



Instituto Politécnico Nacional

Escuela Superior de Cómputo

Profesor: García Floriano Andrés

Alumno: Hernández Jiménez Erick Yael

Patiño Flores Samuel

Robert Garayzar Arturo 5BV1

Practica 6: Clasificadores de distancia mínima y

1NN

Índice

1. Introducción	2
1.1. Clasificador de Distancia Mínima	2
1.2. Clasificador 1-Nearest Neighbor (1-NN)	2
2. Desarrollo	2
2.1. Clasificador de distancia mínima	2
2.1.1. Resultados	2
2.2. 1NN (One Nearest Neighbor)	3
2.2.1. Resultados	4
3. Conclusiones	5
3.1. Hernández Jiménez Erick Yael	5
3.2. Patiño Flores Samuel	5
3.3. Robert Garayzar Arturo	5

1. Introducción

1.1. Clasificador de Distancia Mínima

El clasificador de distancia mínima funciona calculando la distancia entre una muestra nueva y el centroide de cada clase conocida en el conjunto de entrenamiento. Los centroides se calculan como el promedio de los puntos de cada clase, lo cual representa un "punto medio" de cada clase. La clasificación de la muestra nueva se realiza asignándola a la clase cuyo centroide esté más cerca, usando una métrica de distancia (generalmente la distancia Euclidiana). Este clasificador es simple y eficiente cuando las clases son compactas y separables, pero su precisión disminuye si los datos presentan mucha variabilidad o si las clases tienen formas complejas.

Ventajas y desventajas:

- **Ventajas:** Simplicidad, fácil de implementar y rápido de ejecutar.
- **Desventajas:** No funciona bien en datos con clases no compactas o que tienen formas complejas y distribuciones variadas.

1.2. Clasificador 1-Nearest Neighbor (1-NN)

El clasificador 1-NN asigna una clase a una muestra nueva encontrando el punto de entrenamiento más cercano a ella. En lugar de calcular el centroide de cada clase, 1-NN busca el vecino más próximo (el punto de entrenamiento más cercano) y asigna la clase de este vecino a la muestra nueva. Se basa en el supuesto de que las muestras cercanas en el espacio de características probablemente pertenezcan a la misma clase. Este método es un caso particular de k-Nearest Neighbors (k-NN) con $k=1$.

Ventajas y desventajas:

- **Ventajas:** Puede adaptarse a datos con formas de clases complejas, ya que no asume una forma específica para cada clase.
- **Desventajas:** Sensible a ruido, lento en grandes volúmenes de datos, y puede ser menos preciso si los datos están muy dispersos o si hay variabilidad en las distancias entre clases.

2. Desarrollo

Para el desarrollo de esta práctica se implementó el clasificador de distancia mínima y 1NN por medio de clases en Python.

2.1. Clasificador de distancia mínima

Para este clasificador se utilizaron dos métodos de la clase: `fit` y `predict`.

- `fit`: Este método primero define las clases que tenemos de entrada, tomando como referencia nuestra entrada `Y` y los valores únicos de esta. Una vez definidas las clases, se calcula el centroide por cada clase utilizando el método `mean` de `numpy`.
- `predict`: Una vez que ya tenemos nuestro modelo entrenado, podemos hacer la predicción de datos de entrada. Para esto, calculamos la distancia a todos los centroides y tomamos el centroide más cercano.

