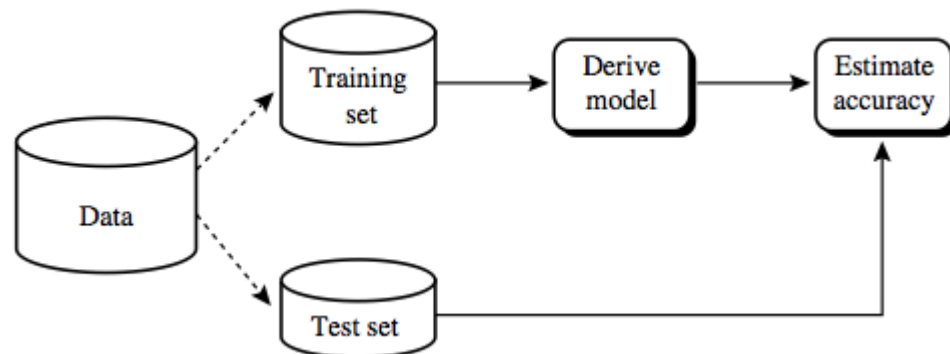


SISTEMAS INTELIGENTES



DEFINICIÓN

Un Sistema inteligente se basa en cálculos matemáticos para tomar decisiones propias sin la necesidad de preguntar al usuario. De esta manera, los algoritmos de aprendizaje de maquina han evolucionado para asemejarse al comportamiento humano. En las ultimas décadas, las redes neuronales han sido uno de los mayores avances en la programación. Estos algoritmos, puede ser: (i) no supervisados, los mismos que no necesitan tener datos anteriores para entrenarse (memoria) y (ii) supervisados. Estos necesitan de un conjunto de datos de entrenamiento y otro para pruebas. Generalmente el 80% de la base de datos recolectada con anticipación es destinada al conjunto de entrenamiento. Además, es necesario conocer el rendimiento de aciertos del modelo.

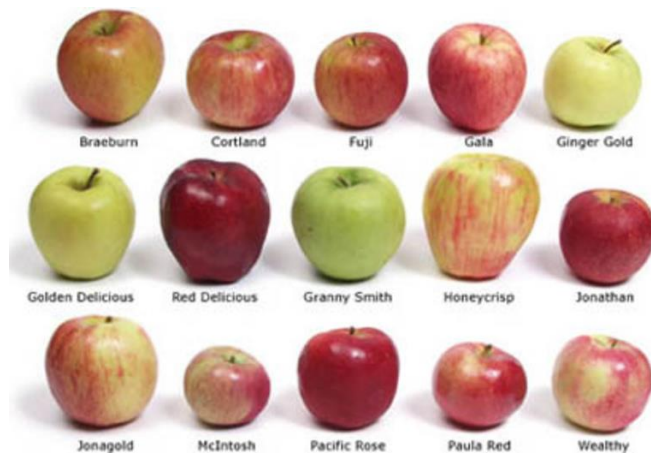


Matriz $A = M_x N$; $M = \text{filas}$ y $N = \text{columnas}$

Vector $D = N$; $N = \text{columnas}$ y 1 fila

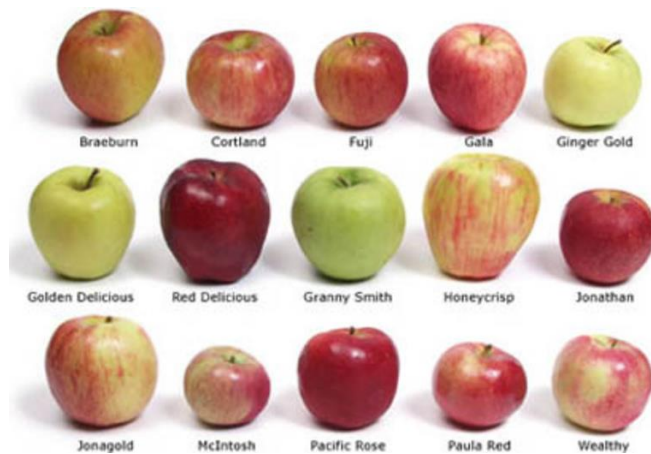
ALGORITMOS DE CLASIFICACIÓN

Los algoritmos de clasificación permiten reconocer a nuevos datos entrantes a un sistema y asignarlos a un conjunto previamente establecido. Para ello, es necesario que el algoritmo sea entrenado con tus las posibles variaciones del conjunto y pueda reconocer adecuadamente.



