

Universidad Nacional Autónoma de México

FACULTAD DE CIENCIAS

REPORTE PRACTICA 3

Inteligencia Artificial

Emilio Francisco Sánchez Martínez - 115004199

Flores Rojas Tenoch Itzin - 318027641

Pintor Muñoz Pedro Joshue - 420053796

Diciembre 2023

1. Regresión:

Para este ejercicio

1. Cargamos el conjunto de datos de diabetes
2. Separamos los datos en conjuntos de entrenamiento y evaluación
3. Instanciamos y entrenamos un modelo de regresión lineal
4. Realizamos predicciones en el conjunto de evaluación
5. Evaluamos el modelo utilizando el Error Cuadrático Medio (MSE) y el coeficiente de determinación R2.

El MSE mide el promedio de los errores al cuadrado entre las predicciones y los valores reales. Cuanto menor sea el MSE, mejor será el modelo. El coeficiente de determinación R2 indica la proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes. Mientras mas cercano sea el R2 al 1 significa que mejor es el modelo.

2. Clasificación:

1. Realizamos la separación del conjunto de datos en conjuntos de entrenamiento y evaluación mediante la función *train_test_split*. El 80 % de los datos se los asignamos al conjunto de entrenamiento, mientras que el 20 % restante se lo dimos al conjunto de evaluación.
2. Entrenamos un modelo de Perceptrón utilizando los datos de entrenamiento. Ajustamos la tasa de aprendizaje y el número máximo de iteraciones de acuerdo con consideraciones adecuadas para el conjunto de datos.
3. Entrenamos un modelo de Árbol de Decisión utilizando los datos de entrenamiento. Ajustamos nuestro modelos para capturar patrones complejos en los datos.
4. Entrenamos un modelo de k-NN utilizando los datos de entrenamiento. Seleccionamos un valor apropiado para el parámetro 'k' que determina la cantidad de vecinos más cercanos a tener en cuenta durante la clasificación o regresión.
5. Por ultimo realizamos la evaluación de los datos usando *classification report* y comparamos los resultados de los 3 modelos utilizados. Para nuestros modelos, el mejor de los tres fue el método de k-vecinos más cercano, sus resultados son los más cercanos a 1, aunque Perceptron también daba resultados cercanos a 1, los de K-NN son más cercanos.

3. Agrupamiento:

Generamos un conjunto de datos para agrupamiento utilizando la función *make_blobs* con 1000 datos y 3 centroides. Posteriormente, aplicamos el algoritmo de k-Medias para agrupar los datos en tres clusters. El algoritmo ha asignado cada punto de datos al cluster cuyo centroide es el más cercano, minimizando la suma de los cuadrados de las distancias.

Creamos una visualización para representar los datos generados y los grupos obtenidos mediante el algoritmo. Cada punto de datos tiene un color de acuerdo con el cluster al que se asigno.

Para el algoritmo es importante tener en cuenta que la efectividad del algoritmo dependerá de la distribución real de los datos y de la elección del número de clusters. En este caso, al haber generado intencionalmente los datos con tres centroides, el algoritmo fué de identificar y asignar los puntos a estos grupos.