2.1.1. Resultados

Aplicación de Hold-Out

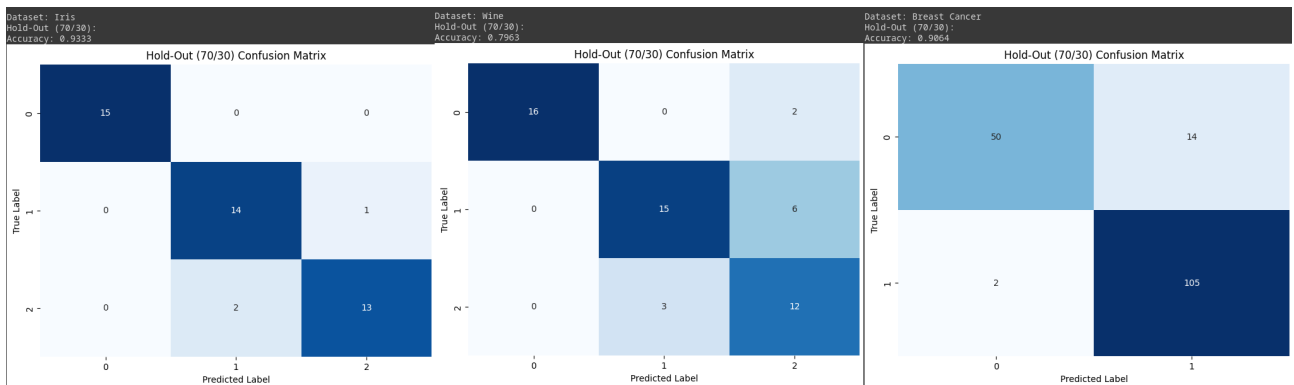


Figura 1: Matriz de confusión Hold Out

Aplicación de 10-Fold

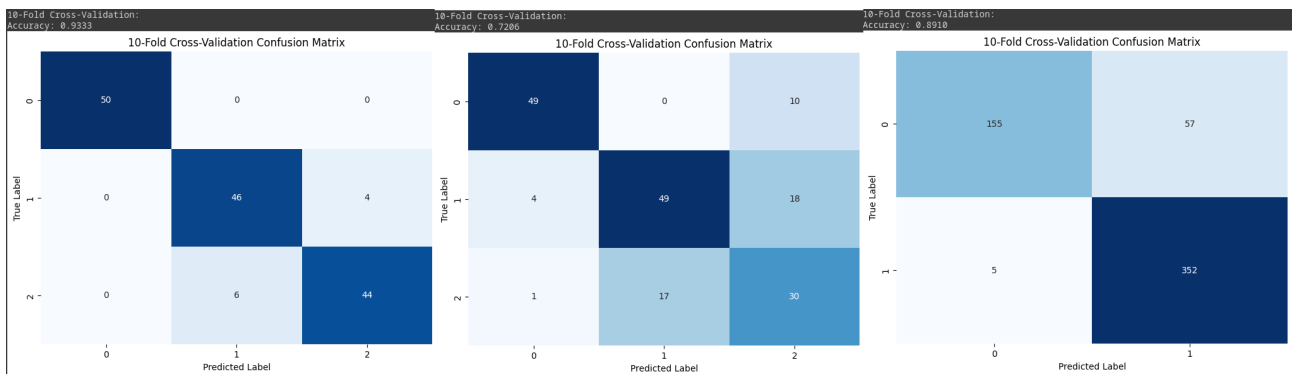


Figura 2: Matriz de confusión 10-Fold

Aplicación de Leave-One-Out

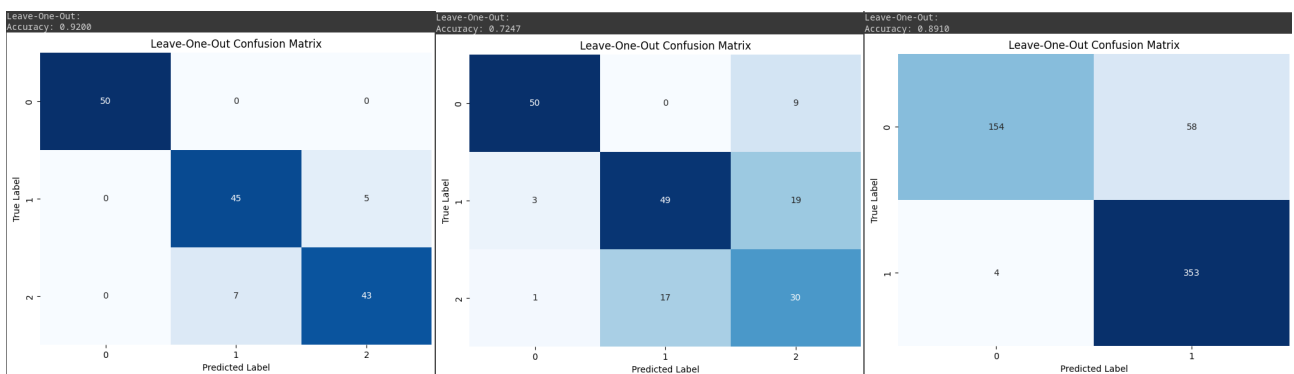


Figura 3: Matriz de confusión Leave-One-Out

2.2. 1NN (One Nearest Neighbor)

Como se ha mencionado, 1NN es un caso particular de KNN, por lo que su implementación es similar.

- **fit:** Se guardan las características junto con su etiquetado.
- **predict:** Se calculan las distancias del punto dado a todos los puntos cargados previamente, y se etiqueta conforme al punto más cercano al punto ingresado.