ALGORITMOS DE CLASIFICACIÓN

Los algoritmos de clasificación permiten reconocer a nuevos datos entrantes a un sistema y asignarlos a un conjunto previamente establecido. Para ello, es necesario que el algoritmo sea entrenado con tus las posibles variaciones del conjunto y pueda reconocer adecuadamente.



CONJUNTO DE DATOS DE ENTRENAMIENTO

Con el ejemplo anterior, si queremos un buen sistema que clasifique las frutas debemos elegir adecuadamente el conjunto de datos a entrenar. Ya que se parte de un criterio que comenta :”**No todos los datos brindan información al algoritmo**”. Por esta razón, existen algoritmos que permiten reconocer al mejor conjunto. Como resultado, se optimiza la memoria, los tiempos de ejecución y respuesta, entre otros. Los mas usados son:

- Selección de prototipos
- Balanceo de datos

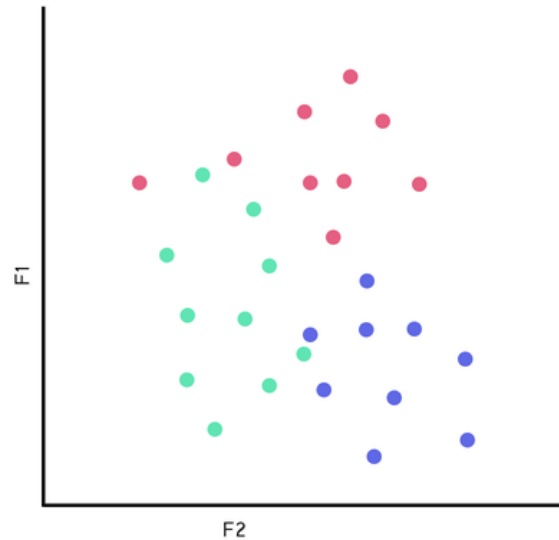
Para estos ejemplos la mayoría de ellos es necesario la normalización de datos y la eliminación de datos negativos.

Normalización: encontrar el valor mas alto de la columna y dividirlo para el resto.

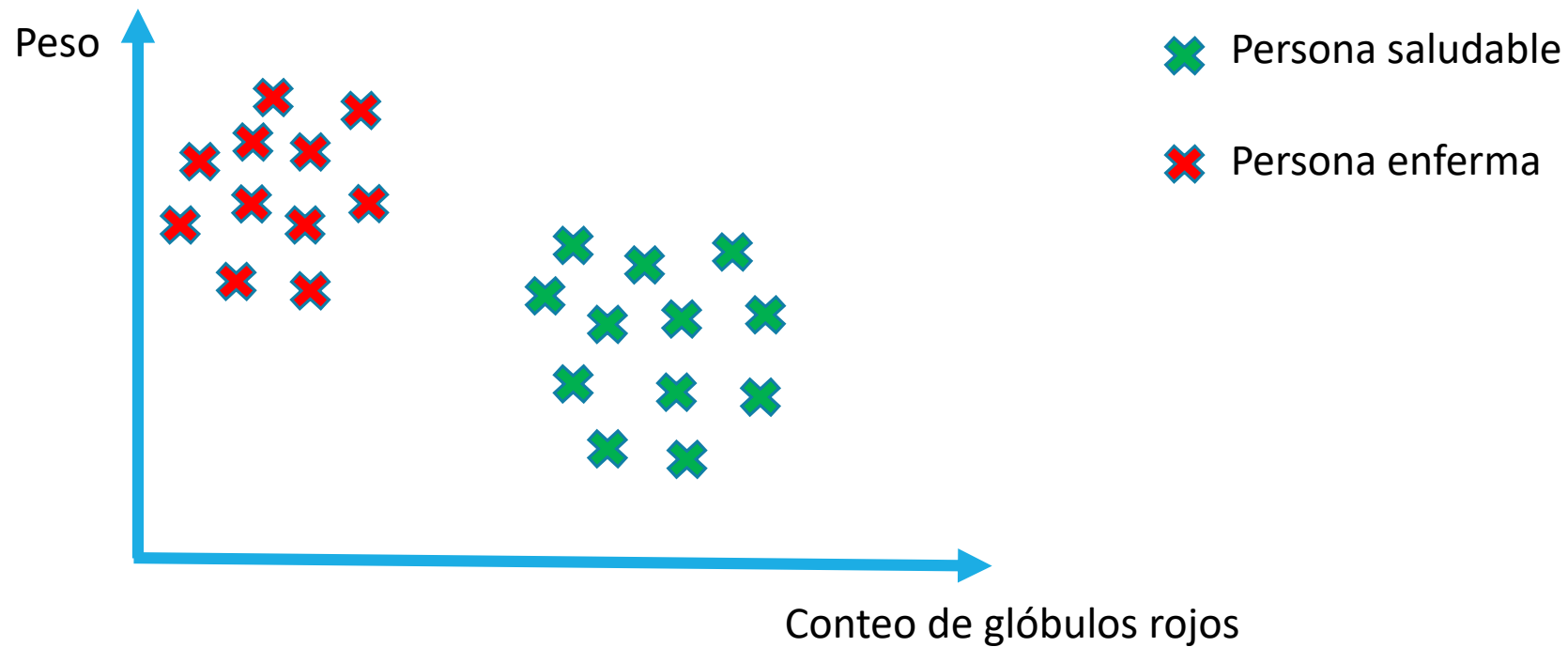
Eliminación de datos negativos: Encontrar el valor negativo mas alto y retarlo al resto.

