Universidad Nacional de Colombia

Métodos numéricos

Proyecto final

Profesor:

Edgar Miguel Vargas Chaparro

Integrantes:

Juan Diego Preciado Mahecha

Jorge Aurelio Morales Manrique

Ricardo Andrés Calvo Mendez

Bogotá D.C. 16 de Diciembre 2020

Proyecto final: Clasificación lineal en el aprendizaje de máquina.

Definición del problema.

Dado un conjunto de puntos, los cuales representan objetos con diferentes características, se busca determinar a qué clase pertenece cada uno de ellos. Cabe aclarar que el alcance del problema se limita a clasificar puntos en dos categorías utilizando un espacio en R^2, las máquinas de soporte de vectores proporcionan herramientas para trabajar con diferentes números de dimensiones en relación al número de características. Por tanto, se requiere encontrar la recta óptima que separe los dos conjuntos de datos de acuerdo a características comunes. La recta resultante permitirá clasificar nuevos conjuntos de información en alguna de las diferentes categorías. Este tipo de resultados esperados se definen como funciones discriminantes lineales las cuales tienen la forma $g(x) = w^t x + w_0$ dónde w se conoce como el vector de pesos y w_0 como el sesgo. La recta óptima de separación será representada como una función discriminante lineal cuya gráfica, junto con los puntos representados, en el plano de R2 mostrará la separación obtenida y se podrá realizar un análisis con el fin de determinar los parámetros que definen cada una de las clases.

Este tipo de algoritmos de clasificación son utilizados ampliamente en el aprendizaje supervisado aplicado en el área del aprendizaje de máquina en inteligencia artificial, algunas de las aplicaciones más populares de este tipo de procedimientos son detección de rostros, clasificación de texto e hipertexto, clasificación de imágenes, actividades en el campo de la bioinformática, reconocimiento de escritura a mano, entre otros.

Modelo matemático.

El modelo requiere como entrada un conjunto de n puntos de la forma $(x_i,\ y_i)\ con\ i \in [1,\ n]$. Para cada punto de entrada se conoce su respectiva clase indicada con los valores $\{1,\ -1\}$ de acuerdo a sus características. Como se mencionó anteriormente se quiere obtener una función discriminante lineal con la forma:

$$g(x) = w^t x + w_0$$

La cual se va a transformar en:

$$g(x) = \begin{bmatrix} w_0 & w^t \end{bmatrix} \begin{bmatrix} 1 \\ x \end{bmatrix} = a^t y = g(y)$$

Debido a que se requiere hallar no solo la pendiente sino también el punto de corte.

Se requieren resolver n ecuaciones:

$$\begin{cases} a^t y_1 = b_1 \\ \vdots \\ a^t y_n = b_n \end{cases}$$

Utilizando notación matricial:

$$\underbrace{\begin{bmatrix} y_1^{(0)} & y_1^{(1)} & \cdots & y_1^{(d)} \\ y_2^{(0)} & y_2^{(1)} & \cdots & y_2^{(d)} \\ \vdots & & & \vdots \\ y_n^{(0)} & y_n^{(1)} & \cdots & y_n^{(d)} \end{bmatrix}}_{\boldsymbol{Y}} \underbrace{\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_d \end{bmatrix}}_{\boldsymbol{a}} = \underbrace{\begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ \vdots \\ b_n \end{bmatrix}}_{\boldsymbol{b}}$$

Por tanto, se requiere resolver el sistema de ecuaciones lineales $Ya = b$. Cada una de las filas de la matriz Y tendrá 3 posiciones, que representarán la clase, la coordenada en x y la coordenada en y. Si la clase del punto es igual a -1 las coordenadas del mismo se multiplicarán por -1, esto con el fin de indicar que deberán estar a un lado específico de la recta en contraposición a los puntos de la clase opuesta. El error asociado a una función discriminante lineal se representa por medio de la siguiente ecuación:

$$e = Ya - b$$

Siendo e un vector, se busca que la longitud del mismo sea tan mínima como sea posible es decir se tiene que minimizar la siguiente ecuación, la cual corresponde al error cuadrático medio:

$$J_s(a) = \left| Ya - b \right|^2 = \sum_{i=1}^{n} (a^t y_i - b_i)^2$$

Se busca estimar los valores de los vectores a y b tales que el error asociado sea mínimo. Los valores del vector a representan la recta final, el cual tiene 3 posiciones que representan el valor del término independiente, el coeficiente de x y el coeficiente de y de la recta asociada.

