

Explorando la Regresión Logística con Regularización

A continuación, se muestran algunos ejemplos del Modelo de Regresión Logística, obtenidos de la página de ScikitLearn:

https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression

En total son 5 ejemplos, cada uno con su código fuente y su explicación. Deberás elegir al menos dos ejemplos y llevar a cabo las instrucciones que se piden. Finalmente se te solicitará realizar una [adaptación a un Problema del Mundo Real](#).

Nota. Por cada actividad se proporciona el enlace donde se encuentra el ejemplo

Ejemplo 1. Penalización L1 y Modelo disperso en la Regresión Logística

https://scikit-learn.org/stable/auto_examples/linear_model/plot_logistic_l1_l2_sparsity.html#sphx-glr-auto-examples-linear-model-plot-logistic-l1-l2-sparsity-py

En esta actividad, explorarás el efecto de la regularización L1 en modelos de regresión logística. La regularización L1 puede conducir a modelos dispersos, donde algunos coeficientes son exactamente cero. Ejecuta el código proporcionado y observa el patrón del modelo disperso en los coeficientes.

Instrucciones:

- Ejecuta el código proporcionado en el ejemplo.
- Modifica la fuerza de regularización y observa los cambios en el modelo disperso.
- Discute el impacto del modelo disperso en el rendimiento e interpretabilidad del modelo.
 - ¿Cómo afecta la dispersión a la precisión del modelo?
 - ¿Qué ventajas tiene un modelo disperso en términos de interpretabilidad?
 - ¿Cómo se compara la regularización L1 con la regularización L2?
 - ¿Qué impacto tiene la dispersión en la capacidad del modelo para generalizar?
 - ¿Cómo podrías ajustar la fuerza de regularización para mejorar el rendimiento del modelo?

Ejemplo 2. Trayectoria de Regularización de la Regresión Logística con L1

https://scikit-learn.org/stable/auto_examples/linear_model/plot_logistic_path.html#sphx-glr-auto-examples-linear-model-plot-logistic-path-py

En esta actividad, visualizarás la trayectoria de regularización de un modelo de regresión logística con penalización L1. La trayectoria de regularización muestra cómo cambian los coeficientes con diferentes niveles de regularización.

Instrucciones:

- Ejecuta el código proporcionado en el ejemplo.
- Grafica la trayectoria de regularización y analiza los cambios en los coeficientes.
- Experimenta con diferentes conjuntos de datos y niveles de regularización.
 - ¿Qué observas en la trayectoria de regularización cuando aumentas la fuerza de regularización?
 - ¿Cómo afectan los diferentes niveles de regularización a los coeficientes del modelo?
 - ¿Qué patrones de cambio en los coeficientes puedes identificar?
 - ¿Cómo podrías utilizar la trayectoria de regularización para seleccionar la fuerza de regularización óptima?
 - ¿Qué impacto tiene la regularización en la capacidad del modelo para generalizar?

Ejemplo 3. Fronteras de Decisión de la Regresión Logística Multinomial y One-vs-Rest

https://scikit-learn.org/stable/auto_examples/linear_model/plot_logistic_multinomial.html#sphx-glr-auto-examples-linear-model-plot-logistic-multinomial-py

En esta actividad, compararás las fronteras de decisión de la regresión logística multinomial y la regresión One-vs-Rest. Las fronteras de decisión nos ayudan a entender cómo el modelo separa las diferentes clases.

Instrucciones:

- Ejecuta el código proporcionado en el ejemplo.
- Visualiza las fronteras de decisión para ambos enfoques.
- Discute las diferencias en las fronteras de decisión y sus implicaciones.
 - ¿Qué diferencias observas en las fronteras de decisión entre los dos enfoques?
 - ¿Cómo afectan las fronteras de decisión al rendimiento del modelo?
 - ¿Qué ventajas tiene la regresión logística multinomial sobre la regresión One-vs-Rest?
 - ¿Cómo podrías mejorar las fronteras de decisión para una mejor clasificación?
 - ¿Qué impacto tienen las fronteras de decisión en la capacidad del modelo para generalizar?