SELECCIÓN DE PROTOTIPOS

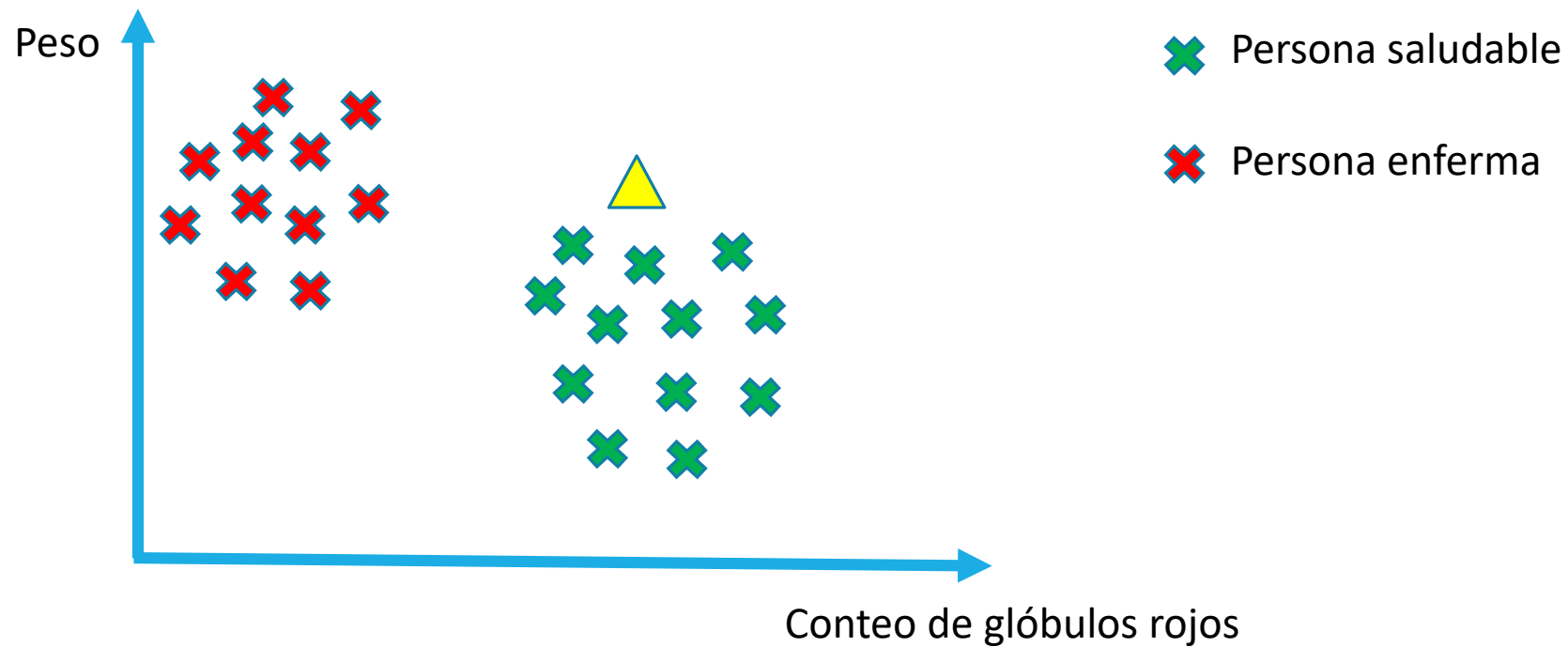
Surgen como métodos para minimizar el impacto de grandes conjuntos de datos en el comportamiento de los algoritmos, reduciendo el tamaño de los datos sin dañar el calidad del conocimiento intrínseco almacenado inicialmente. Es decir, los requisitos de manejo de datos se reducen mientras que la capacidad predictiva de los algoritmos se mantiene. El criterio usado es del vecino mas cercano.