Solución analitica y aproximada.

Al resolver problemas de clasificación lineal se pueden presentar dos casos distintos, que el conjunto de puntos sea linealmente separable o no separable. En el primer caso pueden existir infinitas rectas que separan las clases, por esta razón se hace uso del error cuadrático medio con el fin de encontrar la recta óptima. Dado lo anterior, es complejo encontrar un procedimiento análitico que resuelva el problema. En el segundo caso se dice que no existe ninguna recta que represente la función discriminante lineal asociada al problema.

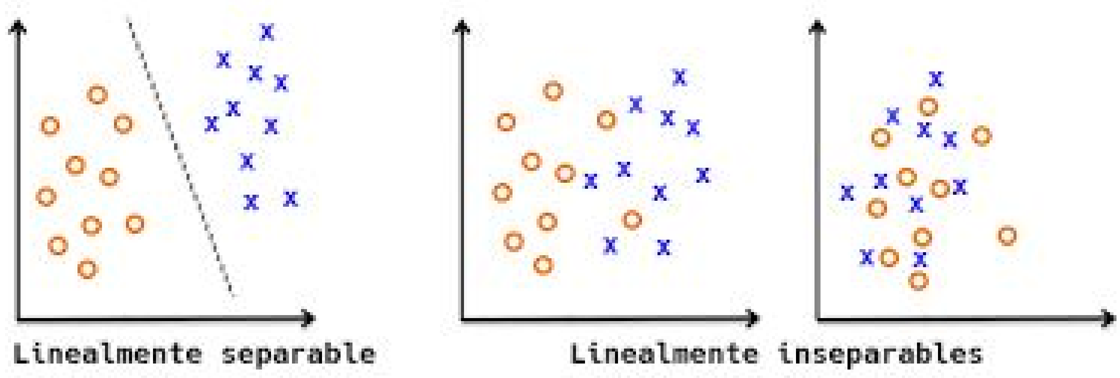

Imagen 1. Casos en la búsqueda de un discriminante lineal.

Solución numérica.

El método que realiza la aproximación numérica posee el nombre "Ho-Kashyap" el cual es iterativo, obteniendo en cada iteración un valor distinto para los vectores a y b por lo cual se usa la siguiente notación para determinar el valor obtenido en la respectiva iteración $a^{(i)}, b^{(i)}, e^{(i)}$. Así mismo se utiliza una variable η la cual representa la tasa de aprendizaje (qué tan rápido serán los cambios entre las iteraciones) con un valor entre $0 < \eta < 1$ y la variable k_{max} la cual indica el número máximo de iteraciones a realizar. A continuación se muestra la secuencia de pasos:

1. Iniciar con valores arbitrarios el vector $a^{(1)}$, el vector $b^{(1)}$ deberá tener valores mayores que cero y la variable $k = 1$ indicando que se va a realizar la primera iteración.
2. Hallar $e^{(k)} = Y a^{(k)} - b^{(k)}$
3. Hallar $b^{(k+1)}$ usando $a^{(k)}$ y $b^{(k)}$

4. Hallar $a^{(k+1)}$ usando $b^{(k+1)}$
5. Hacer $k = k + 1$

El fin del método será determinado por alguna de las siguientes condiciones:

