

Proyecto de investigación 1

Regresión

1 OBJETIVO:

Realizar un proyecto de investigación de regresión lineal y logística con un reporte documentando los hallazgos y el código empleado para llegar a los resultados.

2 INSTRUCCIONES:

2.1 OBTENER LA DATA

Descargar los siguientes 2 datasets:

- a) Wine Quality: <https://archive.ics.uci.edu/ml/datasets/wine+quality>
- b) Dry Bean: <https://archive.ics.uci.edu/ml/datasets/Dry+Bean+Dataset>

2.2 ANALIZAR LA DATA

Leer el tipo de datos con los cuales se compone cada dataset. Mencionar los tipos de datos que se cuenta, valores, significado y si se encuentran anomalías en estas. Indicar cuáles son las variables independientes (features) y cuál es la variable dependiente (target).

Se puede realizar gráficos (boxplot, scatterplot, pairplot, histogramas) y describir lo que se encuentra.

Se puede realizar un análisis de correlación de las variables.

2.3 REGRESIÓN

Utilizar los modelos de regresión logística y lineal según corresponda. Analizar el resultado empleando las principales métricas de evaluación. R^2 para la regresión lineal; matriz de confusión, recall, F1, etc para la regresión logística.

2.4 PRUEBAS Y VARIANTES

Realizar cambios en el procesamiento de la data y en los modelos. Documentar las observaciones y análisis de los cambios observados en el resultado de la predicción.

Algunos ejemplos de pruebas:

- Descartar los valores outliers (por ejemplo, mayor a 3 desviaciones estándar)
- Realizar una selección de columnas (features) y compararlo contra un modelo sin selección.
- Estandarizar los valores (por ejemplo, el StandardScaler de scikit-learn)
- Agregar regularización L1 (Lasso) o L2 (Ridge)
- Cambiar los grados del polinomio y ver cuándo empieza a hacer overfitting

3 REPORTE

Preparar un reporte con las siguientes partes:

- Bases teóricas de los modelos de machine learning usados.
- El análisis descriptivo de los datasets.
- El preprocesamiento realizado.
- Las técnicas de regresión empleadas, las métricas obtenidas y el significado.
- Las pruebas y experimentos adicional y los efectos de estos sobre las métricas en los modelos.
- Concluir con sus observaciones más relevantes.

Utilizar correctamente las referencias en formato APA para todas las fuentes empleadas para el trabajo, incluyendo porciones de código tomadas de otros.

4 HERRAMIENTAS

La herramienta para realizar el proyecto es libre. Sin embargo, se sugiere emplear Python con las librerías vistas en clase.

5 ENTREGA

- Fecha límite: domingo 25 de septiembre 2022 a las 11:59pm
- Informe: Reporte del proyecto en PDF
- Código: Archivo Zip cargado en el BlackBoard

6 PARTICIPACIÓN INDIVIDUAL

Adicional a la entrega del reporte grupal, se realizará una encuesta anónima para calificar la participación de cada uno de los integrantes del grupo y obtener una evaluación individual.

La nota final será 70% nota grupal y 30% nota individual