



4 DE MARZO DE 2024

ACTIVIDAD 2

MÉTODOS DE PARTICIÓN DE DATA SETS

CRUZ GALLEGOS RAMSES AARÓN

SEM INTELIGENCIA ARTIFICIAL 2

MTRO: CAMPOS PEÑA DIEGO



INTRODUCCIÓN

En la siguiente practica se implementa distintos tipos de particionamiento de data sets para en consiguiente ponerlos a prueba y compararlos entre sí, con el objetivo de observar si hay diferencias entre separarlo de una u otra manera.

DESARROLLO

En esta ocasión solo explicare de manera breve, los métodos de separación, ya que el perceptrón simple es el mismo de la actividad pasada y la codificación de los métodos se encuentra en mi GitHub. En mi caso personal los desarrolle yo mismo, ya que se me fue indiciado que idear 5 maneras de realizar la partición y simplemente programe las maneras que se me vinieron a la mente.

Todas las funciones implementan la misma lógica, reciben el porcentaje en el cual se quiere dividir el data set y simplemente se realizan los cálculos.

Los métodos que utilice fueron:

1. Dividir de manera aleatoria el data set.
2. Dividir de manera ordenada, primero tomando el ochenta por ciento de los datos para el entrenamiento y el veinte por ciento restantes para los datos de prueba.
3. Dividir de manera ordena de forma que los datos de prueba son seleccionados primero, y el ochenta por ciento restantes se utiliza para los datos de entrenamiento.
4. Alternando en bloques de 10 en 10, el programa toma los primeros diez datos del data set, luego los últimos 10 y así se va recorriendo sucesivamente hasta llegar al porcentaje que se pidió para los datos de entrenamiento, el restante lo toma para datos de prueba.
5. Alternando en bloques de 20 en 20, funciona de igual manera que el anterior solo que ahora los bloques de datos son más extensos.

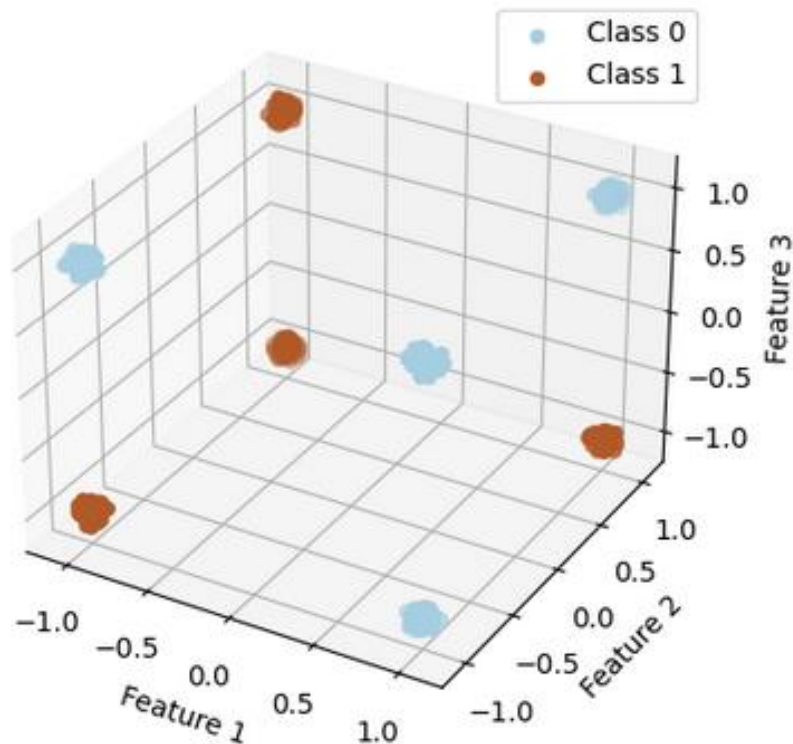
Aunque parezca que las dos últimas formas son iguales se me hizo interesante poner ambas ya que los resultados si fueron distintos y notables a nivel gráfico.

