Ejercicios Tema 4

Percepción

Curso 2021/2022

1. Sea $\mathbf{y}$ un punto en $\mathbb{R}^2$; dibuja todos los puntos que se encuentran a distancia 1 según la distancia L_2, L_1 y L_0
2. ¿Es la distancia euclídea ponderada una métrica?
3. Sea el conjunto de prototipos $X = \{(\mathbf{x}_1, c_1), \ldots, (\mathbf{x}_{10}, c_{10})\}$, con los siguientes valores y etiquetas de clase:

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$	$\mathbf{x}_7$	$\mathbf{x}_8$	$\mathbf{x}_9$
$\mathbf{x}$	$(1,2)$	$(1,3)$	$(2,2)$	$(2,3)$	$(5,2)$	$(5,3)$	$(5,4)$	$(4,3)$	$(4,4)$
c	A	A	A	A	B	B	B	B	B

Sea el punto de test $\mathbf{y} = (3,1)$, se pide:

a) Clasifica $\mathbf{y}$ por el vecino más cercano con distancia euclídea
b) Clasifica $\mathbf{y}$ por los 3-vecinos más cercanos con distancia euclídea
c) ¿Cuántos prototipos de X están a distancia L_1 igual a 2?
d) ¿Cuántos prototipos de X están a distancia L_1 menor igual que 3?

4. Sea $X = \{(\mathbf{x}_1, c_1), \ldots, (\mathbf{x}_{10}, c_{10})\}$ con

	$\mathbf{x}_1$	$\mathbf{x}_2$	$\mathbf{x}_3$	$\mathbf{x}_4$	$\mathbf{x}_5$	$\mathbf{x}_6$	$\mathbf{x}_7$	$\mathbf{x}_8$	$\mathbf{x}_9$	$\mathbf{x}_{10}$
$\mathbf{x}$	$(2,2)$	$(6,4)$	$(7,1)$	$(8,2)$	$(4,2)$	$(8,3)$	$(6,1)$	$(5,2)$	$(3,4)$	$(7,5)$
c	A	B	A	B	A	B	A	B	A	B

Se pide:

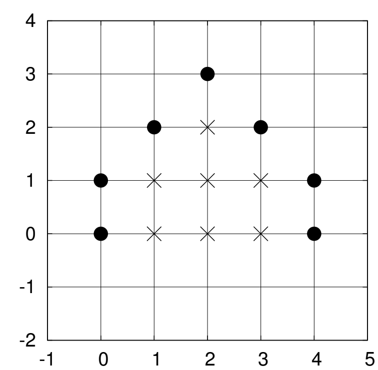
a) Realizar una edición de Wilson y luego un CNN, ambos con k=1 y distancia euclídea

b) Realizar directamente un CNN con k=1 y distancia euclídea

Notas:

- En todos los algoritmos los prototipos se escogen por orden de su índice
- En caso de empate de distancias se clasifica en la clase incorrecta
- Se debe llevar la cuenta de qué prototipos están en los conjuntos S y G en cada paso

5. La figura de la derecha muestra prototipos bidimensionales de 2 clases, $X = \{x_1 = (0,0,\bullet), x_2 = (1,0,\times), x_3 = (1,2,\bullet), x_4 = (1,1,\times), x_5 = (0,1,\bullet), x_6 = (2,2,\times), x_7 = (2,3,\bullet), x_8 = (3,2,\bullet), x_9 = (3,0,\times), x_{10} = (4,0,\bullet), x_{11} = (3,1,\times), x_{12} = (4,1,\bullet), x_{13} = (2,0,\times), x_{14} = (2,1,\times)\}$. Considera la utilización de un clasificador de vecino más cercano en distancia L_1. Se pide:



a) Representa gráficamente la frontera y regiones de decisión.

b) Aplica el algoritmo de condensado de Hart visitando los prototipos por valor de índice creciente. En caso de empate, clasifica en la clase incorrecta.

c) ¿Es posible eliminar más prototipos del conjunto S resultante del apartado anterior sin alterar la frontera y regiones de decisión iniciales? Si es así, indica cuáles.

6. Se tiene el siguiente conjunto de prototipos en $\mathbb{R}^2$ de dos clases:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Prototipo	(1,1)	(2,1)	(2,2)	(3,2)	(3,3)	(4,3)	(3,4)	(5,3)
Clase	A	B	A	B	A	B	A	B

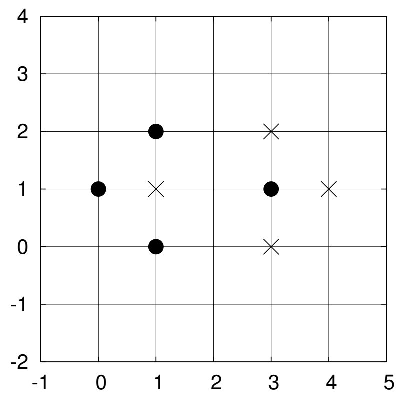
Se pide:

a) Clasificar la muestra (3,1) por vecino más cercano usando distancia euclídea.

b) Aplicar el algoritmo de edición de Wilson sobre el conjunto de prototipos por orden de índice creciente empleando vecino más cercano. En caso de empate por distancias, se clasifica en la clase correcta.

c) Clasificar la misma muestra de test del apartado a) usando el conjunto de prototipos editados.

