste

Al utilizar múltiples divisiones de datos, se reduce la probabilidad de que el modelo se ajuste demasiado a un conjunto de datos específico.

Mayor Utilización de Datos

Aprovecha al máximo los datos disponibles para entrenamiento y prueba en múltiples combinaciones.

La validación cruzada es esencial para obtener una evaluación más fiable del rendimiento del modelo en comparación con una simple división de datos en conjunto de entrenamiento y prueba. Al implementar esta técnica, los practicantes del aprendizaje automático pueden tomar decisiones más informadas sobre la capacidad predictiva de sus modelos.

2. Capacidad del Modelo

la capacidad de un modelo de adaptarse puede llegar a 2 extremos. La insuficiencia ocurre cuando el modelo no puede obtener un valor de error suficientemente bajo en el conjunto de entrenamiento.

El sobreajuste se produce cuando la brecha entre el error de entrenamiento y el error de prueba es demasiado grande.

- **Capacidad de generalización**

Generalización es la capacidad de un algoritmo de obtener un buen desempeño para entradas previamente no observadas.

El error de generalización o error de prueba, se define como el valór esperado del error en una nueva entrada y se estima al medir el rendimiento en un conjunto de pruebas.

Lo que se mide

$$\frac{1}{m^{(\text{train})}}||\boldsymbol{X}^{(\text{train})}\boldsymbol{w} - \boldsymbol{y}^{(\text{train})}||_2^2$$

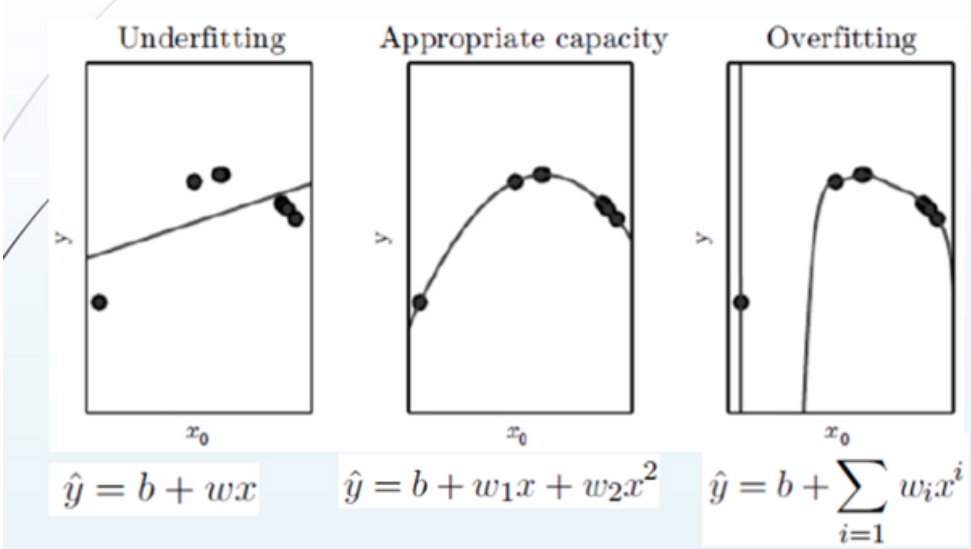
Lo que se quiere mejorar

$$\frac{1}{m^{(\text{test})}}||\boldsymbol{X}^{(\text{test})}\boldsymbol{w} - \boldsymbol{y}^{(\text{test})}||_2^2$$

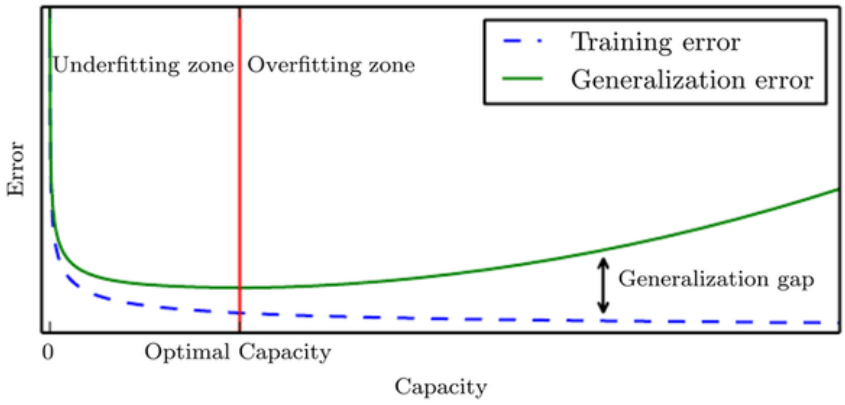
- **Sobreajuste**

El sobreajuste (overfitting) es un problema común en el aprendizaje automático que ocurre cuando un modelo se ajusta demasiado a los datos de entrenamiento, capturando tanto el ruido como los patrones reales.

La regresión polinomial es propensa al sobreajuste si no se controla adecuadamente. Veamos un ejemplo:



- **Capacidad vs Error**



3. Selección de Modelos

La selección del modelo adecuado es una etapa crítica en el proceso de construcción de un sistema de aprendizaje supervisado.

Existen varios algoritmos de aprendizaje supervisado, cada uno con sus propias fortalezas y debilidades. La selección del modelo adecuado depende del tipo de problema, la naturaleza de los datos y los objetivos específicos.

Conozcamos algunos aspectos relevantes:

Evaluación de Rendimiento

Diferentes algoritmos pueden funcionar mejor para diferentes conjuntos de datos. La selección del modelo se basa en una evaluación exhaustiva del rendimiento de cada algoritmo en el contexto específico del problema.

Algunos algoritmos pueden ser más eficientes en términos computacionales que otros. La selección del modelo debe tener en cuenta la capacidad computacional disponible y los requisitos de tiempo de ejecución.

Consideraciones Computacionales

- **Proceso de Selección de Modelos:**

1

División de Datos: Se divide el conjunto de datos en conjuntos de entrenamiento, validación y prueba. El conjunto de entrenamiento se utiliza para entrenar los modelos, el de validación para ajustar hiperparámetros y el de prueba para evaluar el rendimiento final.

2

Evaluación de Modelos: Se entrena cada modelo con el conjunto de entrenamiento y se evalúa su rendimiento en el conjunto de validación. Se utilizan métricas como precisión, recall, F1-score, etc., para comparar los modelos.

3

Ajuste de Hiperparámetros: Se ajustan los hiperparámetros del modelo para optimizar su rendimiento en el conjunto de validación. Esto puede incluir la selección de la tasa de aprendizaje, la profundidad del árbol, el número de capas en una red neuronal, etc.

3

Comparación y Selección: Se comparan los modelos según las métricas de rendimiento y se selecciona el modelo que mejor se ajuste al problema. También se realiza una evaluación final en el conjunto de prueba para validar la generalización del modelo.

- **Consideraciones en la Selección:**

Tipo de Problema

Problemas de clasificación, regresión, o incluso tareas más específicas como detección de anomalías, pueden requerir enfoques de modelado diferentes.

La naturaleza de las características, la presencia de datos desbalanceados, la cantidad de datos, entre otros, influyen en la elección del modelo.

Características del Conjunto de Datos

Interpretabilidad

Algunos modelos son más interpretables que otros. Dependiendo de los requisitos del problema, puede ser crucial comprender cómo toma decisiones el modelo.

La selección de modelos en el aprendizaje supervisado es un proceso iterativo y crucial. La elección adecuada impulsa el rendimiento del modelo y su capacidad para generalizar a nuevos datos. Se debe tener en cuenta la naturaleza del problema y las características específicas del conjunto de datos para tomar decisiones informadas.