Ejemplo 4. Regresión Logística Dispersa Multiclasificación en el Conjunto 20newsgroups

https://scikit-learn.org/stable/auto_examples/linear_model/plot_sparse_logistic_regression_20newsgroups.html#sphx-glr-auto-examples-linear-model-plot-sparse-logistic-regression-20newsgroups-py

En esta actividad, aplicarás regresión logística dispersa al conjunto de datos 20newsgroups. La regresión logística dispersa puede ser útil para conjuntos de datos de alta dimensión como los de texto.

Instrucciones:

- Ejecuta el código proporcionado en el ejemplo.
- Analiza el patrón de dispersión en los coeficientes.
- Experimenta con diferentes niveles de regularización y observa el impacto en el rendimiento del modelo.
 - ¿Cómo afecta la dispersión a la precisión del modelo en conjuntos de datos de texto?
 - ¿Qué ventajas tiene un modelo disperso en términos de interpretabilidad en conjuntos de datos de texto?
 - ¿Cómo se compara la regularización L1 con la regularización L2 en conjuntos de datos de texto?
 - ¿Qué impacto tiene la dispersión en la capacidad del modelo para generalizar en conjuntos de datos de texto?
 - ¿Cómo podrías ajustar la fuerza de regularización para mejorar el rendimiento del modelo en conjuntos de datos de texto?

Ejemplo 5. Clasificación MNIST Usando Regresión Logística Multinomial + L1

https://scikit-learn.org/stable/auto_examples/linear_model/plot_sparse_logistic_regression_mnist.html#sphx-glr-auto-examples-linear-model-plot-sparse-logistic-regression-mnist-py

En esta actividad, clasificarás dígitos del conjunto MNIST usando regresión logística multinomial con regularización L1. MNIST es un conjunto de datos popular para tareas de clasificación de imágenes.

Instrucciones:

- Ejecuta el código proporcionado en el ejemplo.
- Evalúa el rendimiento del modelo en el conjunto de datos MNIST.
- Experimenta con diferentes niveles de regularización y observa los cambios en el rendimiento del modelo.
 - ¿Cómo afecta la regularización L1 al rendimiento del modelo en la clasificación de imágenes?
 - ¿Qué ventajas tiene la regularización L1 en términos de interpretabilidad en la clasificación de imágenes?
 - ¿Cómo se compara la regularización L1 con la regularización L2 en la clasificación de imágenes?
 - ¿Qué impacto tiene la regularización L1 en la capacidad del modelo para generalizar en la clasificación de imágenes?
 - ¿Cómo podrías ajustar la fuerza de regularización para mejorar el rendimiento del modelo en la clasificación de imágenes?

Adaptación a un Problema del Mundo Real

En esta actividad, adaptarás los ejemplos de regresión logística a un problema del mundo real. Elige un conjunto de datos relacionado con un tema actual (por ejemplo, sentiment analysis, disease prediction) y aplica regresión logística con cambios mínimos en el código.

Instrucciones:

- Elige un conjunto de datos relevante para un problema del mundo real.
- Modifica el código proporcionado para ajustarlo al nuevo conjunto de datos.
- Analiza el rendimiento del modelo y discute las implicaciones de tus hallazgos.

Algunos Conjuntos de Datos para Regresión Logística

Dataset Name	Source	Type	Description	Application
Logistic Regression Dataset	Kaggle	Clasificación binaria	Datos linealmente separables ideales para practicar regresión logística básica.	Problemas de clasificación simples, como detección de fraude o diagnóstico médico binario.
Estimation of Obesity Levels	UCI Repository	Clasificación multiclas	Datos sobre hábitos alimenticios y condición física de personas en México, Perú y Colombia.	Predicción de niveles de obesidad, ideal para temas de salud pública.
Student Performance	UCI Repository	Clasificación / Regresión	Datos sobre el rendimiento académico de estudiantes de secundaria.	Predicción de éxito escolar o abandono académico.
Online Retail II	UCI Repository	Clasificación / Clustering / Series de tiempo	Transacciones reales de una tienda en línea durante dos años.	Predicción de comportamiento de compra, segmentación de clientes.