

CURSO: MACHINE LEARNING CON PYTHON

TRABAJO PRÁCTICO MÓDULO 1

OBJETIVO

Este trabajo tiene como objetivo evaluar el dominio de los contenidos de las unidades 1 a 4 del módulo 1

EJERCICIOS

1. Regresión lineal

Un restaurant tiene locales en varias ciudades y está buscando expandirse. Para determinar las ciudades en las que el negocio puede ser rentable, se desea obtener un modelo de regresión lineal para estimar las ganancias del restaurant en función del número de habitantes de la ciudad. El modelo debe determinarse utilizando data obtenida en las ciudades donde el restaurant ya tiene operaciones.

Antes de desarrollar el Jupyter notebook, inspeccione los archivos `datos1_entrenamiento.csv` y `datos1_validacion.csv`. Cada archivo tiene dos columnas etiquetadas `x`, `y`, que corresponden al número de habitantes en decenas de miles (variable de entrada) y a las ganancias en decenas de miles de USD (variable de salida), respectivamente.

Desarrolle un Jupyter notebook que

- 1.1. En una celda de texto, muestre los nombres de los participantes en negrita.
- 1.2. Cargue el archivo `datos1_entrenamiento.csv`, y cree los datos de entrenamiento como dos arreglos 2D de 1 columna que contengan la variable de entrada (`x`) y la variable de salida (`y`), respectivamente.
- 1.3. Cargue el archivo `datos1_validacion.csv`, y cree los datos de validación como dos arreglos 2D de 1 columna que contengan la variable de entrada (`x`) y de salida (`y`), respectivamente.
- 1.4. Realice un gráfico de dispersión y vs. `x` de los datos de entrenamiento

1.5. Ajuste un modelo de regresión lineal con los datos de entrenamiento

1.6. Realice, en una misma figura, un gráfico de dispersión y vs. x de los datos de entrenamiento y un gráfico de línea de las salidas obtenidas al evaluar el modelo en los datos de entrenamiento

1.7. Realice, en una misma figura, un gráfico de dispersión y vs. x de los datos de validación y un gráfico de línea de las salidas obtenidas al evaluar el modelo en los datos de validación

1.8. Para el modelo de regresión lineal, calcule el error absoluto medio, el error cuadrático medio, el error absoluto mediano, la puntuación de varianza explicada y la puntuación R^2 sobre los datos de validación

1.9. Ajuste un modelo de regresión lineal ridge con los datos de entrenamiento

1.10. Para el modelo de regresión lineal ridge, calcule el error absoluto medio, el error cuadrático medio, el error absoluto mediano, la puntuación de varianza explicada y la puntuación R^2 sobre los datos de validación

1.11. En una celda de texto, escriba un breve análisis comparativo de los resultados obtenidos en los apartados 1.7 y 1.9

2. Regresión logística

Se desea implementar un modelo de clasificación binaria basado en regresión logística, para estimar la posibilidad de un solicitante de ser admitido(a) como estudiante en una universidad, en base a las calificaciones obtenidas en dos exámenes. El modelo debe entrenarse con data histórica de solicitantes previos.

Antes de desarrollar el Jupyter notebook, inspeccione los archivos `datos2_entrenamiento.csv` y `datos2_validacion.csv`. Cada archivo tiene tres columnas etiquetadas `Examen1`, `Examen2`, `Decision`, que corresponden, respectivamente, a la calificación obtenida en el examen 1, la calificación obtenida en el examen 2, y la decisión (0 cuando el solicitante no fue admitido y 1 cuando fue admitido)

Desarrolle un Jupyter notebook que

- 2.1. En una celda de texto, muestre los nombres de los participantes en negrita.
- 2.2. Cargue el archivo datos2_entrenamiento.csv en un dataframe de Pandas
- 2.3. Cargue el archivo datos2_validacion.csv en un dataframe de Pandas
- 2.4. Muestre las primeras 5 filas del dataframe con los datos de entrenamiento
- 2.5. Despliegue la descripción (número de valores, información estadística y valores mínimo y máximo) para las columnas Examen1 y Examen2 del dataframe de los datos de entrenamiento
- 2.6. Realice un gráfico que muestre en un plano, calificación del Examen1 vs calificación del Examen2, los puntos del conjunto de entrenamiento. Utilice marcadores diferentes para los puntos correspondientes las clases 0 y 1 (Sugerencia: separe los puntos en dos vectores, y luego grafique ambos vectores en la misma figura)
- 2.7. Determine un modelo de clasificación basado en regresión logística, utilizando los datos de entrenamiento
- 2.8. Determine y grafique la matriz de confusión sobre los puntos de validación
- 2.9. En una celda de texto, escriba un breve comentario sobre la matriz de confusión

INFORMACIÓN IMPORTANTE

- El trabajo puede realizarse individualmente, o en grupos de 2 participantes
- Debe entregarse un archivo comprimido con los Jupyter notebooks desarrollados
 - El archivo comprimido debe generarse con la aplicación 7-Zip, que puede descargarse gratuitamente desde <https://www.7-zip.org/>
 - El nombre del archivo comprimido debe tener el formato TP_Modulo1_Apellidos. Por ejemplo, TP_Modulo1_Canelon o TP_Modulo_Barreto_Canelon.
 - El nombre de cada Jupyter notebook debe tener el formato

Ejer#_Apellidos.ipynb. Por ejemplo, Ejer1_Canelon.ipynb o Ejer1_Barreto_Canelon.ipynb

- El archivo comprimido debe subirse a la plataforma
- La fecha límite de entrega es el 17/04/2024, a las 23:59

CRITERIOS DE EVALUACIÓN

- La calificación total del trabajo estará en función del número de incisos implementados correctamente. Si algún inciso no funciona de manera correcta o genera un error en el Jupyter notebook, se restarán puntos del total correspondiente a ese inciso