7. La figura de la derecha muestra prototipos bidimensionales de 2 clases, $X = \{x_1 = (1,1,\times),\ x_2 = (3,1,\bullet),\ x_3 = (3,0,\times),\ x_4 = (1,0,\bullet),\ x_5 = (4,1,\times),\ x_6 = (0,1,\bullet),\ x_7 = (3,2,\times),\ x_8 = (1,2,\bullet)\}$. Considera la utilización de un clasificador de vecino más cercano en distancia L_1. Se pide:



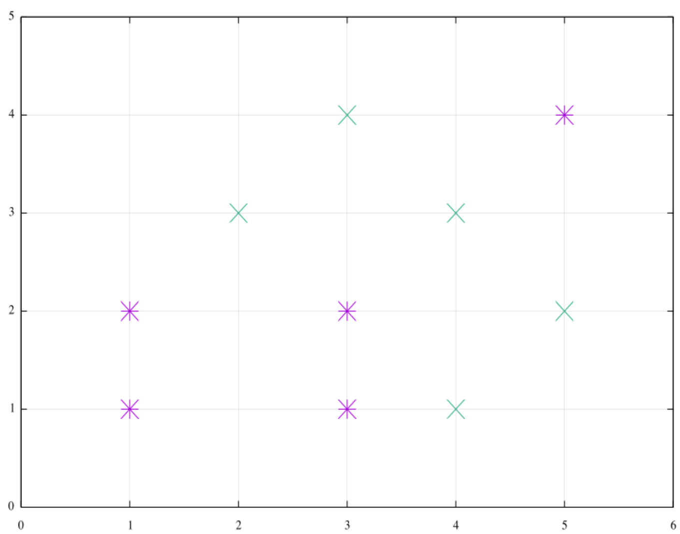
a) Representa gráficamente la frontera y regiones de decisión

b) Aplica el algoritmo de edición de Wilson visitando los prototipos por valor de índice creciente y representa gráficamente la frontera y regiones de decisión resultantes. En caso de empate, clasifica en la clase correcta

c) Sobre el conjunto de prototipos resultantes del apartado anterior, aplica el algoritmo de condensado de Hart visitando los prototipos por valor de índice creciente y representa gráficamente la frontera y regiones de decisión resultantes. En caso de empate, clasifica en la clase correcta

d) Aplica de nuevo el algoritmo de edición de Wilson, pero siendo los prototipos x_1 y x_2 los últimos en ser visitados. Compara la frontera y regiones de decisión resultantes con las del apartado anterior

8. Dado el conjunto de aprendizaje $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_8, \mathbf{x}_9, \mathbf{x}_{10}\}$ con la distribución de clases que muestra la figura, realiza una ejecución del algoritmo de condensado de Hart con distancia euclídea y $k = 1$.



Clase A

$\mathbf{x}_1 = (1, 1)$

$\mathbf{x}_3 = (1, 2)$

$\mathbf{x}_5 = (3, 2)$

$\mathbf{x}_9 = (5, 4)$

$\mathbf{x}_{10} = (3, 1)$

Clase B

$\mathbf{x}_2 = (4, 1)$

$\mathbf{x}_4 = (5, 2)$

$\mathbf{x}_6 = (2, 3)$

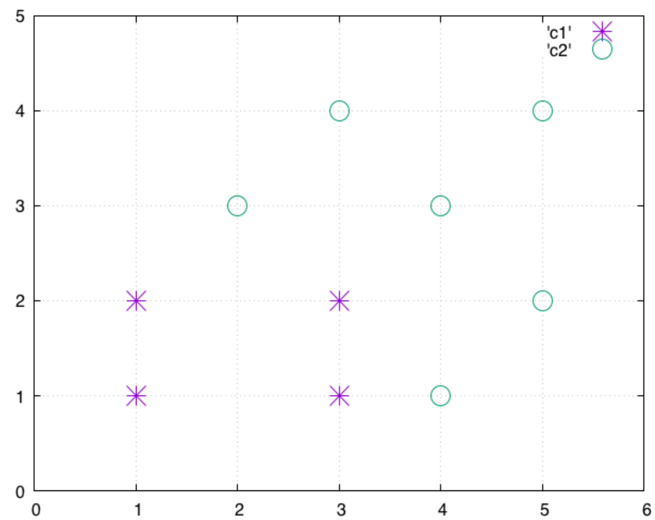
$\mathbf{x}_7 = (4, 3)$

$\mathbf{x}_8 = (3, 4)$

El orden de recorrido de los prototipos en el algoritmo de Hart es **decreciente** con el índice de los mismos: $\mathbf{x}_{10} \ldots \mathbf{x}_1$.

