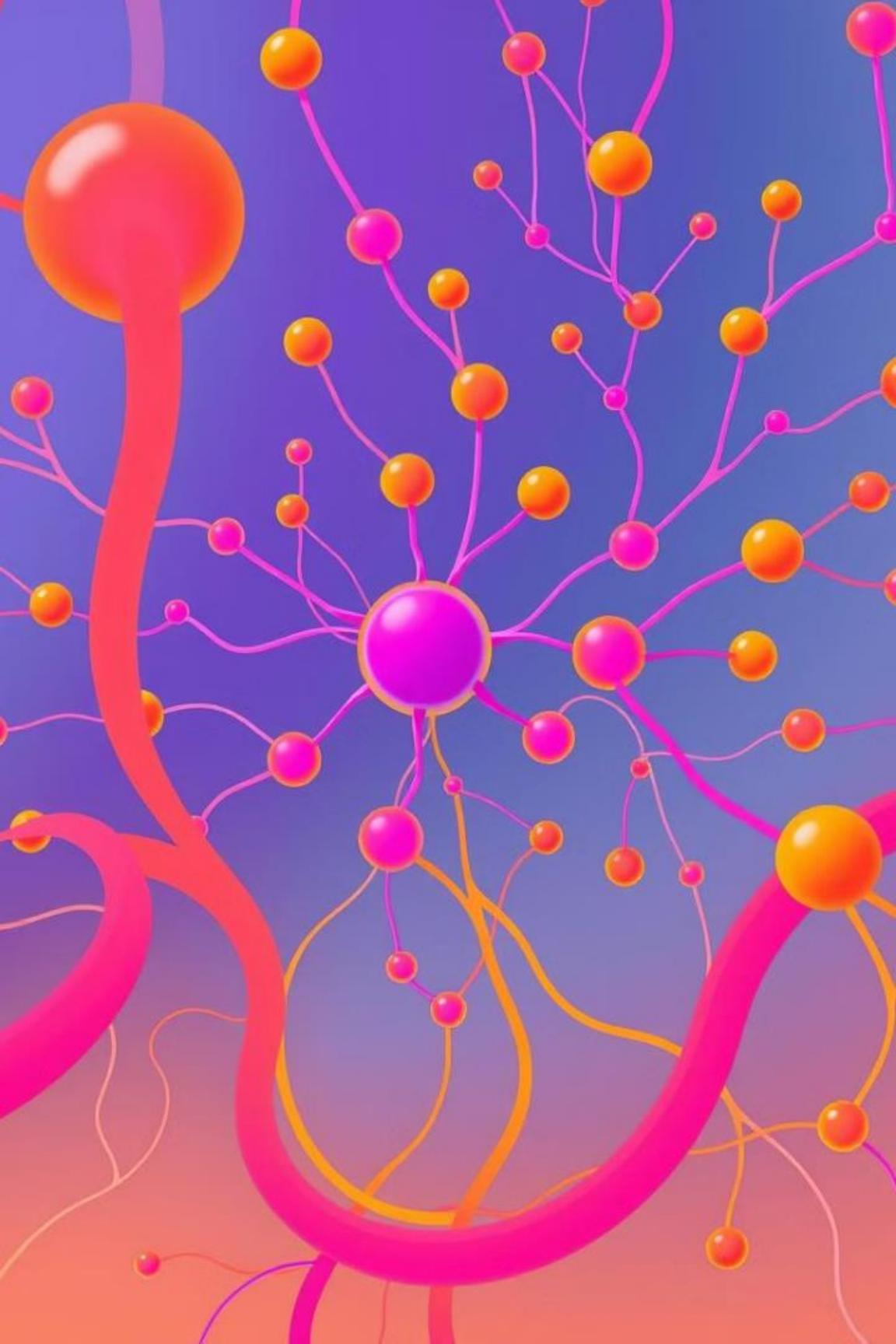


Métricas de rendimiento (evaluación)

Existen varias métricas para evaluar el rendimiento de los algoritmos de aprendizaje de forma experimental.



Métricas en regresión (evaluación de la predicción)

Supongamos que se desea hacer una predicción de un valor para una característica objetivo en el ejemplo x :

- y es el valor observado de la característica objetivo en el ejemplo x .
- $\hat{y} = h(x)$ es el valor previsto de la característica objetivo en el ejemplo x .

si \hat{y} difiere de y existe un error. Entonces el error se define por: $e = \hat{y} - y$

Existen varias formas de medir el error.

