

Lectura 5

Implementación de Regresión Lineal y Logística usando Scikit-Learn

SI3015 - Fundamentos de Aprendizaje Automático

2026

1. Ejercicio en clase:

Descargar uno de estos dos datasets:

Link dataset de películas

Link dataset del Titanic

Nota: también puede probar con un conjunto de datos propio en el que pueda aplicar las actividades a continuación.

Cargar el conjunto de datos y explorelo gráficamente para realizar las siguientes actividades:

1.0.1. Regresión Lineal

- Determine la columna a la cual aplicarle una regresión lineal.
- Divida el dataset en entrenamiento y prueba.
- Grafique el conjunto de entrenamiento y prueba. Diferencielos con colores.
- Defina los pipelines con Ridge y Lasso.
- Defina las distribuciones de parámetros sobre las que hará la búsqueda.
- Aplique búsqueda aleatoria y cross-validation.
- Entrene los modelos (Ridge y Lasso).
- Obtenga los mejores parámetros de Ridge y Lasso.
- Obtenga el R y el MAE (Mean Absolute Error) para Ridge y Lasso.
- Grafique el modelo predicho para Ridge y Lasso.

1.0.2. Regresión Logística

- Determine la columna a la cual aplicarle una regresión logística. Los datos deben tener tendencia a clasificarse en dos.
- Divida el dataset en entrenamiento y prueba.
- Defina el pipeline.
- Defina las distribuciones de parámetros.
- Aplique búsqueda aleatoria y cross-validation.
- Entrene el modelo.
- Obtenga los mejores parámetros.
- Obtenga la accuracy y el f1-score de prueba.
- Grafique el modelo predicho.
- Grafique la matriz de confusión.

2. Sugerencias:

- Cerciorese de realizar la mayor limpieza, exploración y transformación posible del conjunto de datos. Recuerde que un conjunto de datos suficientemente limpio y transformado aumenta significativamente la probabilidad de obtener muy buenos resultados al aplicar el modelo de aprendizaje automático.
- Recuerde basarse en los ejemplos `ej_regresionLineal.ipynb` y `ej_regresionLogistica.pdf` para realizar el ejercicio.