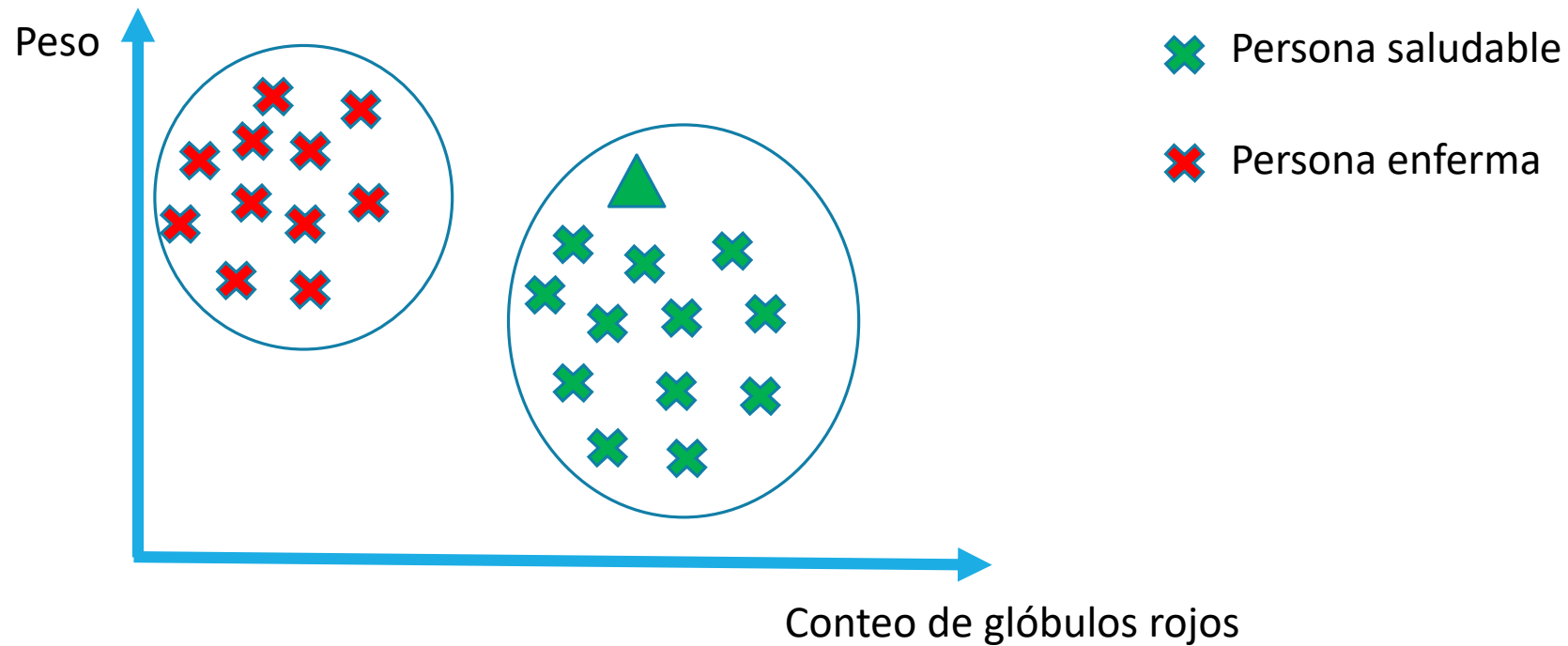
REGLA DEL VECINO MÁS CERCANO (NN)