- Todos lo valores del vector $e^{(k)}$ son mayores o iguales que cero
- Se ha alcanzado el límite de iteraciones k_{max}
- Todos los valores del vector e son menores que cero, por ende se cumple que $b^{(k+1)} = b^k$

Para la aplicación del método se deben tener en cuenta los siguientes escenarios:

- **El conjunto de puntos (entre ambas clases) forman una línea recta.** En este caso la matriz Y tendrá dos filas iguales, por tanto no tendrá pseudoinversa y el cálculo de b^k no se podrá efectuar.
- **El conjunto de puntos no es linealmente separable.** En este caso se alcanzará la tercera condición de parada mencionada anteriormente, es decir, se cumplirá que $b^{(k+1)} = b^k$ y por ende la variación entre las iteraciones será de cero.
- **El conjunto de puntos es linealmente separable.** En este caso el procedimiento encontrará la recta que más se aproxima al resultado esperado, teniendo en cuenta el valor de la tasa de aprendizaje η y el número máximo de iteraciones k_{max}.

Análisis de resultados.

El algoritmo que aplica el método “Ho-Kashyap” fue implementado haciendo uso de MATLAB y el entorno gráfico que este provee. A continuación se muestran los resultados obtenidos al aplicar el algoritmo a diferentes conjuntos de datos y su respectivo análisis.

- **Conjunto de puntos linealmente separables.**
  - Clase A: (-1, 0), (-1, 1)
  - Clase B: (0, 0), (1, 2)
  - $a = (1,\ 1,\ 1)$, $\eta = 0.9$, $k_{max} = 200$