RESULTADOS

DATASET SPHERES2D10 CON PARTICIÓN RANDOM

```
Seleccione el conjunto de datos:
1. spheres2d10.csv
2. spheres2d50.csv
3. spheres2d70.csv
Ingrese el número correspondiente al conjunto de datos deseado: 1
Seleccione el tipo de partición:
1. Partición Random
2. Partición en Orden
3. Partición en Orden (Test Primero)
4. Partición Alternando Bloques de 10
5. Partición Alternando Bloques de 20
Ingrese el número correspondiente al tipo de partición deseado: 1
Ingrese el porcentaje de datos para entrenamiento (ej. 0.8): 0.8
Ingrese el porcentaje de datos para prueba (ej. 0.2): 0.2
Seleccionaste: Partición Random
Perceptrón entrenado con éxito.
Accuracy del perceptrón en datos de prueba (Accuracy): 0.487
Precisión del perceptrón en datos de prueba (Precision): 0.604
Recall del perceptrón en datos de prueba (Recall): 0.487
F1 del perceptrón en datos de prueba (F1): 0.5392263978001833
```

Conjunto de Datos spheres2d10.csv - Técnica: Partición Random



Seleccione el conjunto de datos:

1. spheres2d10.csv
2. spheres2d50.csv
3. spheres2d70.csv

Ingrese el número correspondiente al conjunto de datos deseado: 1

Seleccione el tipo de partición:

1. Partición Random
2. Partición en Orden
3. Partición en Orden (Test Primero)
4. Partición Alternando Bloques de 10
5. Partición Alternando Bloques de 20

Ingrese el número correspondiente al tipo de partición deseado: 5

Ingrese el porcentaje de datos para entrenamiento (ej. 0.8): 0.8

Ingrese el porcentaje de datos para prueba (ej. 0.2): 0.2

Seleccionaste: Partición Alternando Bloques de 20

Perceptrón entrenado con éxito.

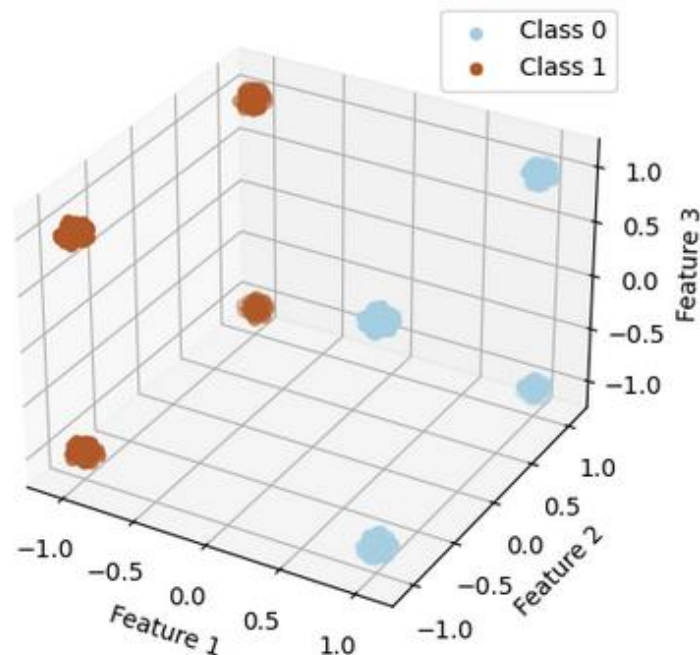
Accuracy del perceptrón en datos de prueba (Accuracy): 0.5

