




Regresión Polinómica

Se desea predecir el precio de venta de un auto en base a características como tipo de transmisión, kilometraje, y valor de mercado actual entre otras.


(a) Exploración de datos:

1. Cargar el archivo `car_data.csv`. Descartar las columnas correspondientes al nombre del auto y su dueño.
2. Definir los conjuntos de entrenamiento, testeo y validación en base a la información en la columna `Set`. : Una vez hecho este paso la información de esa feature no será de utilidad para la regresión.
3. Indicar la proporción de las variables categóricas representando probabilidades. : Si no sabe qué tipo de variable aleatoria es la categórica, deberá buscar dicha información.

(b) Pre-procesamiento:

1. Utilizar el comando `OneHotEncoder` (sklearn) para codificar las variables categóricas de más de dos clases como one-hot y las categóricas de dos clases como binarias.
2. Utilice el comando `PolynomialFeatures` (sklearn) para crear un mapa polinómico de orden 3 sobre las variables numéricas (exceptuando el precio de venta que será la variable a predecir).
3. Combinar y posteriormente normalizar las salidas de las operaciones anteriores. : Herramientas recomendadas son `ColumnTransformer` y `StandardScaler` (sklearn).

(c) Regresión lineal:

1. Utilizar `Pipeline` (sklearn) para combinar las operaciones anteriores con una regresión lineal. : Se recomienda utilizar `LinearRegression` (sklearn).
2. ¿Cuántos parámetros tiene el modelo? Explicar por qué.
3. Utilizar `display` para hacer un diagrama del sistema definido.
4. Entrenar el sistema descrito anteriormente.
5. Reportar el error cuadrático medio de entrenamiento y validación.

(d) Regularización

1. Combinar `Ridge` (sklearn) con `Pipeline` para implementar una regresión polinómica regularizada.
2. Entrenar una regresión regularizada para diferentes valores de $10^{-6} < \lambda < 10^{-2}$.
3. Graficar el error cuadrático medio de entrenamiento y validación en función de λ .
4. Reportar el λ que minimiza el error cuadrático medio de validación.
5. Reportar el error cuadrático medio de testeo para el λ obtenido anteriormente.