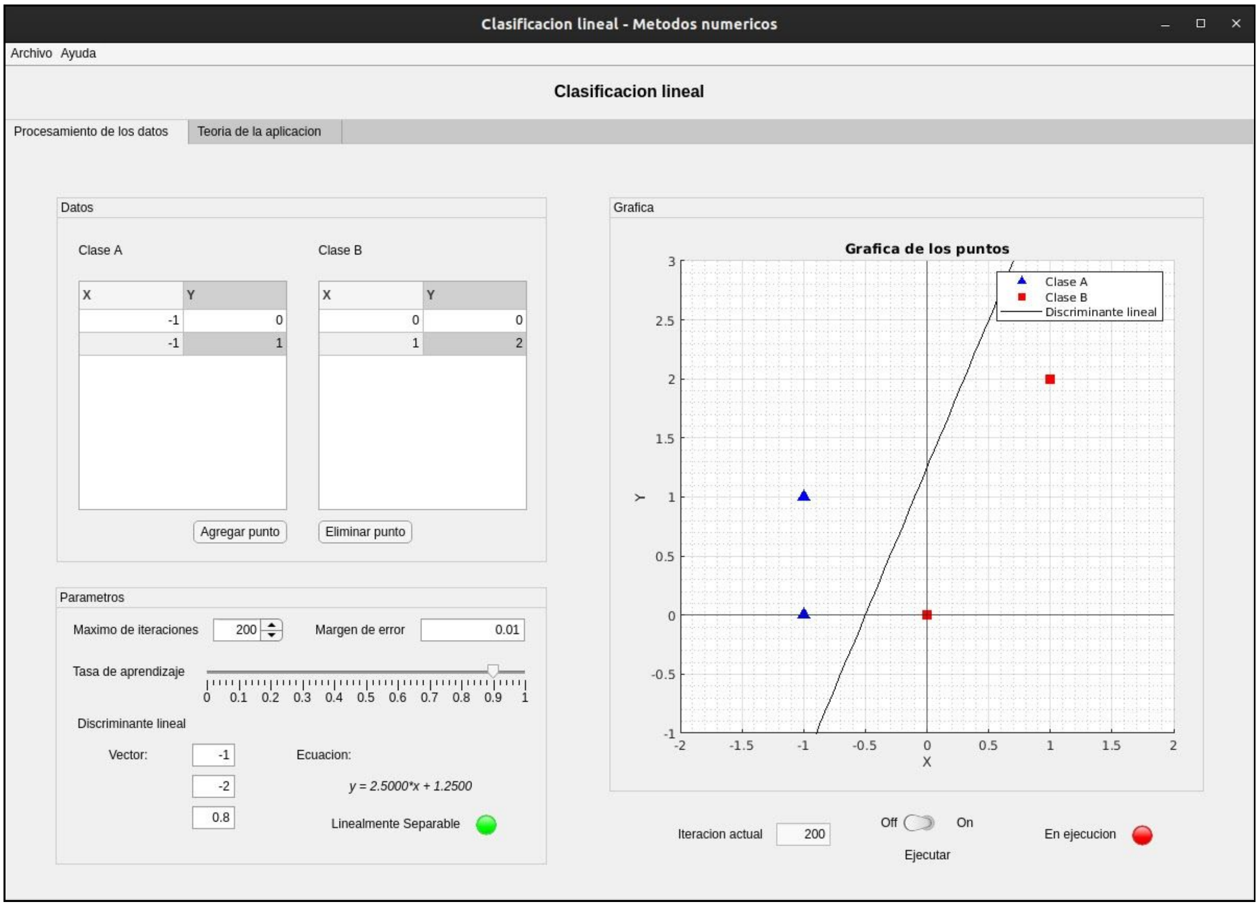

Imagen 2. Primer ejemplo - Conjunto de puntos linealmente separables, a = (1, 1, 1).

Análisis. Se puede observar que las clases son linealmente separables, así que, tras completar 200 iteraciones el programa fue capaz de encontrar una recta que separa ambas clases, sin embargo, en el caso de haber realizado más iteraciones, el método encontraría una recta que tenga un error cuadrático medio menor.

- **Conjunto de puntos linealmente separables.**
    - Clase A: (5, 9), (3, 8)
    - Clase B: (2, 9), (3, 10)
    - $a = (1,\ 2,\ -1)$, $\eta = 0.9$, $k_{max} = 20$