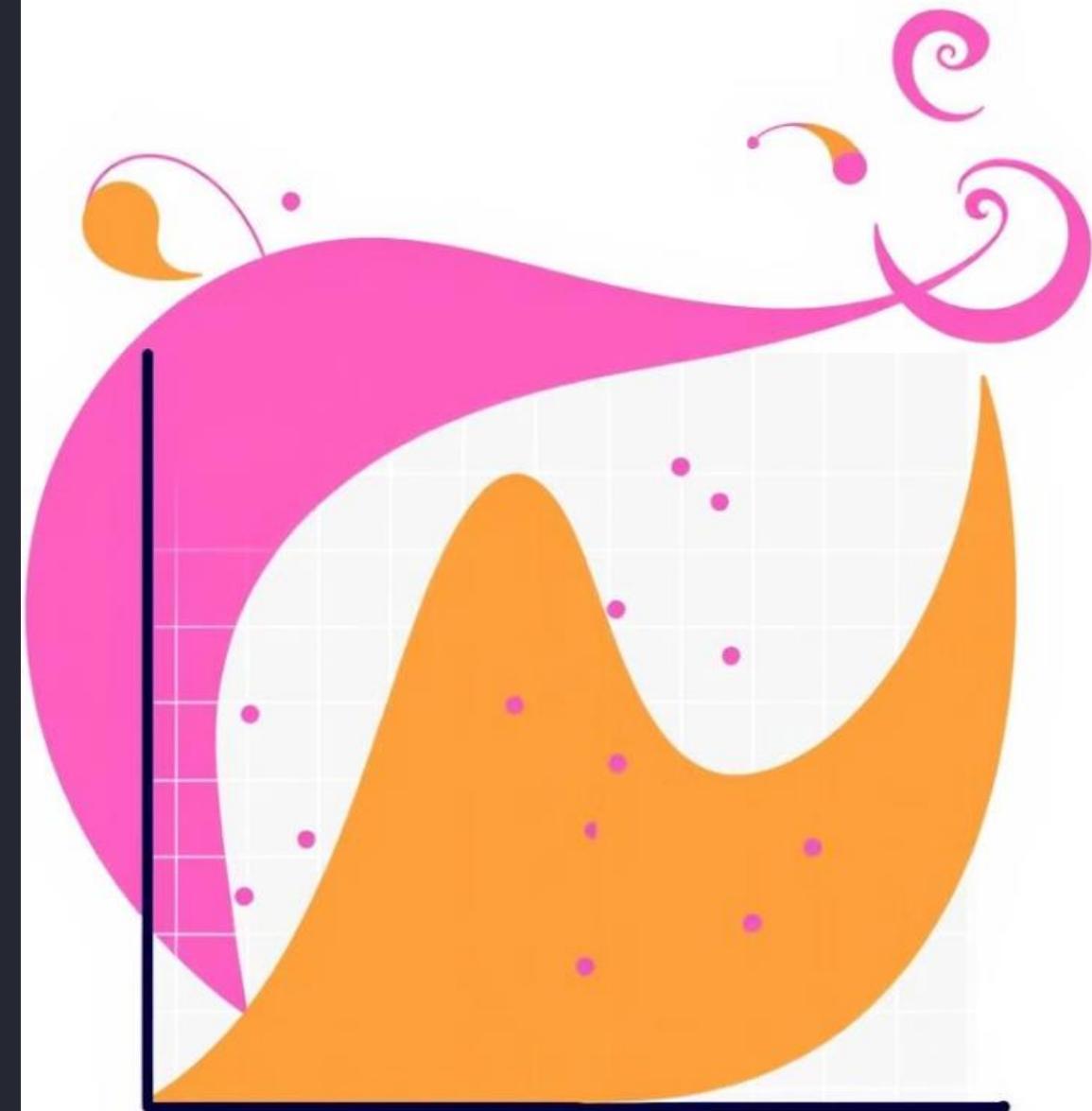
A continuación, se muestran algunas formas de medir el error.

Error absoluto medio (mean absolute error - MAE)

Es la diferencia entre el valor pronosticado y el valor real en cada punto pronosticado, conceptualmente la métrica de evaluación más fácil para problemas de regresión. Responde a la pregunta: “¿En qué medida te equivocaste en tus predicciones, en promedio?”.