En caso de **empate** de distancias de prototipos de clases diferentes se clasifica en la clase incorrecta.

9. Dado el conjunto de aprendizaje $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_8, \mathbf{x}_9, \mathbf{x}_{10}\}$ con la distribución de clases que muestra la figura, realiza una ejecución del algoritmo de condensado de Hart con distancia euclídea y $k = 1$.



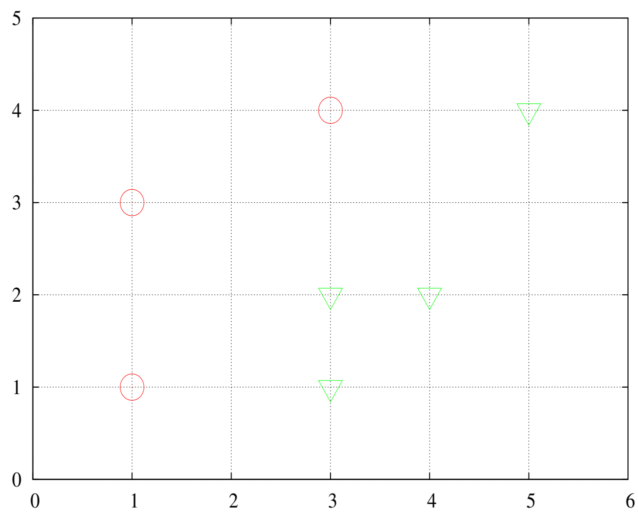
Clase A
$\mathbf{x}_1 = (1,1)$
$\mathbf{x}_3 = (1,2)$
$\mathbf{x}_5 = (3,2)$
$\mathbf{x}_{10} = (3,1)$

Clase B
$\mathbf{x}_2 = (4,1)$
$\mathbf{x}_4 = (5,2)$
$\mathbf{x}_6 = (2,3)$
$\mathbf{x}_7 = (4,3)$
$\mathbf{x}_8 = (3,4)$
$\mathbf{x}_9 = (5,4)$

El orden de recorrido de los prototipos en el algoritmo de Hart es **creciente** con el índice de los mismo: $\mathbf{x}_1 \ldots \mathbf{x}_{10}$.

En caso de **empate** de distancias se clasifica en la clase incorrecta.

10. Dado el conjunto de aprendizaje $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7\}$ con la distribución de clases que muestra la figura, realiza una ejecución del algoritmo de condensado de Hart con distancia euclídea y $k = 1$.



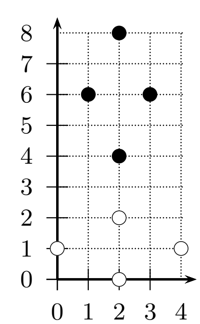
Clase A
$\mathbf{x}_1 = (1,1)$
$\mathbf{x}_3 = (1,3)$
$\mathbf{x}_5 = (3,4)$

Clase B
$\mathbf{x}_2 = (3,1)$
$\mathbf{x}_4 = (4,2)$
$\mathbf{x}_6 = (3,2)$
$\mathbf{x}_7 = (5,4)$

El orden de recorrido de los prototipos en el algoritmo de Hart es **decreciente** con el índice de los mismo: $\mathbf{x}_7 \ldots \mathbf{x}_1$.

En caso de **empate** de distancias se clasifica a la clase incorrecta.

11. La figura de la derecha muestra prototipos bidimensionales de 2 clases: $\{\mathbf{x}_1 = (2,0,\circ), \mathbf{x}_2 = (0,1,\circ), \mathbf{x}_3 = (4,1,\circ), \mathbf{x}_4 = (2,2,\circ), \mathbf{x}_5 = (2,4,\bullet), \mathbf{x}_6 = (1,6,\bullet),$ $\mathbf{x}_7 = (3,6,\bullet), \mathbf{x}_8 = (2,8,\bullet)\}$ utilizados en un clasificador por el vecino más cercano ($k = 1$). Se pide:



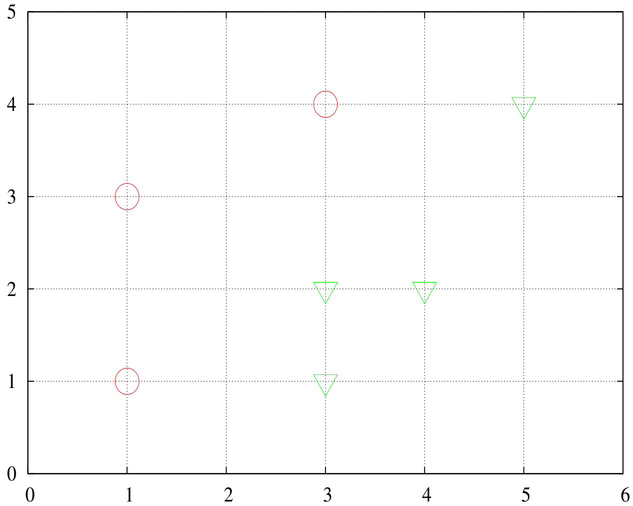
a) Clasificar las muestras $\{\mathbf{y}_1 = (2,3),\ \mathbf{y}_2 = (1,3),\ \mathbf{y}_3 = (0,3)\}$ por distancia Euclídea

b) Clasificar las muestras del apartado a) por distancia Mahalanobis-diagonal por clase

c) Identificar dos puntos de la frontera de decisión asociada al clasificador de distancia Euclídea

d) Ídem al apartado c) pero para el clasificador de distancia Mahalanobis-diagonal por clase

12. Dado el conjunto de aprendizaje $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7\}$ con la distribución de clases que muestra la figura. Realiza una ejecución del algoritmo de condensado de Hart.