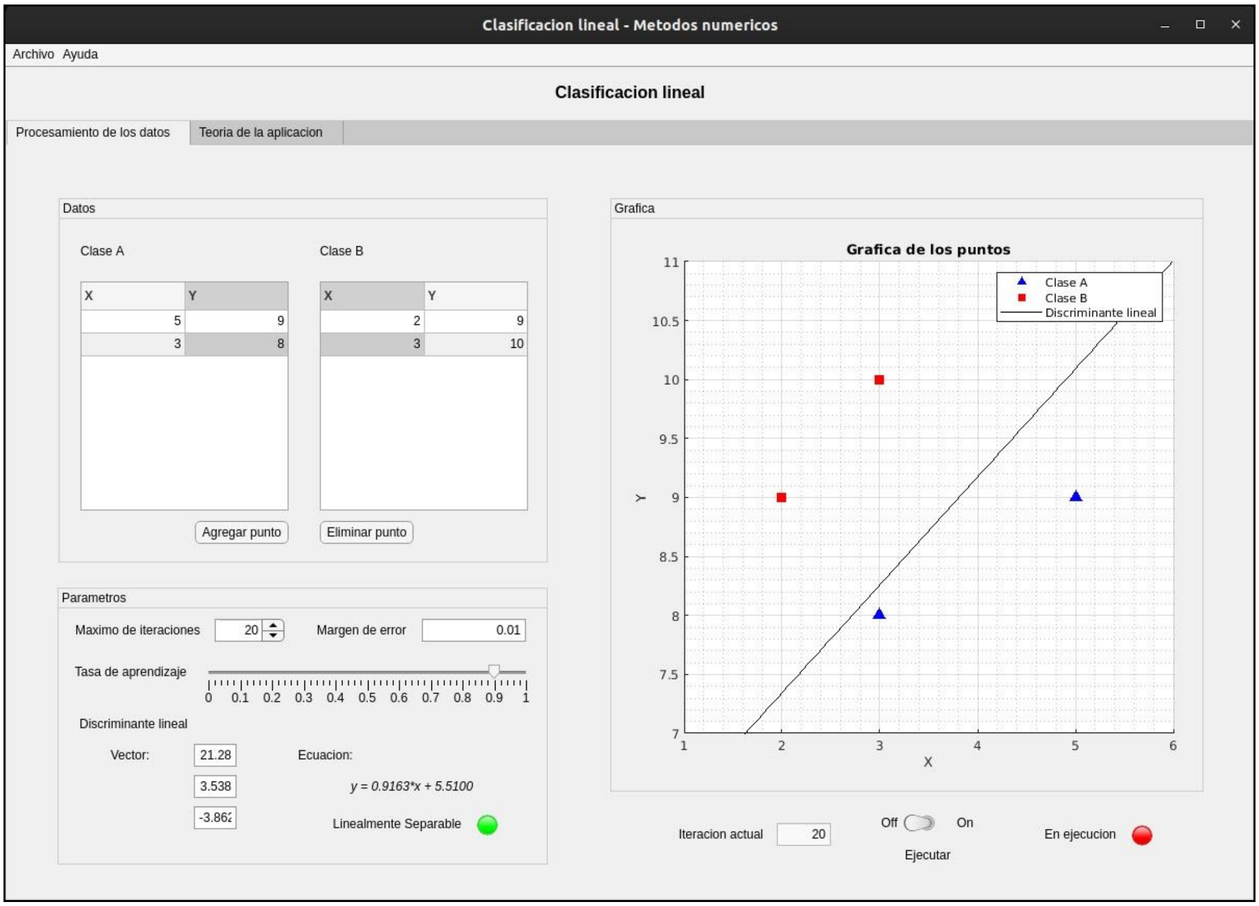

Imagen 3. Segundo ejemplo - Conjunto de puntos linealmente separables, a = (1, 2, -1).

Análisis. Al igual que el ejemplo anterior el conjunto de puntos de entrada corresponde a un conjunto de puntos linealmente separables, con la diferencia de que se realizan como máximo 20 iteraciones, esto con el fin de mostrar que en entre más pequeño sea el valor de k_{max} el valor del error cuadrático medio será mayor, por ende es necesario determinar la cantidad adecuada de iteraciones para garantizar un resultado lo más aproximado al ideal.

- **Conjunto de puntos linealmente no separables**
  - Clase A: (1, 0), (-1, 1)
  - Clase B: (0, 0), (1, 2)
  - $a = (1,\ 1,\ 1)$, $\eta = 0.9$, $k_{max} = 20$

