



Aplicación de Modelos de Regresión en Sklearn

Aplicación de Modelos de Regresión en Sklearn

Guía paso a paso para que un estudiante aplique modelos de regresión en Scikit-learn a una base de datos tipo CSV descargada de plataformas como Kaggle o UCI Machine Learning Repository:

Descargar y explorar los datos:

1 Descarga el archivo CSV de la plataforma deseada, como Kaggle o UCI ML Repository.

2 Explora el archivo CSV para comprender la estructura de los datos, las características disponibles y la variable objetivo que se desea predecir.

Preprocesamiento de datos:

1 Limpia los datos eliminando filas con valores faltantes o duplicados.

2 Divide los datos en características (X) y la variable objetivo (y).

División de datos:

1 Utiliza la función `train_test_split` de Scikit-learn para dividir los datos en conjuntos de entrenamiento y prueba. Esto permitirá evaluar el rendimiento del modelo en datos no vistos.



Selección de modelo:

1 Elige un modelo de regresión adecuado para tu problema. Algunas opciones comunes incluyen regresión lineal, regresión polinómica, regresión Ridge, regresión Lasso.

2 Importa el modelo correspondiente de Scikit-learn.

Entrenamiento del modelo:

1 Ajusta el modelo a los datos de entrenamiento utilizando el método fit.

2 Puedes ajustar diferentes hiperparámetros del modelo para mejorar su rendimiento utilizando la validación cruzada u otras técnicas de optimización de hiperparámetros.

Evaluación del modelo:

1 Utiliza métricas de evaluación adecuadas para medir el rendimiento del modelo en los datos de prueba. Algunas métricas comunes para la regresión incluyen el error cuadrático medio (MSE), el coeficiente de determinación (R^2) y el error absoluto medio (MAE).

2 Visualiza los resultados del modelo utilizando gráficos como diagramas de dispersión de las predicciones vs. los valores reales.

3 Al seguir estos pasos, un estudiante puede aplicar con éxito modelos de regresión en Scikit-learn a una base de datos tipo CSV descargada de plataformas de aprendizaje automático como Kaggle o UCI Machine Learning Repository.

