

Ejercicio: Modelado Predictivo con Técnicas de Regresión

En este ejercicio, aplicarás diferentes técnicas de regresión para predecir una variable objetivo a partir de un conjunto de datos dado. A continuación, se detallan los pasos que debes seguir para resolver el ejercicio:

1. Exploración de Datos (EDA)

- Realiza una exploración inicial del conjunto de datos. Esto incluye la inspección de la estructura del dataset (número de filas, columnas, tipos de datos), la detección de valores nulos o atípicos (outliers), y el análisis descriptivo de las variables (medidas estadísticas básicas como media, mediana, desviación estándar, etc.).
- Visualiza las relaciones entre las variables mediante gráficos (distribuciones, diagramas de dispersión, matrices de correlación, etc.) para identificar posibles patrones o tendencias.
- Comenta cualquier observación relevante sobre la distribución de las variables, la presencia de colinealidad o cualquier otra característica que pueda influir en los modelos.

2. División del Dataset

- Divide el dataset en dos subconjuntos: **training** y **test**. Utiliza una proporción adecuada, como 80%-20% o 70%-30%, dependiendo del tamaño del dataset.
- Asegúrate de que la división sea aleatoria y que mantenga la representatividad de la variable objetivo en ambos subconjuntos.
- Justifica la elección de la proporción utilizada y evalúa si es necesario aplicar técnicas adicionales, como la estratificación, en función de la distribución de la variable objetivo.

3. Modelado con Regresión Lineal

- Construye un modelo de regresión lineal simple utilizando el subconjunto de training. Asegúrate de incluir la normalización de las variables si es necesario.
- Evalúa el desempeño del modelo en el subconjunto de test utilizando métricas adecuadas (MSE, RMSE, R^2 , etc.).
- Discute los resultados obtenidos y cualquier limitación del modelo de regresión lineal en este contexto.

4. Modelado con Regresión Ridge

- Aplica la técnica de regresión Ridge sobre el conjunto de datos de training. Recuerda que este método incluye una regularización que penaliza la magnitud de los coeficientes, reduciendo así el riesgo de sobreajuste.
- Ajusta el hiperparámetro de regularización (alpha) mediante validación cruzada para encontrar el valor óptimo.
- Evalúa el modelo en el conjunto de test y compara su rendimiento con el modelo de

regresión lineal.

5. Modelado con Regresión Lasso

- Implementa un modelo de regresión Lasso, que también incluye un término de regularización, pero que además puede conducir a la eliminación de algunas variables (coeficientes cero).
- Ajusta el hiperparámetro (α) utilizando validación cruzada para optimizar el modelo.
- Evalúa y compara el desempeño del modelo en el conjunto de test, tomando en cuenta la posible reducción en el número de variables significativas.

6. Comparación de Resultados

- Compara los resultados obtenidos de los tres modelos (Regresión Lineal, Ridge, y Lasso) utilizando las métricas de evaluación en el conjunto de test.
- Analiza las ventajas y desventajas de cada modelo en función de su rendimiento y complejidad.
- Discute cuál modelo sería más adecuado para el problema planteado y justifica tu elección.
- Si es pertinente, considera la interpretación de los coeficientes de cada modelo y su relación con las variables predictoras.