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

donde y es el valor verdadero, \hat{y} es el valor predicho y n es el tamaño de la muestra.

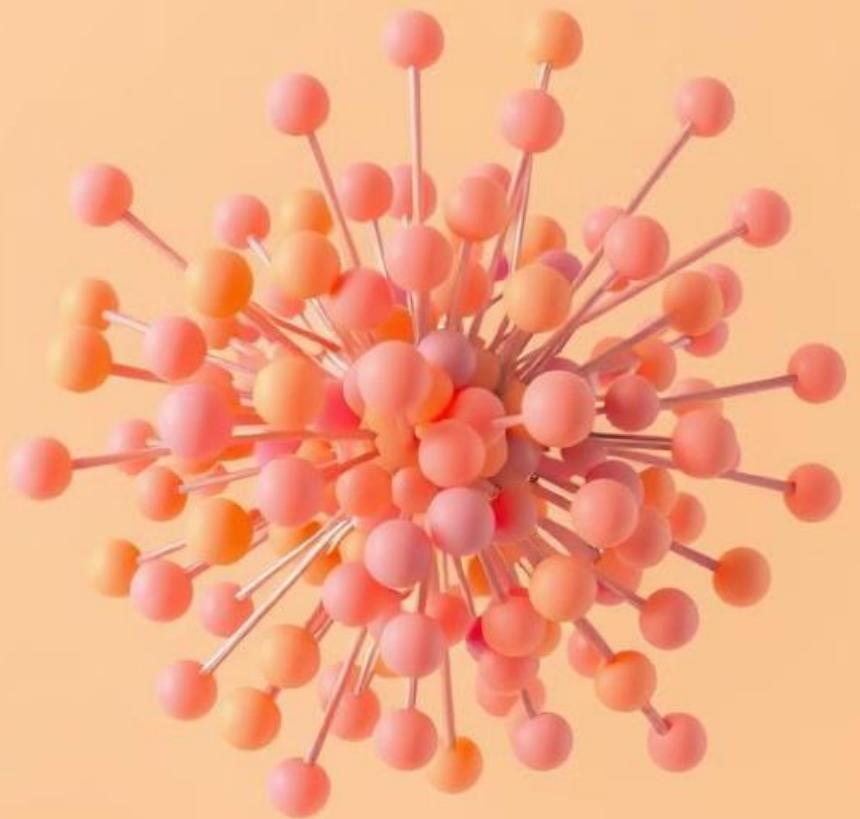


Suma error cuadrático medio (mean squared error - MSE)

Representa la distancia al cuadrado entre los valores reales y predichos. realizamos al cuadrado para evitar la cancelación de términos negativos y es útil elevar al cuadrado el error porque da un mayor peso a los valores atípicos, lo que da como resultado un gradiente suave para errores pequeños.

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

donde y es el valor verdadero, \hat{y} es el valor predicho y n es el tamaño de la muestra.