Clase A

$\mathbf{x}_1 = (1, 1)$

$\mathbf{x}_3 = (1, 3)$

$\mathbf{x}_5 = (3, 4)$

Clase B

$\mathbf{x}_2 = (3, 1)$

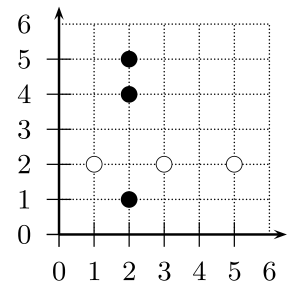
$\mathbf{x}_4 = (4, 2)$

$\mathbf{x}_6 = (3, 2)$

$\mathbf{x}_7 = (5, 4)$

El orden de recorrido de los prototipos en el algoritmo de Hart es el del índice de los mismo: $\mathbf{x}_1 \ldots \mathbf{x}_7$

13. La figura de la derecha muestra prototipos bidimensionales de 2 clases: $x_1 = (1, 2, \circ), x_2 = (2, 1, \bullet), x_3 = (3, 2, \circ), x_4 = (2, 4, \bullet), x_5 = (5, 2, \circ), x_6 = (2, 5, \bullet)$. Aplica el algoritmo de edición de Wilson sobre el conjunto de prototipos por orden de índice utilizando vecino más próximo ($k = 1$) en distancia:

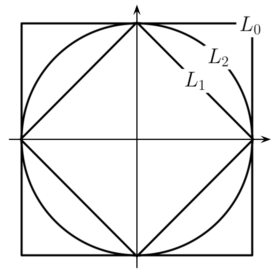


   a) Euclídea

   b) Mahalanobis-diagonal por clase

   Nota: Si es necesario, asume que $\infty \cdot 0 = 0$.

Soluciones

1. 

2. En general, para cualquier valor de los pesos, no. Si los pesos son cero no cumple la propiedad de positiva o nula. Para pesos positivos y estrictamente mayor que cero, sí que sería una distancia.

3. a) A b) Empate (A ó B) c) 1 d) 5

4. *a*) Primeramente, aplicamos el algoritmo de Wilson:

$$\begin{aligned}
&\mathbf{x}_1 \rightarrow \text{Acierto}\\
&\mathbf{x}_2 \rightarrow \text{Acierto}\\
&\mathbf{x}_3 \rightarrow \text{Acierto}\\
&\mathbf{x}_4 \rightarrow \text{Acierto}\\
&\mathbf{x}_5 \rightarrow \text{Error}\\
&\mathbf{x}_6 \rightarrow \text{Acierto}\\
&\mathbf{x}_7 \rightarrow \text{Acierto}\\
&\mathbf{x}_8 \rightarrow \text{Error}\\
&\mathbf{x}_9 \rightarrow \text{Acierto}\\
&\mathbf{x}_{10} \rightarrow \text{Acierto}
\end{aligned}$$

Tras la primera iteración $X = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_9, \mathbf{x}_{10}\}$. En la segunda iteración todas las muestras son clasificadas correctamente.
Una vez aplicado el algoritmo de edición de Wilson, aplicamos el algoritmo CNN. Primera parte, construcción del conjunto S y G:

$$\begin{aligned}
&\mathbf{x}_1 \rightarrow S\\
&\mathbf{x}_2, \text{Error} \rightarrow S\\
&\mathbf{x}_3, \text{Error} \rightarrow S\\
&\mathbf{x}_4, \text{Error} \rightarrow S\\
&\mathbf{x}_6, \text{Acierto} \rightarrow G\\
&\mathbf{x}_7, \text{Acierto} \rightarrow G\\
&\mathbf{x}_9, \text{Acierto} \rightarrow G\\
&\mathbf{x}_{10}, \text{Acierto} \rightarrow G
\end{aligned}$$

$S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$ y $G = \{\mathbf{x}_7, \mathbf{x}_9, \mathbf{x}_{10}\}$

Segunda parte, recorrido de G:

$$\begin{aligned}
&\mathbf{x}_6, \text{Acierto, no mover a } S\\
&\mathbf{x}_7, \text{Acierto, no mover a } S\\
&\mathbf{x}_9, \text{Acierto, no mover a } S\\
&\mathbf{x}_{10}, \text{Acierto, no mover a } S\\
&error = 0 \rightarrow \text{Acabar}
\end{aligned}$$

