**EVALUACIÓN**

**Agosto de 2025**

**CRITERIOS DE EVALUACIÓN:**

**Comprensión y preparación de datos**

* Calidad del análisis exploratorio y cómo se han preparado los datos para el modelado.

**Implementación de modelos de regresión y su validación por medio de métricas de calidad**

* Correcta selección y entrenamiento de modelos de regresión.
* Uso apropiado de métricas para evaluar la calidad del modelo.

**Métodos de regularización**

* Implementación y justificación de métodos de regularización como Ridge y Lasso.

**Métodos de validación cruzada y boostrap**

* Uso efectivo de técnicas de validación cruzada para ajuste de hiperparámetros y evaluación del modelo.

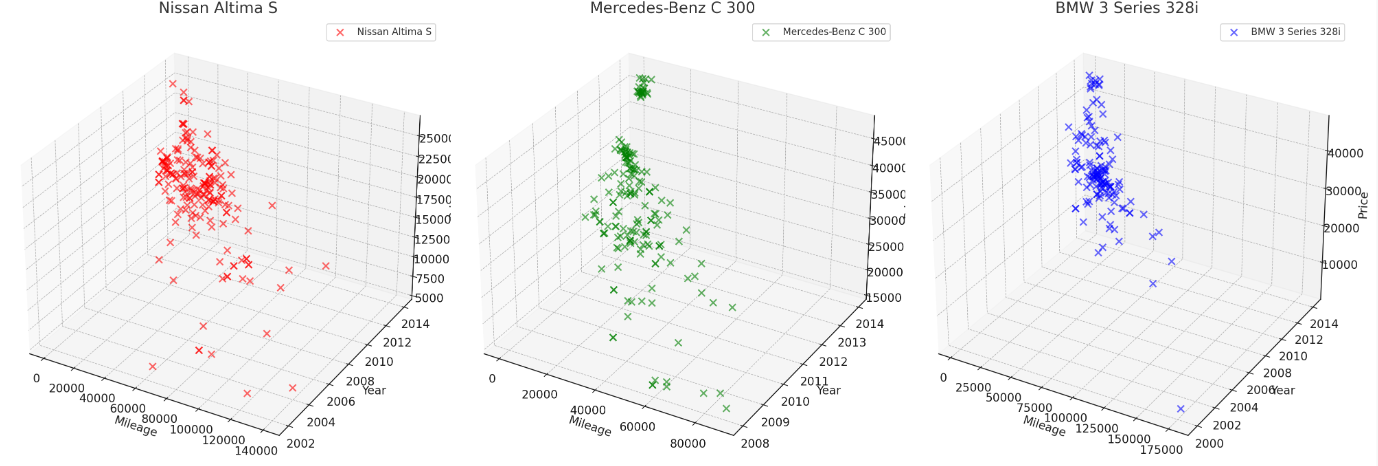
Los criterios para autoevaluación, coevaluación y heteroevaluación, se define a partir de la rúbrica anexa en la plataforma Moodle del curso

**PARTE 1 (30 pts)**

Desarrollar el modelo de regresión para predecir el precio de venta de vehículos de una de las marcas específicas en el conjunto de datos LondonCars2014. Los predictores claves a considerar son el kilometraje del vehículo y el año de fabricación. El rendimiento del modelo se evaluará en función del valor del Error Cuadrático Medio de la Raíz (RMSE). Cuanto menor sea el RMSE, mejor será el modelo

**1.A Selección de la Marca:** Elija una de las marcas y un modelo de automóviles disponibles en el conjunto de datos para centrar su análisis.

|  |  |
| --- | --- |
| **Marca** | **Model** |
| Nissan | Altima S |
| Mercedes-Benz | C 300 |
| BMW | 3 Series 328i |
| Lexus | RX 350 |
| Infiniti | G 37 |
| Mercedes-Benz | E 350 |
| Lexus | ES 350 |
| Mercedes-Benz | ML 350 |
| Honda | Civic LX |



**1.B Selección de Modelo**: Está en libertad de emplear cualquier método de regresión que considere apropiado para lograr el objetivo, incluyendo regularización.

**1.C Validación del Modelo:** Utilice técnicas de validación para asegurarse de que el modelo es robusto y generalizable a nuevos datos.

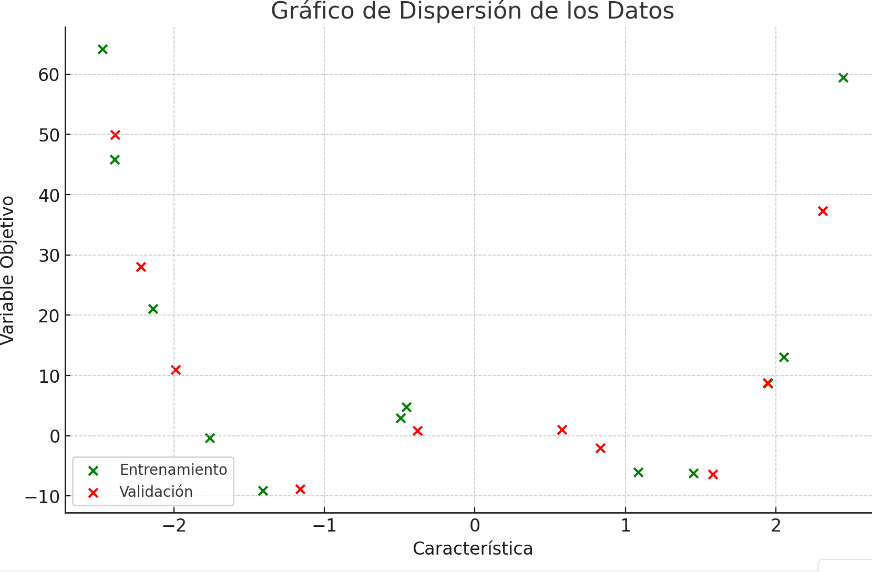
**1.D Presentación de Resultados**: Documente claramente todas las etapas del análisis, desde la exploración de datos hasta la evaluación del modelo, y presente sus resultados de una manera coherente y fácil de seguir.