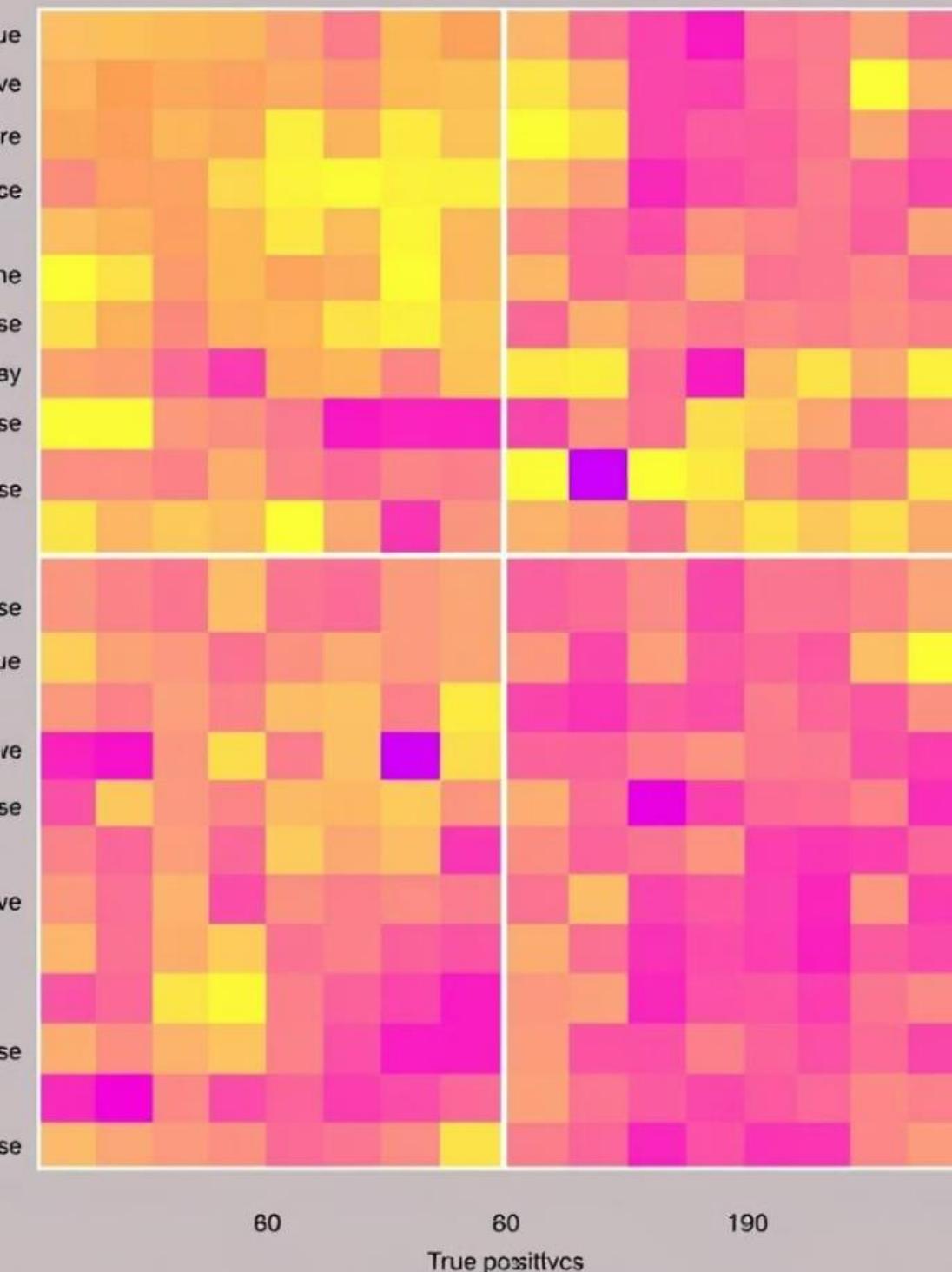
Suma de raíz del error cuadrático medio (root mean squared error - RMSE)

Esta medida da una idea de que nivel de error suele tener el sistema en sus predicciones, dando un peso mayor a los errores grandes. Como la métrica anterior nos da el resultado en unidades cuadradas, para poder interpretarlo más fácilmente sacamos la raíz cuadrada y de esta manera tenemos el valor en las unidades originales.

$$RMSE(y_i, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

donde y es el valor verdadero, \hat{y} es el valor predicho y n es el tamaño de la muestra.

Confusion Matrix



R2 Coeficiente de determinación

R2 es el coeficiente de determinación, nos indica que tanta variación tiene la variable dependiente que se puede predecir desde la variable independiente.

En otras palabras, que tan bien se ajusta el modelo a las observaciones reales que tenemos. Cuando usamos R2 todas las variables independientes que estén en nuestro modelo contribuyen a su valor.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