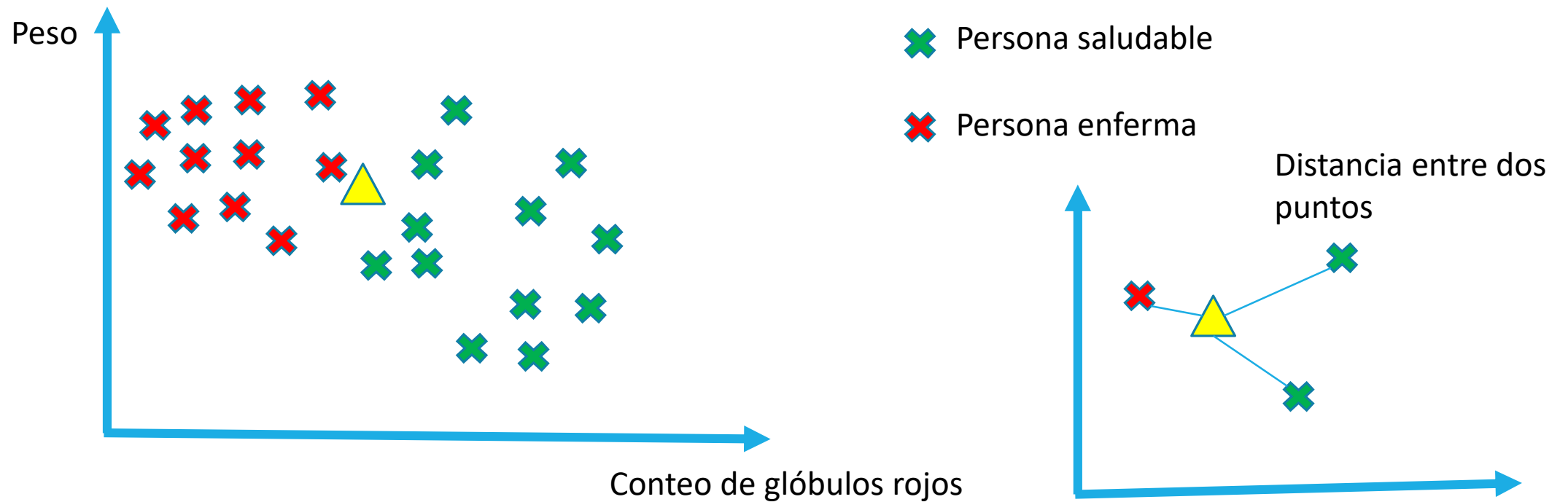
REGLA DEL VECINO MÁS CERCANO (NN)



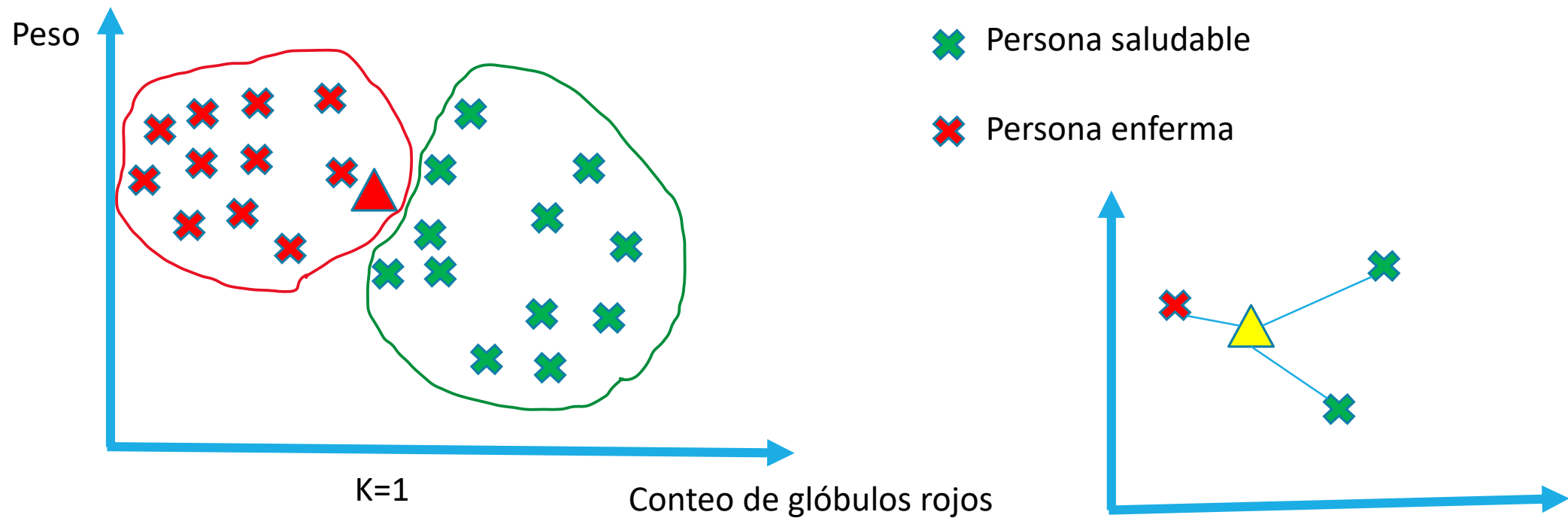
REGLA DEL VECINO MÁS CERCANO (NN)