Acaba con el conjunto reducido: $S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4\}$

b) Primera parte, construcción del conjunto S y G:

$$\begin{aligned}
&\mathbf{x}_1 \to S\\
&\mathbf{x}_2, \text{Error} \to S\\
&\mathbf{x}_3, \text{Error} \to S\\
&\mathbf{x}_4, \text{Error} \to S\\
&\mathbf{x}_5, \text{Acierto} \to G\\
&\mathbf{x}_6, \text{Acierto} \to G\\
&\mathbf{x}_7, \text{Acierto} \to G\\
&\mathbf{x}_8, \text{Error} \to S\\
&\mathbf{x}_9, \text{Acierto} \to G\\
&\mathbf{x}_{10}, \text{Acierto} \to G
\end{aligned}$$

$S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_8\}$ y $G = \{\mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_9, \mathbf{x}_{10}\}$

Segunda parte, recorrido de G, primera iteración:

$$\begin{aligned}
&\mathbf{x}_5, \text{Error} \to S\\
&\mathbf{x}_6, \text{Acierto, no mover a } S\\
&\mathbf{x}_7, \text{Acierto, no mover a } S\\
&\mathbf{x}_9, \text{Acierto, no mover a } S\\
&\mathbf{x}_{10}, \text{Acierto, no mover a } S\\
&\mathit{error} = 1 \to \text{Repetir}
\end{aligned}$$

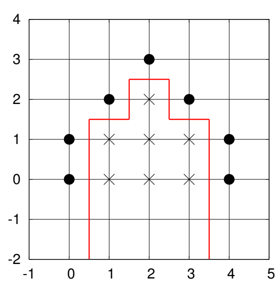
Segunda iteración:

$$\begin{aligned}
&\mathbf{x}_6, \text{Acierto, no mover a } S\\
&\mathbf{x}_7, \text{Acierto, no mover a } S\\
&\mathbf{x}_9, \text{Acierto, no mover a } S\\
&\mathbf{x}_{10}, \text{Acierto, no mover a } S\\
&\mathit{error} = 0 \to \text{Acabar}
\end{aligned}$$

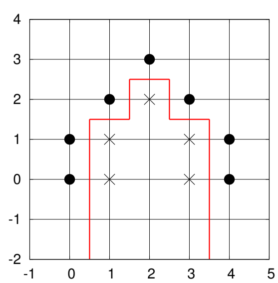
Acaba con el conjunto reducido: $S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_8\}$

5. (Examen Junio 2018)

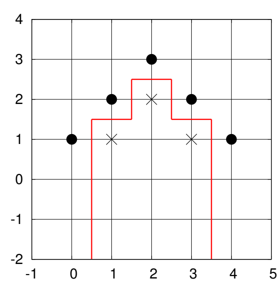
a)



b) $S = X - \{x_{13},\, x_{14}\}$



c) Sí, es posible eliminar los prototipos x_1, x_2, x_9 y x_{10}



6. (Examen Recuperación Junio 2018)

a) La distancia euclídea (al cuadrado) de la muestra (3,1) a cada prototipo es:

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Prototipo	(1,1)	(2,1)	(2,2)	(3,2)	(3,3)	(4,3)	(3,4)	(5,3)
d^2	4	1	2	1	4	5	9	8

Por tanto, se clasifica en la clase B (que es la del prototipo (2,1)).

b) Llamando a cada prototipo x_i, siendo i el índice en la tabla de prototipos, la matriz de distancia euclídea (al cuadrado) entre los prototipos es:

d^2	x_1	x_2	x_3	x_4	x_5	x_6	x_7
x_2	1						
x_3	2	1					
x_4	5	2	1				
x_5	8	5	2	1			
x_6	13	5	5	2	1		
x_7	13	10	5	4	1	2	
x_8	20	13	10	5	4	1	5

Aplicamos edición de Wilson con $k = 1$:

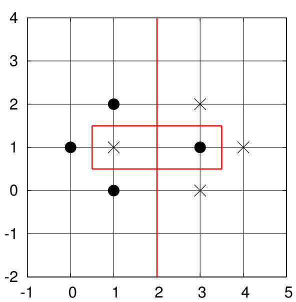
- $x_1 \to x_2$: Error, $X = \{x_2, x_3, x_4, x_5, x_6, x_7, x_8\}$
- $x_2 \to x_3$: Error, $X = \{x_3, x_4, x_5, x_6, x_7, x_8\}$
- $x_3 \to x_4$: Error, $X = \{x_4, x_5, x_6, x_7, x_8\}$
- $x_4 \to x_5$: Error, $X = \{x_5, x_6, x_7, x_8\}$
- $x_5 \to x_6/x_7$: Acierto, $X = \{x_5, x_6, x_7, x_8\}$
- $x_6 \to x_5/x_8$: Acierto, $X = \{x_5, x_6, x_7, x_8\}$
- $x_7 \to x_5$: Acierto, $X = \{x_5, x_6, x_7, x_8\}$
- $x_8 \to x_6$: Acierto, $X = \{x_5, x_6, x_7, x_8\}$