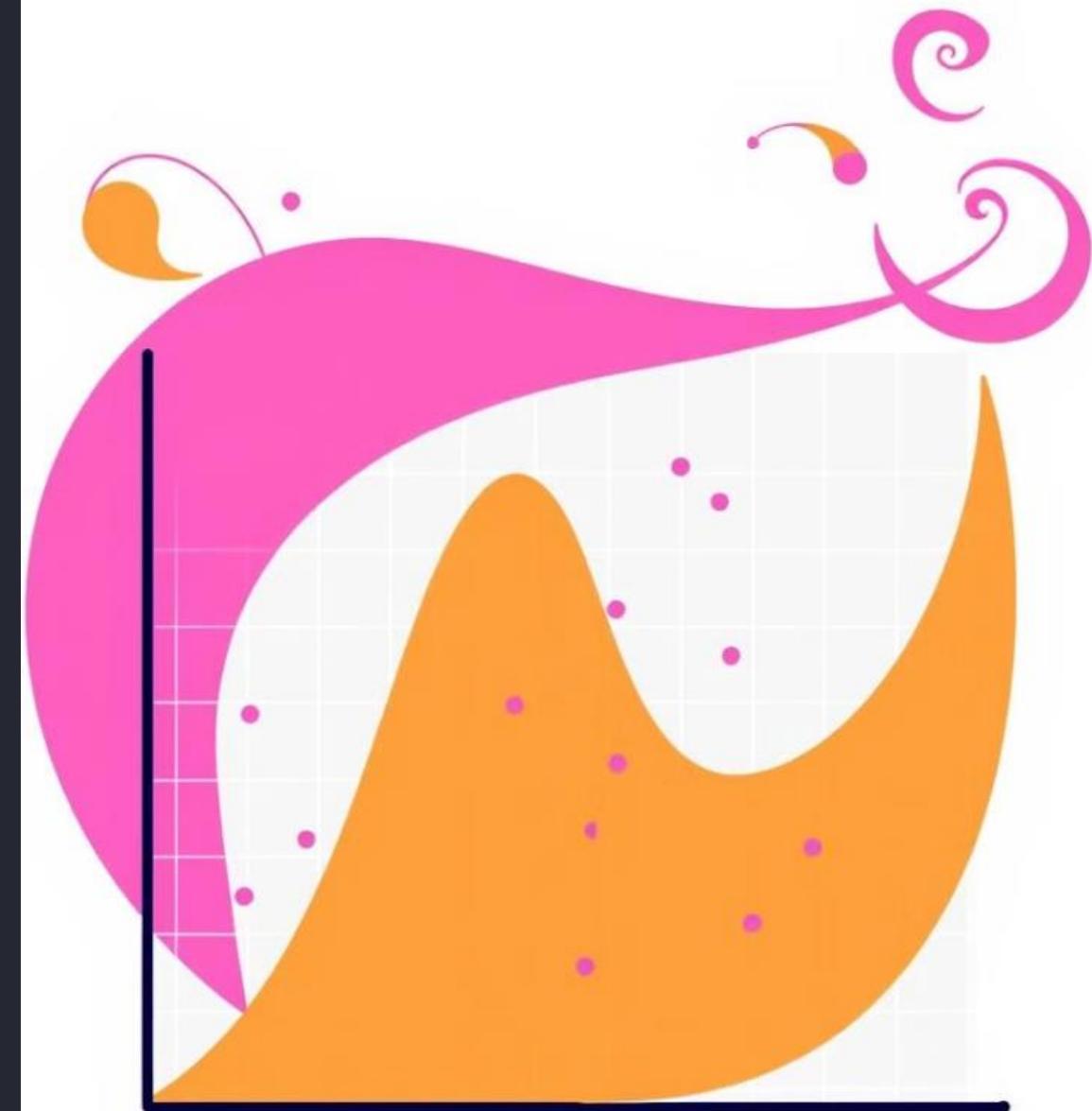
donde y es el valor verdadero, \hat{y} es el valor predicho, \bar{y} es el promedio de los valores verdaderos y n es el tamaño de la muestra.
score, devuelve el coeficiente de determinación R2 de la predicción.

Métricas en clasificación

- Matriz de confusión.
- La curva ROC (receiver operating characteristic).

