

EVALUACIÓN

03 AL 05 DE OCTUBRE

CRITERIOS DE EVALUACIÓN:

Comprensión y preparación de datos

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

Selección y Justificación del Spline

- Justificación de la elección de tipo de spline
- Justificación y metodología para la elección del número y la ubicación de nodos

Métodos de validación cruzada y Bootstrap

- Uso efectivo de técnicas de validación cruzada para ajuste de hiperparámetros y evaluación del modelo.
- Implementación de métodos de bootstrap para estimación de intervalos de confianza y robustez del modelo.

Selección y Justificación del Kernel:

- Correcta selección y entrenamiento de las características usadas para el kernel.
- Ha implementado la metodología adecuada para la selección del kernel.

Evaluación del Modelo:

- Se ha implementado metodologías de evaluación de modelos
- Ha utilizado métricas adecuada para evaluar el rendimiento

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

DATASET A UTILIZAR:

House Prices: Advanced Regression Techniques:

El conjunto de datos contiene 79 variables explicativas que abarcan casi todos los aspectos de las viviendas residenciales, como el tipo de material del techo, el número de habitaciones, el año de construcción, la presencia de ciertas instalaciones (piscina, garaje, etc.), entre otros.

Recuerda: No debemos aceptar ciegamente todos los atributos y sus interpretaciones, sino cuestionar y entender el contexto detrás de ellos.

Link de descargas: está en la página del curso y se llama data.raw, en él encuentra la descripción de cada columna

Nota: Debe hacer todo su proceso de modelado con los datos llamados train.csv. en la evaluación por parte del profesor se usaron unos datos de test

OBJETIVO:

El objetivo principal es predecir el precio de venta (**SalePrice**) de cada vivienda y tener el mejor modelo, tanto para spline como usando kernel

DESARROLLO:

1. Exploración de datos (10 puntos):

- Visualizar las relaciones entre las características y el precio de vivienda e identificar posibles relaciones no lineales
- Identificar y tratar datos faltantes si los hay
- Si es necesario usar variables categóricas
- Evaluar la normalización de características según el método a utilizar
- Dividir el conjunto de datos, según el método de validación a utilizar

2. Modelado con spline (30 puntos):

- Definir qué características va a utilizar para predecir el precio de venta, y aplicar regresión spline.
- Explorar el tipo de spline (por ejemplo, spline cúbico o natural)
- Ajustar el modelo y evaluar su rendimiento en el conjunto de prueba.
- Ajustar el número y la ubicación de los nodos según sea necesario.

3. Modelado con Regresión con Kernel (50 punto):

- Elegir un kernel a utilizar y el número de característica a utilizar
- Ajustar el modelo en el conjunto de entrenamiento, posiblemente usando validación cruzada para encontrar el mejor parámetro de ancho de banda ventana para el kernel.
- Evaluar el rendimiento del modelo.
- Ajustar el modelo según sea necesario, quizás probando diferentes kernels o ajustando parámetros.

4. Comparación y Evaluación (10 puntos):

- Comparar el rendimiento de los modelos de regresión spline y regresión con kernel.
- Usar métricas apropiadas como RMSE, MAE y R^2 para la evaluación.



ENTREGABLES:

Notebook código bien comentado y organizado que utilizó para su análisis y modelado adicional a la siguiente información:

1. **Análisis Exploratorio de Datos y pre-procesamiento:** Respectivas gráficas y descripciones para el análisis realizado. Explicar las consideraciones realizadas y los criterios para el pre-procesamiento y selección de datos
2. **Spline:** Tipo de spline seleccionado y razones para esta elección. Ecuaciones o representación del spline utilizado, Métricas de rendimiento en el conjunto de entrenamiento y validación según sea el caso
3. **kernel:** Tipo de kernel elegido y justificación. Decisión sobre parámetros clave, como el ancho de banda/ Ventana, metodología usada para encontrar este parámetro y código de la misma. Métricas de rendimiento en el conjunto de entrenamiento y validación según sea el caso
- 4 **Análisis de los dos modelos,** Comparar el rendimiento de los modelos de regresión spline y regresión con kernel