**Entregables:**

* **Archivo de Notebook**: Proporcione un archivo de cuaderno Jupyter (**.ipynb**) o Google Colab, nombrado como **parte1.ipynb**. Este archivo debe contener las secciones 3.a, b y c y la Documentación de cada sección debe venir acompañada con una explique clara del enfoque adoptado, los resultados obtenidos y cualquier conclusión relevante

**PARTE 2 (35 pts)**

En la siguiente grafica tenemos los datos aleatorios, Desarrollar un modelo de regresión polinomial para predecir la variable objetivo. Validar el rendimiento del modelo utilizando técnicas de validación cruzada y bootstrap.



Los datos los descargar del Moodle y tiene el nombre de **part2xVal, part2yVal, part2xtrain, part2xtrain** y están distribuido de la siguiente manera:

* Conjunto de entrenamiento de características (x\_train): 12 observaciones
* Conjunto de validación de características (x\_val): 10 observaciones
* Conjunto de entrenamiento de etiquetas (y\_train): 12 observaciones
* Conjunto de validación de etiquetas (y\_val): 10 observaciones

**2.A Regresión y Validación b(15pts)**

* Implementar varis modelos de regresión polinomial de varios grados.
* Utilizar validación para evaluar el rendimiento del modelo en el conjunto de entrenamiento.
* Reportar métricas MSE y el MAE y el mejor modelo

**2.B Mejora del Modelo con Regularización (10pts)**

1. Experimentar con métodos de regularización Ridge o Lasso con los modelos.
2. Utilizar validación para encontrar los mejores parámetros de regularización.
3. Reporta e hiperparametro lamda que mejor resultado da junto con el polinomio

**2..C entregar el intervalo de confianza**

* Implementar el procedimiento de bootstrap para estimar la distribución MSE del modelo seleccionado en los dos pasos anteriores.
* Grafica los resultados del bootstrap para calcular un intervalo de confianza para esta métrica.

**Entregables:**

* **Archivo de Notebook**: Proporcione un archivo de cuaderno de Jupyter (**.ipynb**) o Google Colab, titulado **parte2.ipynb**. Este archivo debe estar estructurado en secciones que correspondan a los ítems 2.a 2.b y 2.c Cada sección debe incluir detalles sobre la selección del modelo empleado.

**PARTE 3 (35 pts)**

**Objetivo**: Desarrolle un modelo de regresión para predecir la popularidad de las canciones basándose en sus características de audio y metadatos. Utilice técnicas de regularización como Ridge (regularización L2) o Lasso (regularización L1) para prevenir el sobreajuste y mejorar la generalización del modelo. Aplique validación para evaluar el rendimiento del modelo y seleccionar los hiperparámetros óptimos de regularización. Use El dataset ***song\_data.csv***

**3.A Análisis exploratorio de los datos (10 puntos):**

Visualizar la relación entre las diferentes características y la variable objetivo ("song\_popularity"). Y verifica la correlación entre las características y la variable objetivo, defina que característica no va tomar como predictores y cuales si

**3.B. Construcción del Modelo de Regresión (15 puntos)**

**Nota importante:** Tienen completa libertad para escoger el método de validación más adecuado (ya sea mediante un conjunto de validación o validación cruzada) y para decidir la mejor manera de dividir los datos para el análisis. Siempre y cuando este muy argumentado

Desarrolle un modelo de regresión, que puede ser lineal múltiple o polinomial, enfocado en predecir la popularidad de las canciones. Es crucial evaluar el desempeño de los modelos generados utilizando métricas como el coeficiente de determinación (R²), el Error Cuadrático Medio Raíz (RMSE) y el Error Absoluto Medio (MAE).

Se espera que script de júpiter se evidencia todos los modelos entrenados y al final indicar cuál es el modelo seleccionado entregando el valor de métricas de calidad y los coeficientes del polinomio o de línea multivariable

**3.C. Validación y entrenamiento final, (15 pts)**

Con el modelo seleccionado vuelva a entrenar con todo el conjunto de datos y entregar los nuevos coeficientes y la métrica de calidad finales esperadas. Con este modelo es con el que realizare las pruebas de test con otro conjunto de datos

**Entregables:**

* **Archivo de Notebook**: Envíe un archivo de cuaderno de Jupyter (**.ipynb**) o Google Colab, titulado **parte3.ipynb**. Este archivo debe dividirse en secciones que aborden los ítems 3.a, 3.b y 3.c. En cada sección, asegúrese de explicar detalladamente tanto la selección de datos como los métodos de validación empleados para cada etapa del análisis.
* Agregar una última sección, donde indique claramente cuál es su modelo seleccionado y las métricas de calidad referidas a partir de la validación