2.2.1. Resultados

Aplicación de Hold-Out

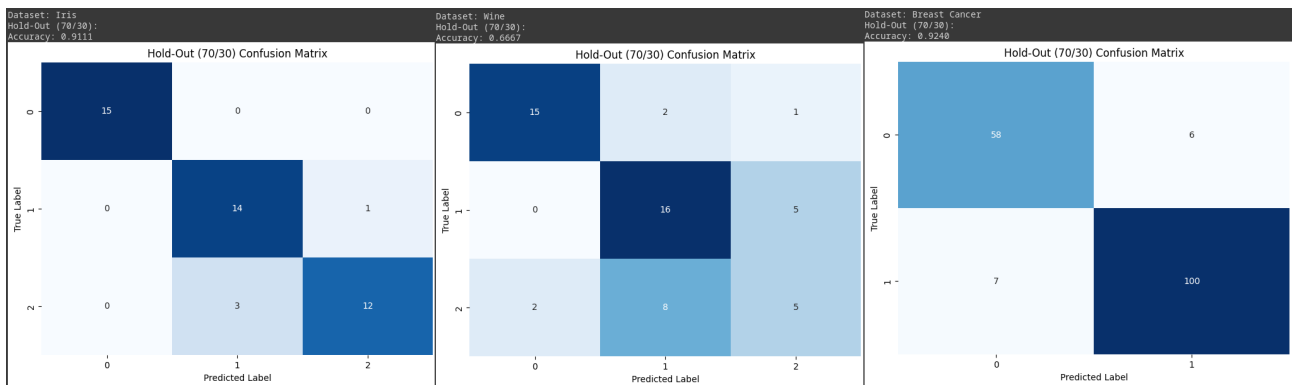


Figura 4: Matriz de confusión con 1NN-Hold-Out

Aplicación de 10-Fold

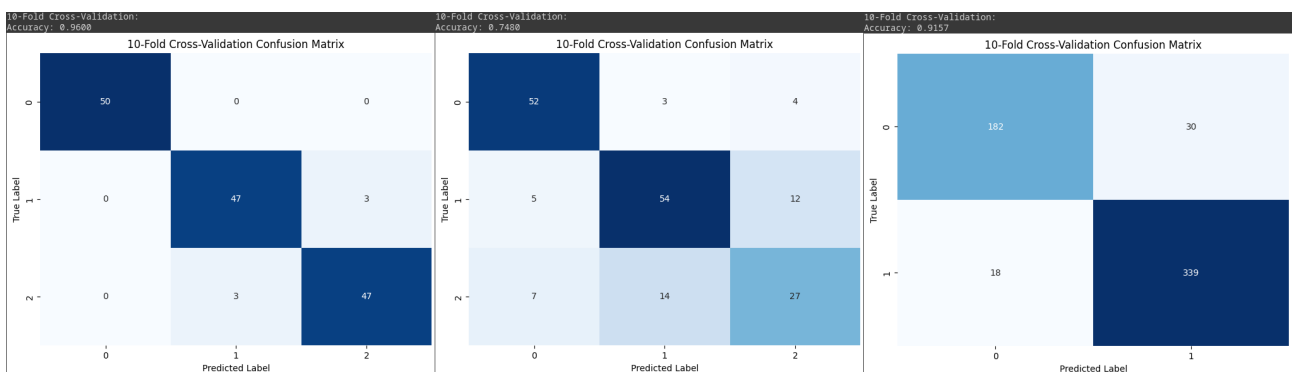


Figura 5: Matriz de confusión con 1NN-10-Fold

Aplicación de Leave-One-Out

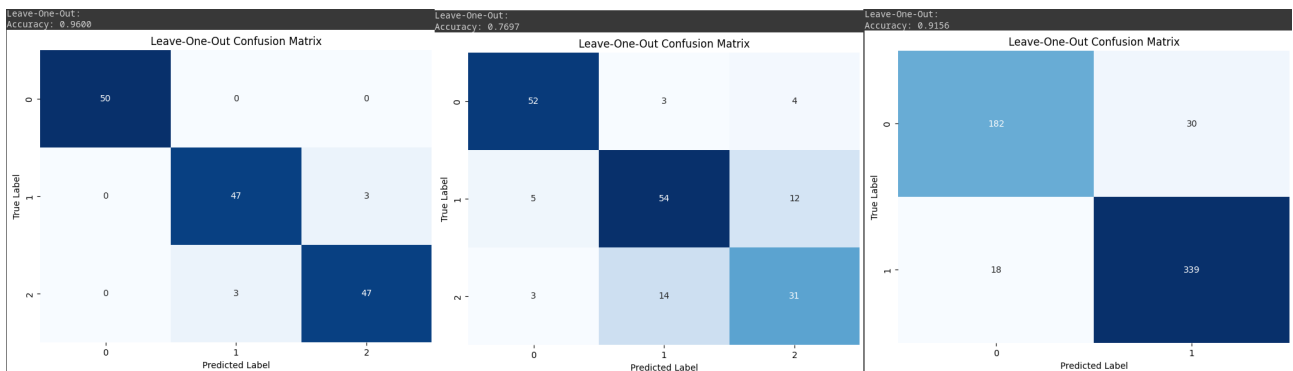


Figura 6: Matriz de confusión con 1NN-Leave-One-Out

3. Conclusiones

3.1. Hernández Jiménez Erick Yael

Con los resultados anteriormente mostrados, se puede apreciar que el modelo 1NN (1 Nearest Neighbor) predice con un buen resultado de precisión mínimo de 72% mientras se elija adecuadamente el método de selección de las instancias de prueba y de entrenamiento. Cabe mencionar que el método Hold-Out es el que muestra mejores resultados, ya que este ofreció predicciones promedio del 87.86% de precisión; por lo cual la elección del método de partición del dataset genera un gran impacto en la efectividad del modelo.

3.2. Patiño Flores Samuel

Los resultados mostraron que, aunque ambos clasificadores son efectivos en la clasificación de datos, el 1-NN tiende a ofrecer una mayor precisión en situaciones donde los datos son más complejos y presentan variabilidad. Esto se debe a que el 1-NN considera el punto más cercano en el conjunto de entrenamiento, lo que le permite adaptarse mejor a la distribución de los datos. Por otro lado, el clasificador de distancia mínima, aunque es más simple y eficiente en clases compactas y separables, puede verse afectado negativamente por la variabilidad y la forma de las clases, lo que se traduce en una disminución de su precisión en escenarios más complejos.

3.3. Robert Garayzar Arturo

En conclusión, esta práctica permitió comparar dos métodos de clasificación: el clasificador de distancia mínima y el de 1-Nearest Neighbor (1-NN). El clasificador de distancia mínima es rápido y simple, pero su precisión baja con datos complejos. En cambio, 1-NN es más flexible para formas variadas de datos, aunque sensible al ruido y más lento con grandes volúmenes. En general, la práctica ayudó a comprender las fortalezas y limitaciones de ambos enfoques.