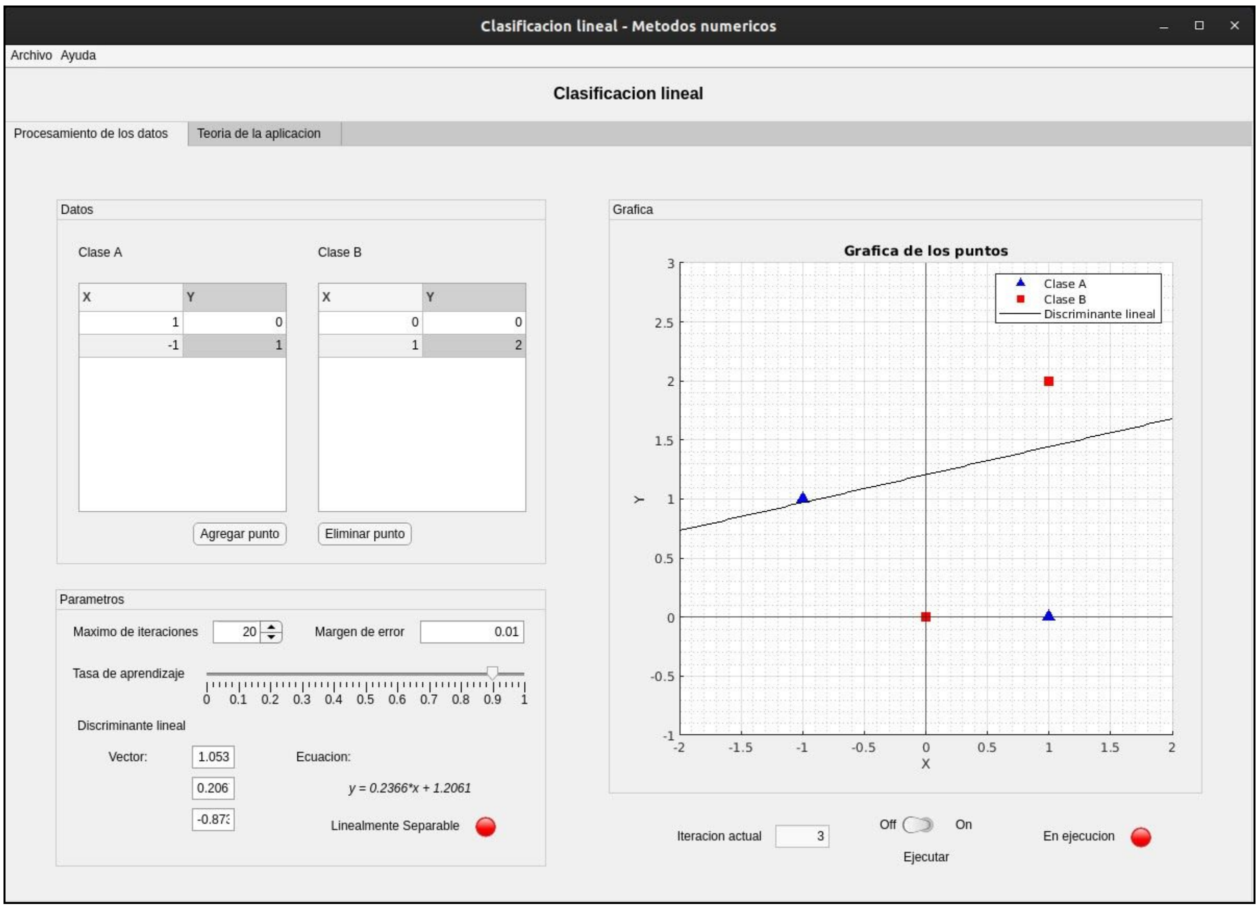

Imagen 4. Tercer ejemplo - Conjunto de puntos linealmente no separables, a = (1, 1, 1).

Análisis. En este caso se puede observar a simple vista que no existe ninguna recta que separe satisfactoriamente las dos clases, Aquí, bastó 3 iteraciones para que el vector e tuviera todas sus componentes negativos, mostrando así que el caso no es linealmente separable.

- **Conjunto de puntos linealmente no separables.**
  - Clase A: (0, 1), (0, -1)
  - Clase B: (-1, 0), (1, 0)
  - $a = (1,\ 1,\ 1)$, $\eta = 0.9$, $k_{max} = 20$