Matriz de confusión

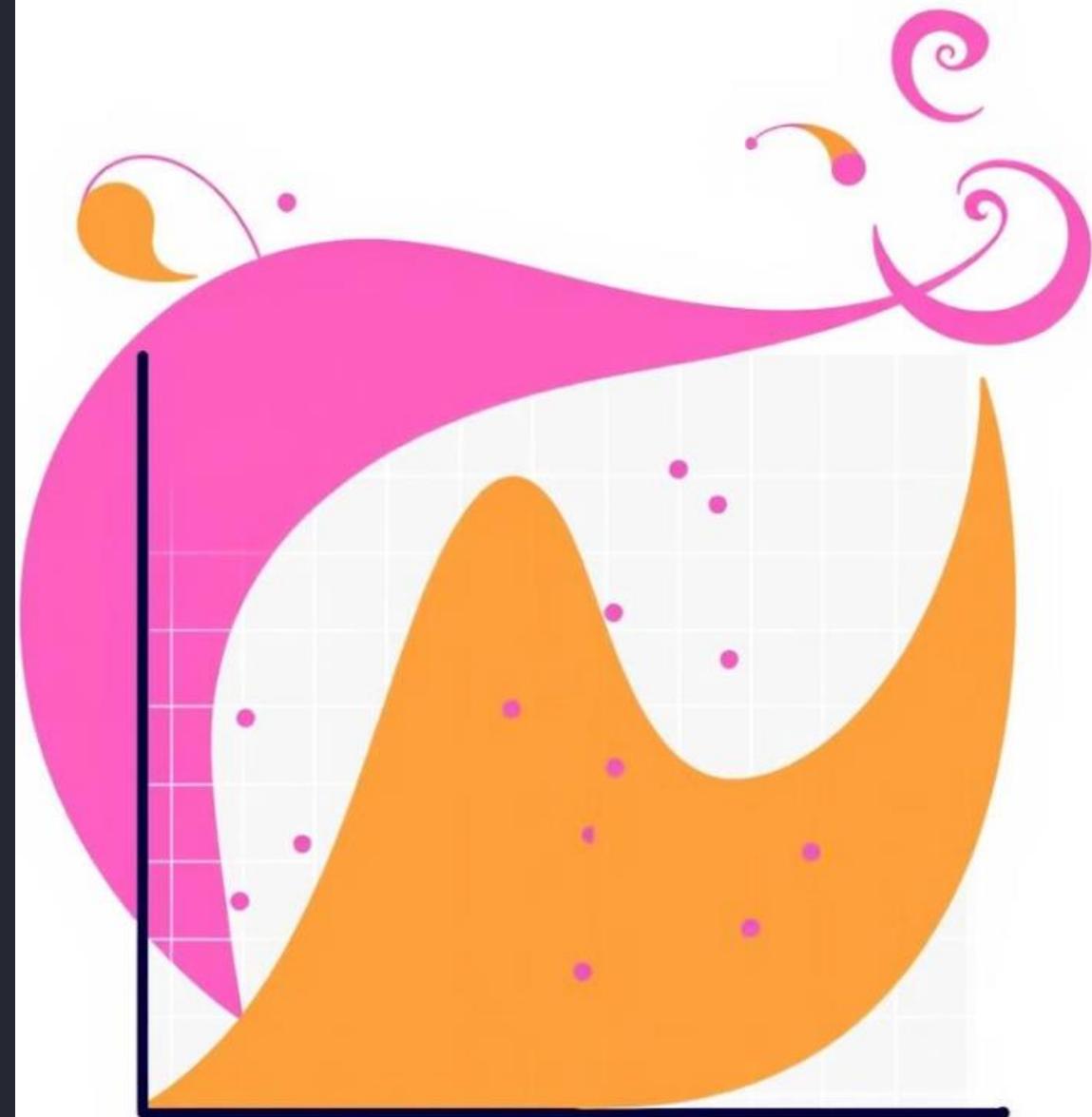
La matriz de confusión es una herramienta fundamental en la evaluación de modelos de clasificación. Esta matriz muestra de manera visual la calidad de las predicciones de un modelo comparadas con los valores reales. Es particularmente útil porque no solo muestra la precisión general del modelo, sino también como se comporta en términos de los diferentes tipos de errores que puede cometer.