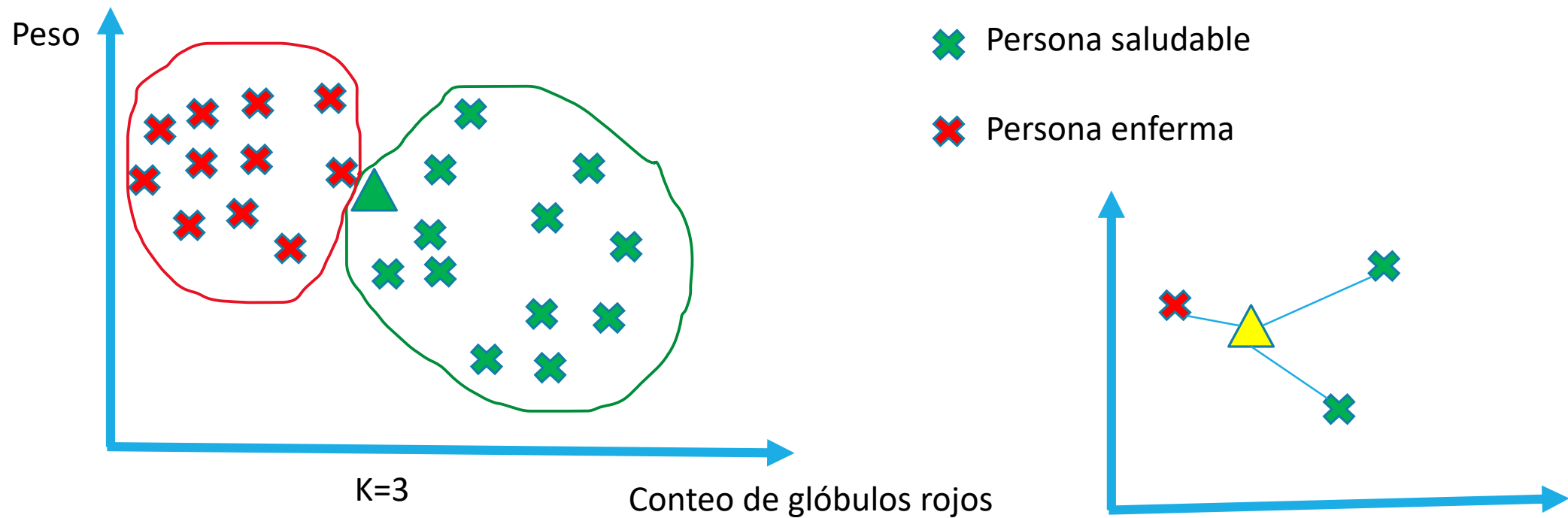
REGLA DEL VECINO MÁS CERCANO (NN)