Por tanto, el conjunto de prototipos final es $X = \{(3,3), (4,3), (3,4), (5,3)\}$

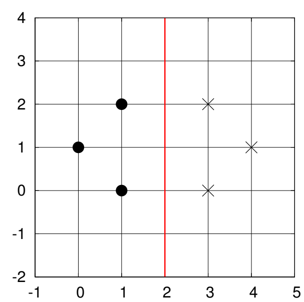
c) Al aplicar vecino más cercano se clasifica en la clase A (la del prototipo (3,3))

7. (Examen Junio 2017)

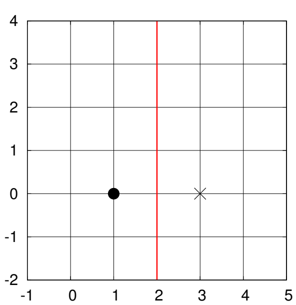
a)



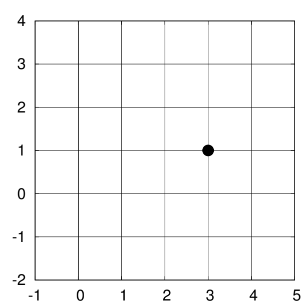
b)



c)



d)



8. (Examen Recuperación Junio 2016)

   Primera parte, construcción del conjunto S y G:

$$\begin{aligned}
&\mathbf{x}_{10} \to S\\
&\mathbf{x}_9, \text{Acierto} \to G\\
&\mathbf{x}_8, \text{Error} \to S\\
&\mathbf{x}_7, \text{Acierto} \to G\\
&\mathbf{x}_6, \text{Acierto} \to G\\
&\mathbf{x}_5, \text{Acierto} \to G\\
&\mathbf{x}_4, \text{Error} \to S\\
&\mathbf{x}_3, \text{Acierto} \to G\\
&\mathbf{x}_2, \text{Error} \to S\\
&\mathbf{x}_1, \text{Acierto} \to G
\end{aligned}$$

   $S = \{\mathbf{x}_{10}, \mathbf{x}_8, \mathbf{x}_4, \mathbf{x}_2\}$ y $G = \{\mathbf{x}_9, \mathbf{x}_7, \mathbf{x}_6, \mathbf{x}_5\mathbf{x}_3, \mathbf{x}_1\}$

   Segunda parte, recorrido de G:

$$\begin{aligned}
&\mathbf{x}_9, \text{Error} \to S\\
&\mathbf{x}_7, \text{Error} \to S\\
&\mathbf{x}_6, \text{Acierto, no mover a } S\\
&\mathbf{x}_5, \text{Acierto, no mover a } S\\
&\mathbf{x}_3, \text{Acierto, no mover a } S\\
&\mathbf{x}_1, \text{Acierto, no mover a } S\\
&error = 1 \to \text{volver a recorrer } G
\end{aligned}$$

   Segunda parte, recorrido de G:

$$\begin{aligned}
&\mathbf{x}_6, \text{Acierto, no mover a } S\\
&\mathbf{x}_5, \text{Acierto, no mover a } S\\
&\mathbf{x}_3, \text{Acierto, no mover a } S\\
&\mathbf{x}_1, \text{Acierto, no mover a } S\\
&error = 0 \to \text{Acabar}
\end{aligned}$$

   Acaba con el conjunto reducido: $S = \{\mathbf{x}_{10}, \mathbf{x}_9, \mathbf{x}_8, \mathbf{x}_7, \mathbf{x}_4, \mathbf{x}_2\}$

9. (Examen Junio 2016)

   Primera parte, construcción del conjunto S y G:

$$\begin{aligned}
&\mathbf{x}_1 \to S\\
&\mathbf{x}_2, \text{Error} \to S\\
&\mathbf{x}_3, \text{Acierto} \to G\\
&\mathbf{x}_4, \text{Acierto} \to G\\
&\mathbf{x}_5, \text{Error} \to S\\
&\mathbf{x}_6, \text{Error} \to S\\
&\mathbf{x}_7, \text{Error} \to S\\
&\mathbf{x}_8, \text{Acierto} \to G\\
&\mathbf{x}_9, \text{Acierto} \to G\\
&\mathbf{x}_{10}, \text{Error} \to S
\end{aligned}$$

$S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_{10}\}$ y $G = \{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_8, \mathbf{x}_9\}$

Segunda parte, recorrido de G:

$$\begin{aligned}
&\mathbf{x}_3, \text{Acierto, no mover a } S\\
&\mathbf{x}_4, \text{Acierto, no mover a } S\\
&\mathbf{x}_8, \text{Acierto, no mover a } S\\
&\mathbf{x}_9, \text{Acierto, no mover a } S\\
&error = 0 \rightarrow \text{Acabar}
\end{aligned}$$