Componentes de la matriz de confusión

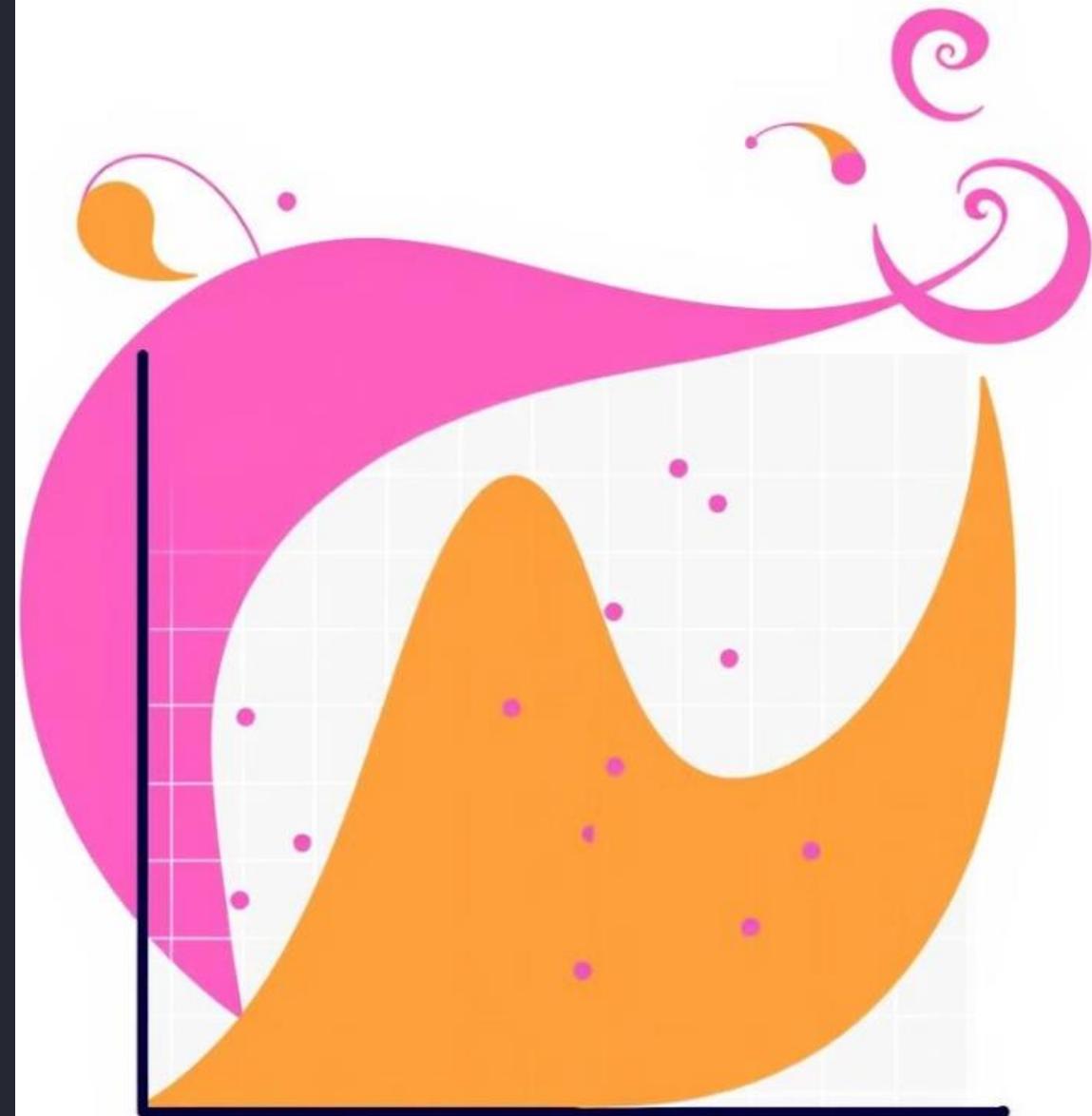
		Predicción		
		Positivo	Negativo	
Actual	Positivo	Verdaderos Positivos	Falsos Negativos	dato real = 1 dato predicho = 0
	Negativo	Falsos Positivos	Verdaderos Negativos	dato real = 0 dato predicho = 0
	dato real = 1 dato predicho = 1	dato real = 0 dato predicho = 1		

Supongamos que estamos trabajando con un problema de clasificación binaria, donde las clases son Positivo (P) y Negativo (N). La matriz de confusión tendrá la siguiente estructura:



Componentes de la matriz de confusión

- Verdaderos positivos (TP): Son los casos en los que el modelo predijo correctamente que la clase es positiva.
- Falsos positivos (FP): Son los casos en los que el modelo predijo incorrectamente que la clase es positiva (pero en realidad es negativa).
- Verdaderos negativos (TN): Son los casos en los que el modelo predijo correctamente que la clase es negativa.
- Falsos negativos (FN): Son los casos en los que el modelo predijo incorrectamente que la clase es negativa (pero en realidad es positiva).