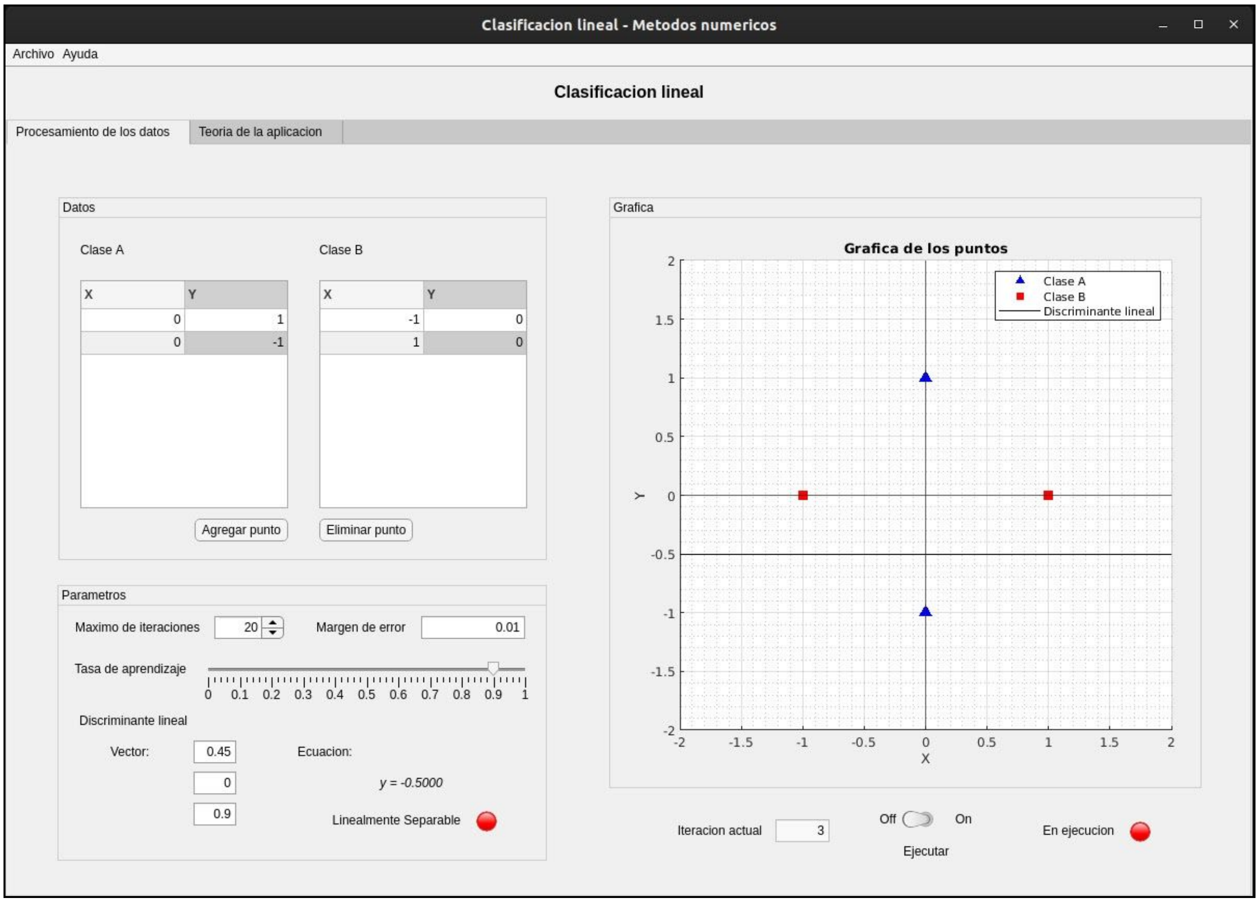

Imagen 5. Cuarto ejemplo - Conjunto de puntos linealmente no separables, a = (1, 1, 1).

Análisis. De manera similar al ejemplo anterior, no se cuenta con una recta que separe correctamente el conjunto de puntos, tomó tres iteraciones para reconocer esto, el programa detuvo su ejecución en dicha iteración.

Conclusiones

- El método de Ho-Kashyap siempre converge a la recta óptima cuando $0 < \eta < 1$. Sin embargo es un método demasiado costoso computacionalmente para la mayoría de casos.
- El método permite reconocer casos linealmente no separables, que es el caso cuando todos los componentes de e son menores que 0.
- El método no siempre converge a la solución en una cantidad finita de pasos, pero siempre converge.

Referencias:

- **Olga Veksler.** https://www.csd.uwo.ca/~oveksler/Courses/CS434a_541a/Lecture9.pdf
- **Olga Veksler.** https://www.csd.uwo.ca/~oveksler/Courses/CS434a_541a/Lecture10.pdf