Acaba con el conjunto reducido: $S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5, \mathbf{x}_6, \mathbf{x}_7, \mathbf{x}_{10}\}$

10. (Examen Junio 2015)
Primera parte, construcción del conjunto S y G:

$$\begin{aligned}
&\mathbf{x}_7 \rightarrow S\\
&\mathbf{x}_6, \text{Acierto} \rightarrow G\\
&\mathbf{x}_5, \text{Error} \rightarrow S\\
&\mathbf{x}_4, \text{Error} \rightarrow S\\
&\mathbf{x}_3, \text{Acierto} \rightarrow G\\
&\mathbf{x}_2, \text{Acierto} \rightarrow G\\
&\mathbf{x}_1, \text{Error} \rightarrow S
\end{aligned}$$

$S = \{\mathbf{x}_7, \mathbf{x}_5, \mathbf{x}_4, \mathbf{x}_1\}$ y $G = \{\mathbf{x}_6, \mathbf{x}_3, \mathbf{x}_2\}$

Segunda parte, recorrido de G:

$$\begin{aligned}
&\mathbf{x}_6, \text{Acierto, no mover a } S\\
&\mathbf{x}_3, \text{Acierto, no mover a } S\\
&\mathbf{x}_2, \text{Acierto, no mover a } S\\
&error = 0 \rightarrow \text{Acabar}
\end{aligned}$$

Acaba con el conjunto reducido: $S = \{\mathbf{x}_7, \mathbf{x}_5, \mathbf{x}_4, \mathbf{x}_1\}$

11. (Examen Recuperación Junio 2014)

   a) La clasificación basada en el vecino más cercano por distancia Euclídea

$$d_E(\mathbf{y}, \mathbf{p}) = \sqrt{\sum_{i=1}^{D} (y_i - p_i)^2}$$

   resulta en un empate por distancia para las muestras $\mathbf{y}_1$ y $\mathbf{y}_2$, así que se clasifican en cualquiera de las dos clases, mientras que $\hat{c}(\mathbf{y}_3) = \circ$.

   b) La clasificación basada en el vecino más cercano por distancia Mahalanobis-diagonal por clase

$$d_M(\mathbf{y}, \mathbf{p}) = \sqrt{\sum_{i=1}^{D} \frac{1}{\sigma_{ic}^2} (y_i - p_i)^2}$$

   donde c es la clase de $\mathbf{p}$, y σ_{ic}^2 es la varianza de la componente i-ésima en la clase c. Para calcular la distancia de Mahalanobis-diagonal por clase entre cada prototipo $\mathbf{p}$ y una muestra $\mathbf{y}$, necesitamos calcular previamente las varianzas por clase y componente:

σ_{ic}^2	$i = 1$	$i = 2$
$c = \circ$	2	0,5
$c = \bullet$	0,5	2

Teóricamente para calcular el vecino más cercano deberíamos calcular la distancia de cada prototipo a la muestra $\mathbf{y}$ a clasificar. En esta solución, calcularemos la distancia para el prototipo de cada clase que es el más cercano a la muestra $\mathbf{y}$, pero en la práctica se calcularía para cualquier prototipo que tuviera opciones de ser el más cercano a la muestra $\mathbf{y}$. Para la muestra $\mathbf{y}_1$:

$$d_M(\mathbf{y}_1, \mathbf{x}_4) = \sqrt{\frac{1}{2}(2-2)^2 + \frac{1}{0{,}5}(3-2)^2} = \sqrt{2}$$
$$d_M(\mathbf{y}_1, \mathbf{x}_5) = \sqrt{\frac{1}{0{,}5}(2-2)^2 + \frac{1}{2}(3-4)^2} = \sqrt{0{,}5}$$

Así que $\hat{c}(\mathbf{y}_1) = \bullet$. Para la muestra $\mathbf{y}_2$:

$$d_M(\mathbf{y}_2, \mathbf{x}_4) = \sqrt{\frac{1}{2}(1-2)^2 + \frac{1}{0{,}5}(3-2)^2} = \sqrt{2{,}5}$$
$$d_M(\mathbf{y}_2, \mathbf{x}_5) = \sqrt{\frac{1}{0{,}5}(1-2)^2 + \frac{1}{2}(3-4)^2} = \sqrt{2{,}5}$$

Se produce un empate por distancia clasificándose en cualquiera de las dos clases. Para la muestra $\mathbf{y}_3$:

$$d_M(\mathbf{y}_3, \mathbf{x}_4) = \sqrt{\frac{1}{2}(0-2)^2 + \frac{1}{0{,}5}(3-2)^2} = \sqrt{4}$$
$$d_M(\mathbf{y}_3, \mathbf{x}_5) = \sqrt{\frac{1}{0{,}5}(0-2)^2 + \frac{1}{2}(3-4)^2} = \sqrt{8{,}5}$$

Así que $\hat{c}(\mathbf{y}_3) = \circ$.

c) Los puntos de la frontera de decisión entre las dos clases son los puntos del espacio donde se producen empates. Así que aprovechando el resultado del apartado a) estos puntos serían $(2,3)$ y $(1,3)$.

d) Al igual que en el apartado anterior sabemos que $(1,3)$ es un punto de la frontera de decisión. Dada la simetría en la distribución de los prototipos, podemos asegurar que otro punto en la frontera de decisión será $(3,3)$.