Precisión del perceptrón en datos de prueba (Precision): 0.621

Recall del perceptrón en datos de prueba (Recall): 0.5

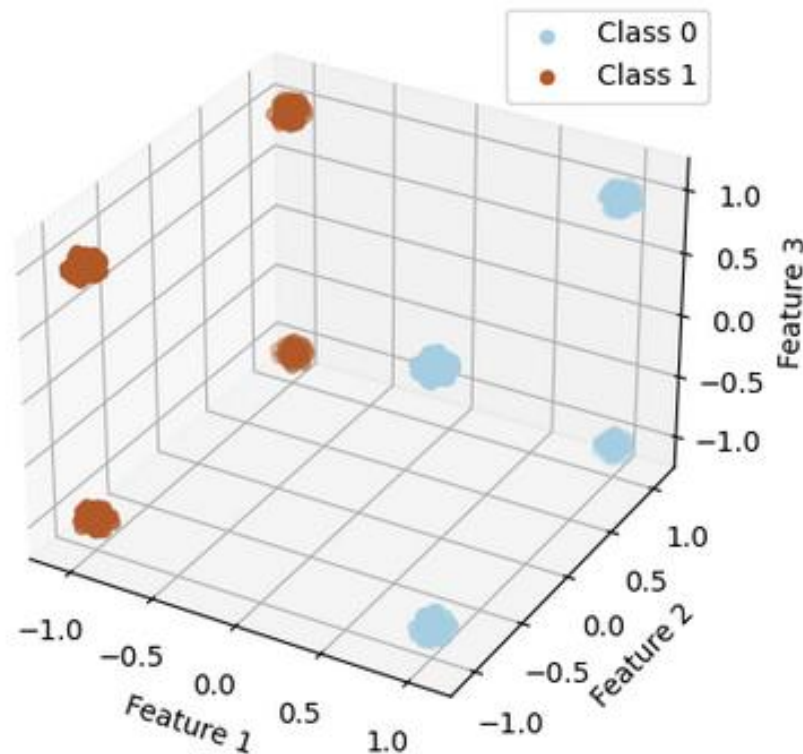
F1 del perceptrón en datos de prueba (F1): 0.5539696699375557

Conjunto de Datos spheres2d10.csv - Técnica: Partición Alternando Bloques de 20



```
Seleccione el conjunto de datos:
1. spheres2d10.csv
2. spheres2d50.csv
3. spheres2d70.csv
Ingrese el número correspondiente al conjunto de datos deseado: 2
Seleccione el tipo de partición:
1. Partición Random
2. Partición en Orden
3. Partición en Orden (Test Primero)
4. Partición Alternando Bloques de 10
5. Partición Alternando Bloques de 20
Ingrese el número correspondiente al tipo de partición deseado: 2
Ingrese el porcentaje de datos para entrenamiento (ej. 0.8): 0.8
Ingrese el porcentaje de datos para prueba (ej. 0.2): 0.2
Seleccionaste: Partición en Orden
Perceptrón entrenado con éxito.
Accuracy del perceptrón en datos de prueba (Accuracy): 0.5
Precisión del perceptrón en datos de prueba (Precision): 0.621
Recall del perceptrón en datos de prueba (Recall): 0.5
F1 del perceptrón en datos de prueba (F1): 0.5539696699375557
```

Conjunto de Datos spheres2d50.csv - Técnica: Partición en Orden



Seleccione el conjunto de datos:

1. spheres2d10.csv
2. spheres2d50.csv
3. spheres2d70.csv

Ingrese el número correspondiente al conjunto de datos deseado: 2

Seleccione el tipo de partición:

1. Partición Random
2. Partición en Orden
3. Partición en Orden (Test Primero)
4. Partición Alternando Bloques de 10
5. Partición Alternando Bloques de 20

Ingrese el número correspondiente al tipo de partición deseado: 4

Ingrese el porcentaje de datos para entrenamiento (ej. 0.8): 0.8

Ingrese el porcentaje de datos para prueba (ej. 0.2): 0.2

Seleccionaste: Partición Alternando Bloques de 10

Perceptrón entrenado con éxito.

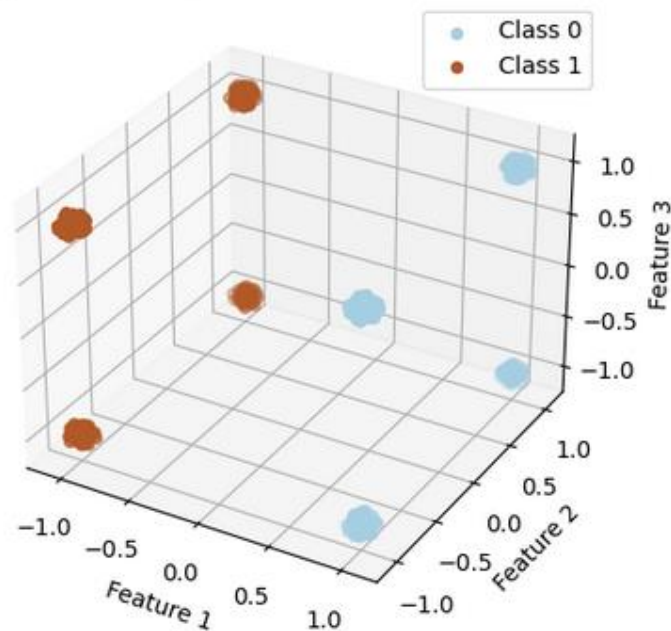
Accuracy del perceptrón en datos de prueba (Accuracy): 0.5