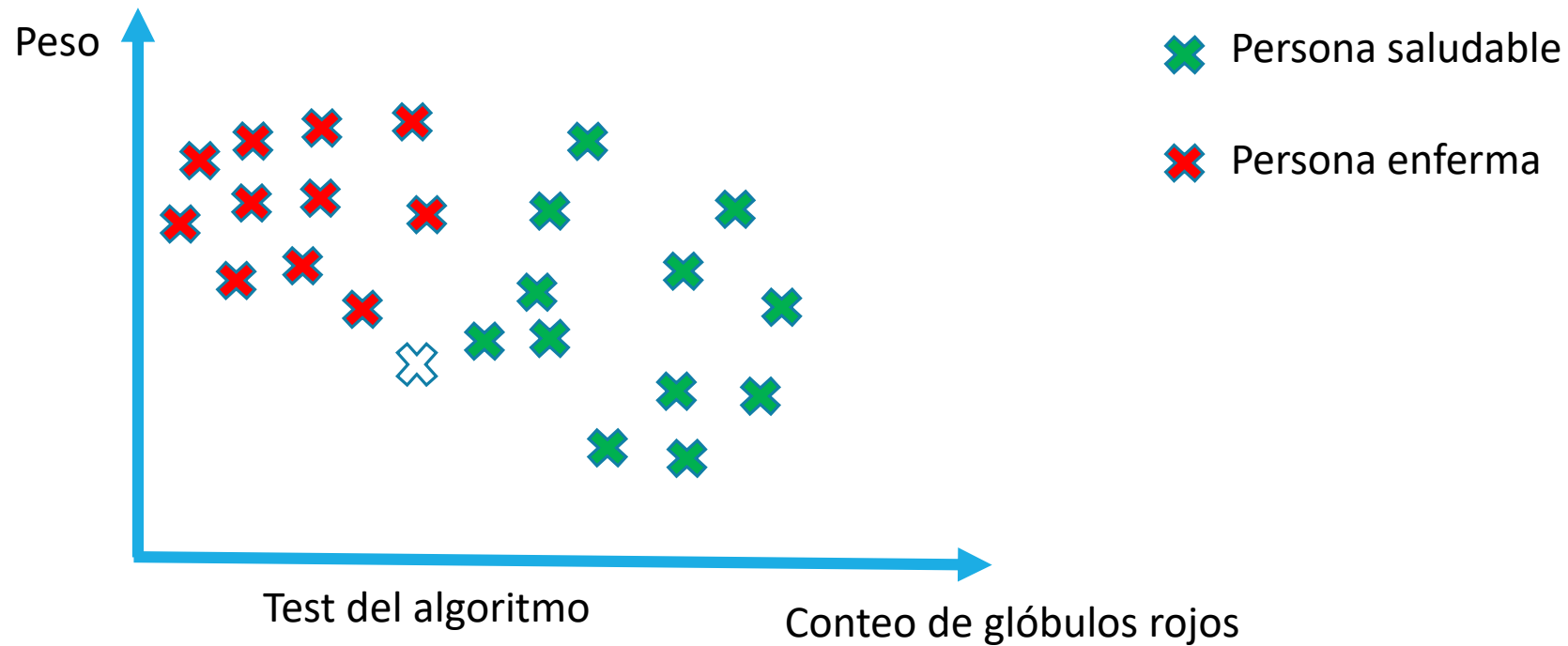
REGLA DEL VECINO MÁS CERCANO (NN)



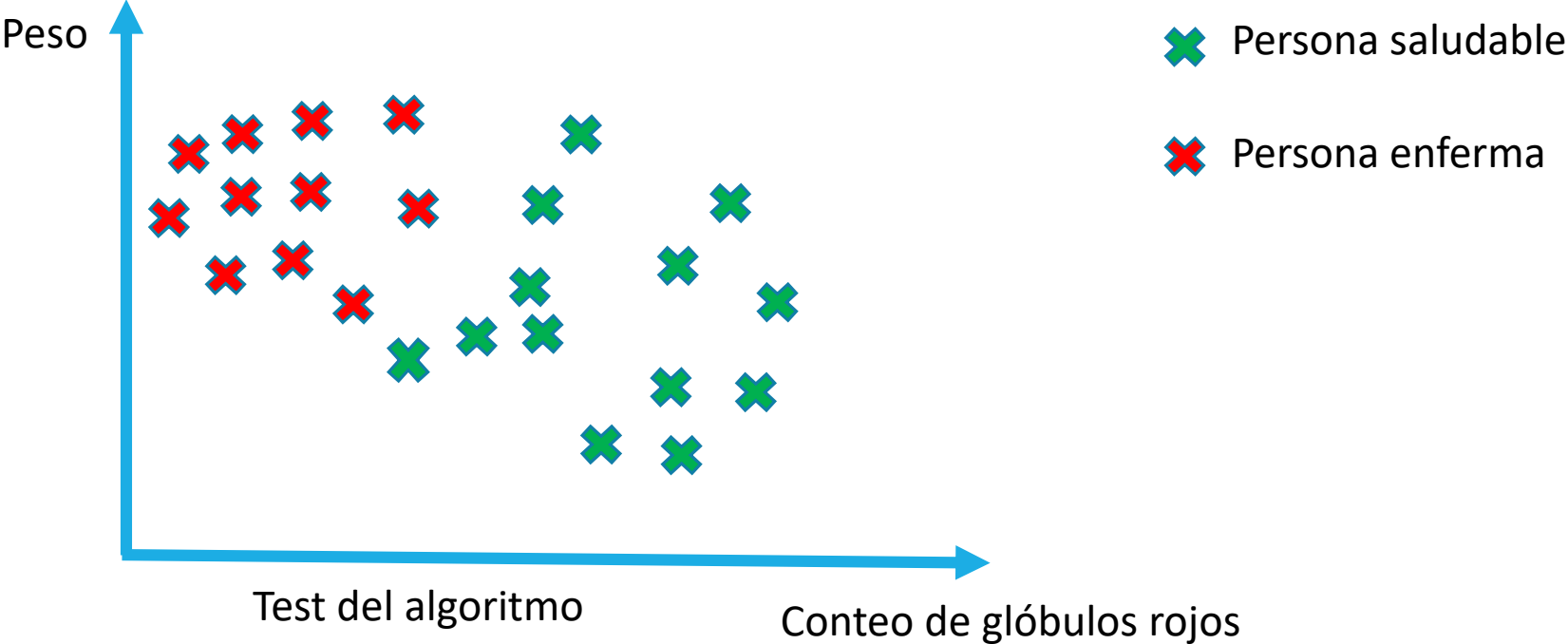
REGLA DEL VECINO MÁS CERCANO (NN)



