

Electiva Aprendizaje automático
Maestría en Gestión de Tecnologías de Información y Conocimiento
Facultad de Ingeniería
Universidad de Nariño
Guía 1
Python y Pandas

1. Lea únicamente los valores numéricos del archivo "Adult_Parte1.xlsx" y guárdelos en un dataframe de Pandas llamado X.
2. Obtenga el tamaño de 'X' en cantidad de filas y columnas.
3. Calcule el valor mínimo de cada columna de 'X'.
4. Ahora obtenga el máximo, pero por cada fila.
5. Realice una escala lineal de 'X' por columna. Para esto deberá a cada valor, restarle el mínimo de la columna y dividirlo por la diferencia entre el máximo y el mínimo de la columna correspondiente
6. Reordene aleatoriamente las filas de 'X'.
7. Separe la última columna del conjunto de datos y almacénelo en una variable llamada 'Y'. 'X' deberá contener el resto de las columnas.
8. Separe el 80% de las filas de 'X' en una variable llamada 'X_train' y el 20% restante en 'X_test'. Haga lo mismo con 'Y' manteniendo el orden obtenido del punto 6.
9. Transponga la matriz 'X'. Tenga en cuenta que si transpone antes de la escala lineal no seguirá obteniendo el mismo resultado en los Ejercicios 5, 6, 7 y 8.
10. Defina un vector 'prediction' del mismo tamaño que 'Y_test' cuyo contenido sean únicamente ceros y unos ubicados aleatoriamente.
11. Considerando que el vector del Ejercicio 10 es el resultado de clasificar los datos de 'X_train' con un modelo predictivo, obtener el número de coincidencias con 'T_test'.
12. Construir una matriz de confusión en base al vector 'T_test' y 'prediction'. Puede revisar https://en.wikipedia.org/wiki/Confusion_matrix para ayudarse en los cálculos.

Actividad

13. Calcule la métrica de precisión de cada una de las clases con los valores obtenidos en la matriz de confusión.
14. Calcule la métrica de recall de cada una de las clases con los valores obtenidos en la matriz de confusión.
15. Calcule la métrica f1-score de cada una de las clases con los valores obtenidos en la matriz de confusión.
16. Calcule el accuracy del modelo.
17. Consultar la fórmula de la métrica accuracy_balanced y calcularla de acuerdo con la matriz de confusión obtenida.