12. (Examen Recuperación Junio 2013)

Primera parte, construcción del conjunto S y G:

$$\begin{aligned}
&\mathbf{x}_1 \to S\\
&\mathbf{x}_2, \text{Error} \to S\\
&\mathbf{x}_3, \text{Acierto} \to G\\
&\mathbf{x}_4, \text{Acierto} \to G\\
&\mathbf{x}_5, \text{Error} \to S\\
&\mathbf{x}_6, \text{Acierto} \to G\\
&\mathbf{x}_7, \text{Error} \to S
\end{aligned}$$

$S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5, \mathbf{x}_7\}$ y $G = \{\mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_6\}$

Segunda parte, recorrido de G:

$$\begin{aligned}
&\mathbf{x}_3, \text{Acierto, no mover a } S\\
&\mathbf{x}_4, \text{Acierto, no mover a } S\\
&\mathbf{x}_6, \text{Acierto, no mover a } S\\
&error = 0 \to \text{Acabar}
\end{aligned}$$

Acaba con el conjunto reducido: $S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_5, \mathbf{x}_7\}$

13. (Examen Junio 2013)

a) El algoritmo de Wilson se reduce a recorrer iterativamente los prototipos tomando uno de ellos como muestra de test y el resto como muestras de entrenamiento, eliminando los prototipos mal clasificados. En este caso es el clasificador de vécino más próximo ($k = 1$) con distancia Euclídea convencional:

	Clase estimada	Conjunto resultante
$x_1 = (1, 2, \circ)$	$\bullet$	$\{x_2, x_3, x_4, x_5, x_6\}$
$x_2 = (2, 1, \bullet)$	$\circ$	$\{x_3, x_4, x_5, x_6\}$
$x_3 = (3, 2, \circ)$	$\circ$	$\{x_3, x_4, x_5, x_6\}$
$x_4 = (2, 4, \bullet)$	$\bullet$	$\{x_3, x_4, x_5, x_6\}$
$x_5 = (5, 2, \circ)$	$\circ$	$\{x_3, x_4, x_5, x_6\}$
$x_6 = (2, 5, \bullet)$	$\bullet$	$\{x_3, x_4, x_5, x_6\}$

b) En este caso el algoritmo de Wilson utiliza el clasificador de vécino más próximo ($k = 1$) con distancia Mahalanobis-diagonal por clase:

$$d(\mathbf{y}, \mathbf{p}) = \sqrt{\sum_{i=1}^{D} \frac{1}{\sigma_{ic}^2}(y_i - p_i)^2}$$

donde c es la clase de $\mathbf{p}$, y σ_{ic}^2 es la varianza de la componente i-ésima en la clase c.

Aplicamos una primera iteración del algoritmo de Wilson. Para calcular la distancia de Mahalanobis-diagonal por clase entre cada prototipo $\mathbf{p}$ y un prototipo excluido $\mathbf{y}$, necesitamos calcular previamente las varianzas por clase y componente que en todos los casos son:

σ_{ic}^2	$i = 1$	$i = 2$
$c = \circ$	$\alpha > 0$	0
$c = \bullet$	0	$\beta > 0$

La distancia entre un prototipo $\mathbf{p}$ de la clase $\circ$ y un prototipo excluido $\mathbf{y}$ de la clase $\bullet$ resulta en:

$$d(\bullet, \circ) = \sqrt{\frac{1}{\sigma_{1\circ}^2}(y_1 - p_1)^2 + \frac{1}{\sigma_{2\circ}^2}(y_2 - p_2)^2} = \sqrt{\frac{1}{\alpha}(y_1 - p_1)^2 + \frac{1}{0}(y_2 - p_2)^2} = +\infty$$

Lo mismo ocurre cuando calculamos la distancia Mahalanobis-diagonal por clase entre un prototipo $\mathbf{p}$ de la clase $\bullet$ y un prototipo excluido $\mathbf{y}$ de la clase $\circ$. Sin embargo, no ocurre lo mismo cuando calculamos esta distancia entre prototipos de la misma clase. Por ejemplo, para dos prototipos de la clase $\circ$:

$$d(\circ, \circ) = \sqrt{\frac{1}{\sigma_{1\circ}^2}(y_1 - p_1)^2 + \frac{1}{\sigma_{2\circ}^2}(y_2 - p_2)^2} = \sqrt{\frac{1}{\alpha}(y_1 - p_1)^2 + \frac{1}{0}(0)^2} = \sqrt{\frac{1}{\alpha}(y_1 - p_1)^2}$$

Por tanto, todos los prototipos son correctamente clasificados y el algoritmo de Wilson no elimina ninguno.