Precisión del perceptrón en datos de prueba (Precision): 0.621

Recall del perceptrón en datos de prueba (Recall): 0.5

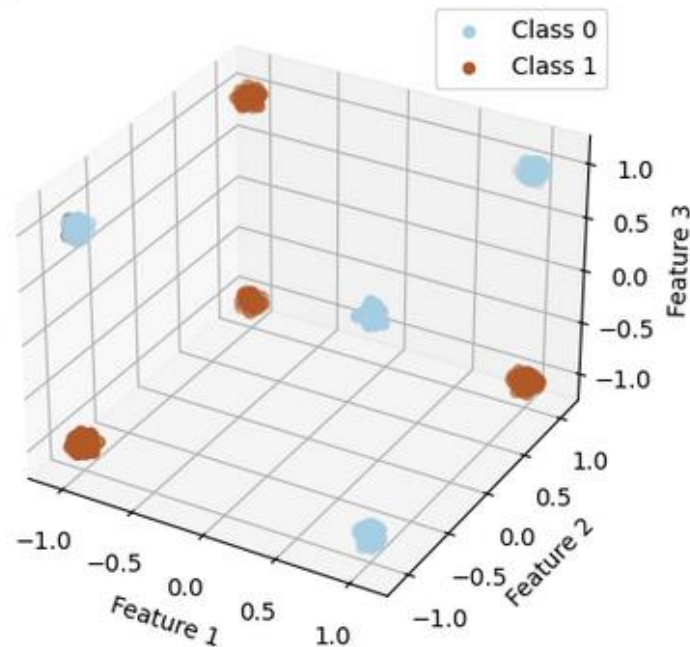
F1 del perceptrón en datos de prueba (F1): 0.5539696699375557

Conjunto de Datos spheres2d50.csv - Técnica: Partición Alternando Bloques de 10




```
Seleccione el conjunto de datos:
1. spheres2d10.csv
2. spheres2d50.csv
3. spheres2d70.csv
Ingrese el número correspondiente al conjunto de datos deseado: 3
Seleccione el tipo de partición:
1. Partición Random
2. Partición en Orden
3. Partición en Orden (Test Primero)
4. Partición Alternando Bloques de 10
5. Partición Alternando Bloques de 20
Ingrese el número correspondiente al tipo de partición deseado: 3
Ingrese el porcentaje de datos para entrenamiento (ej. 0.8): 0.8
Ingrese el porcentaje de datos para prueba (ej. 0.2): 0.2
Seleccionaste: Partición en Orden (Test Primero)
Perceptrón entrenado con éxito.
Accuracy del perceptrón en datos de prueba (Accuracy): 0.546
Precisión del perceptrón en datos de prueba (Precision): 0.635
Recall del perceptrón en datos de prueba (Recall): 0.546
F1 del perceptrón en datos de prueba (F1): 0.5871464860287892
```

Conjunto de Datos spheres2d70.csv - Técnica: Partición en Orden (Test Primero)



CONCLUSIÓN

En general la practica me resulto interesante en algunos puntos, primeramente, aclarando que no se si los resultados obtenidos sean los correctos, la mayoría de ellos no presentan gran cambio en cuanto a la clasificación final de los datos, llegando a mi conclusión personal que no importa mucho el método de partición del data set, lo que si es importante son los porcentajes de entrenamiento y pruebas, por ello utilice un 80 y 20, ya que investigando usualmente es lo que se usa.

GITHUB CON EL CÓDIGO:

[SEM-IA-2/PRACTICA 2 - SPLIT DATA SET at main · CRUZITO4O4/SEM-IA-2 \(github.com\)](https://github.com/CRUZITO4O4/SEM-IA-2)