REGLA DEL VECINO MÁS CERCANO (NN)



REGLA DEL VECINO MÁS CERCANO (NN)



ALGORITMOS DE CLASIFICACIÓN

•KNN

1. Elige el número de vecinos (k).
2. Realizar la distancia entre dos puntos de una nueva instancia con toda la base de entrenamiento.
3. Seleccionar los k datos de la base de entrenamiento que más cercanos a la nueva instancia.
4. Seleccionar el grupo donde se encuentre el mayor número de k-vecinos.
5. Asignar a la nueva instancia al grupo que pertenece.

$$d_E(P, Q) = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

ALGORITMOS DE CLASIFICACIÓN- Rendimiento

KNN

Matriz de confusión: Es una herramienta que permite la visualización del desempeño de un algoritmo que se emplea en aprendizaje supervisado.

		Clasificación	
		Positivo	Negativo
Verdad Terreno	Positivo	Positivos Ciertos	Negativos Falsos
	Negativo	Positivos Falsos	Negativos Ciertos

ALGORITMOS DE CLASIFICACIÓN-

Rendimiento

TP – True Positives: Son el número verdaderos positivos, es decir, de predicciones correctas para la clase +.

FN – False Negatives: Son el número de falsos negativos, es decir, la predicción es negativa cuando realmente el valor tendría que ser positivo. A estos casos también se les denomina errores de tipo II.

FP – False Positives: Son el número de falsos positivos, es decir, la predicción es positiva cuando realmente el valor tendría que ser negativo. A estos casos también se les denomina errores de tipo I.

TN – True Negatives: Son el número de verdaderos negativos, es decir, de predicciones correctas para la clase -.

ALGORITMOS DE CLASIFICACIÓN-

Rendimiento

Sensibilidad: también se la llama recall o tasa de verdaderos positivos. Nos da la probabilidad de que, dada una observación realmente positiva, el modelo la clasifique así.

$$\text{Sensibilidad} = \frac{TP}{TP+FN}$$

Especificidad: también llamado *ratio de verdaderos negativos*. Nos da la probabilidad de que, dada una observación realmente negativa, el modelo la clasifique así.

$$\text{Especificidad} = \frac{TN}{TN+FP}$$

Precisión: también llamado *valor de predicción positiva*. Nos da la probabilidad de que, dada una predicción positiva, la realidad sea positiva también.

$$\text{Presición} = \frac{TP}{TP+FP}$$

ALGORITMOS DE CLASIFICACIÓN- Rendimiento

Error de clasificación: Porcentaje de errores del modelo.

$$\text{Error} = \frac{FP+FN}{TP+TN+FP+FN}$$

Accuracy: Porcentaje total de los aciertos de nuestro modelo.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}$$