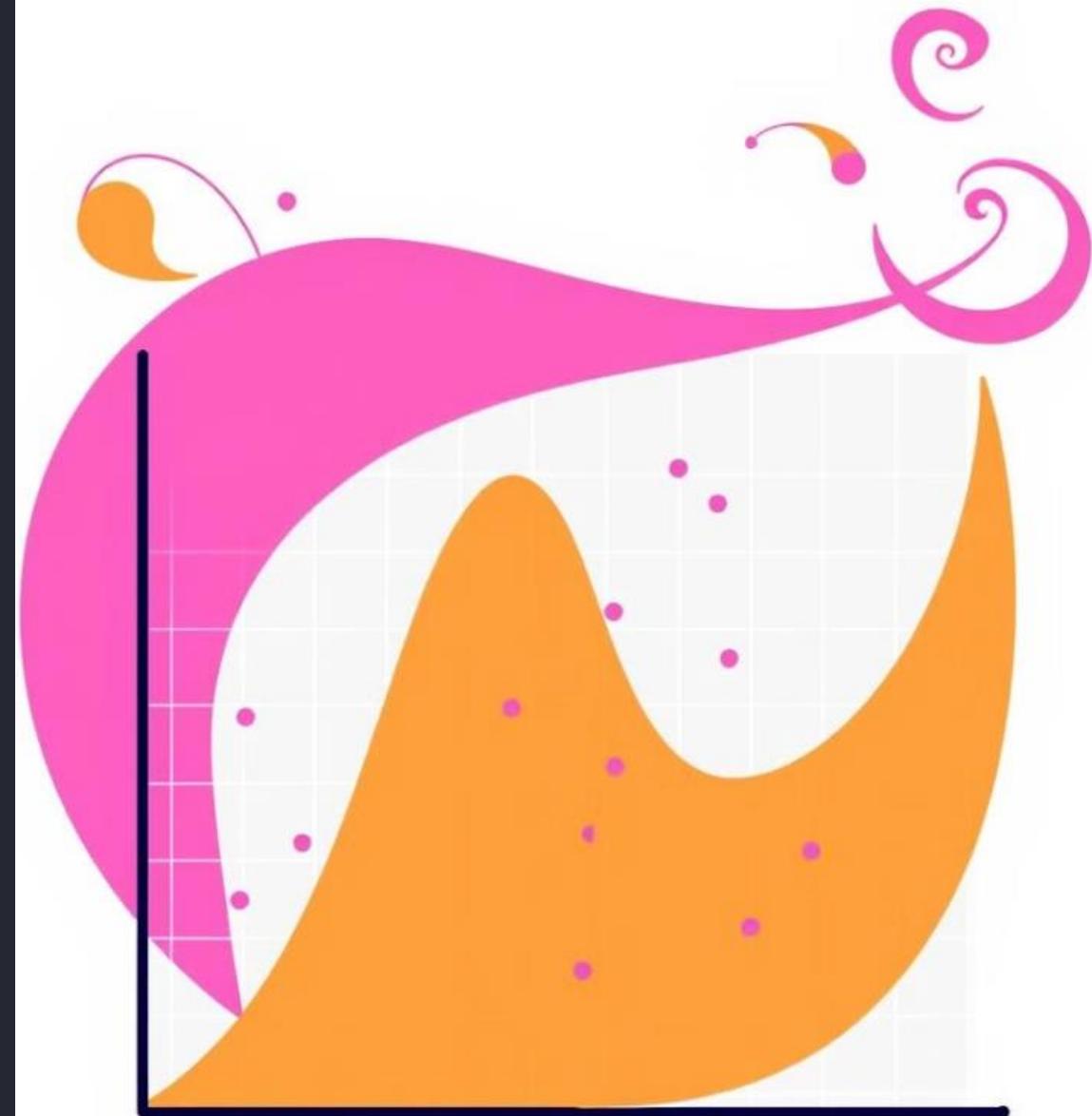


Componentes de la matriz de confusión

- Exactitud (Accuracy): Accuracy o exactitud es una métrica fundamental en problemas de clasificación que mide la proporción de predicciones correctas que hizo el modelo sobre el total de predicciones realizadas. Accuracy se calcula utilizando la siguiente fórmula:

$$Accuracy = \frac{\text{Predicciones correctas}}{\text{Total de predicciones}} = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy responde a la pregunta: “De todos los casos que el modelo predijo, ¿cuantos predijo correctamente?”. En otras palabras, mide la exactitud global del modelo en predecir tanto las instancias positivas como las negativas.

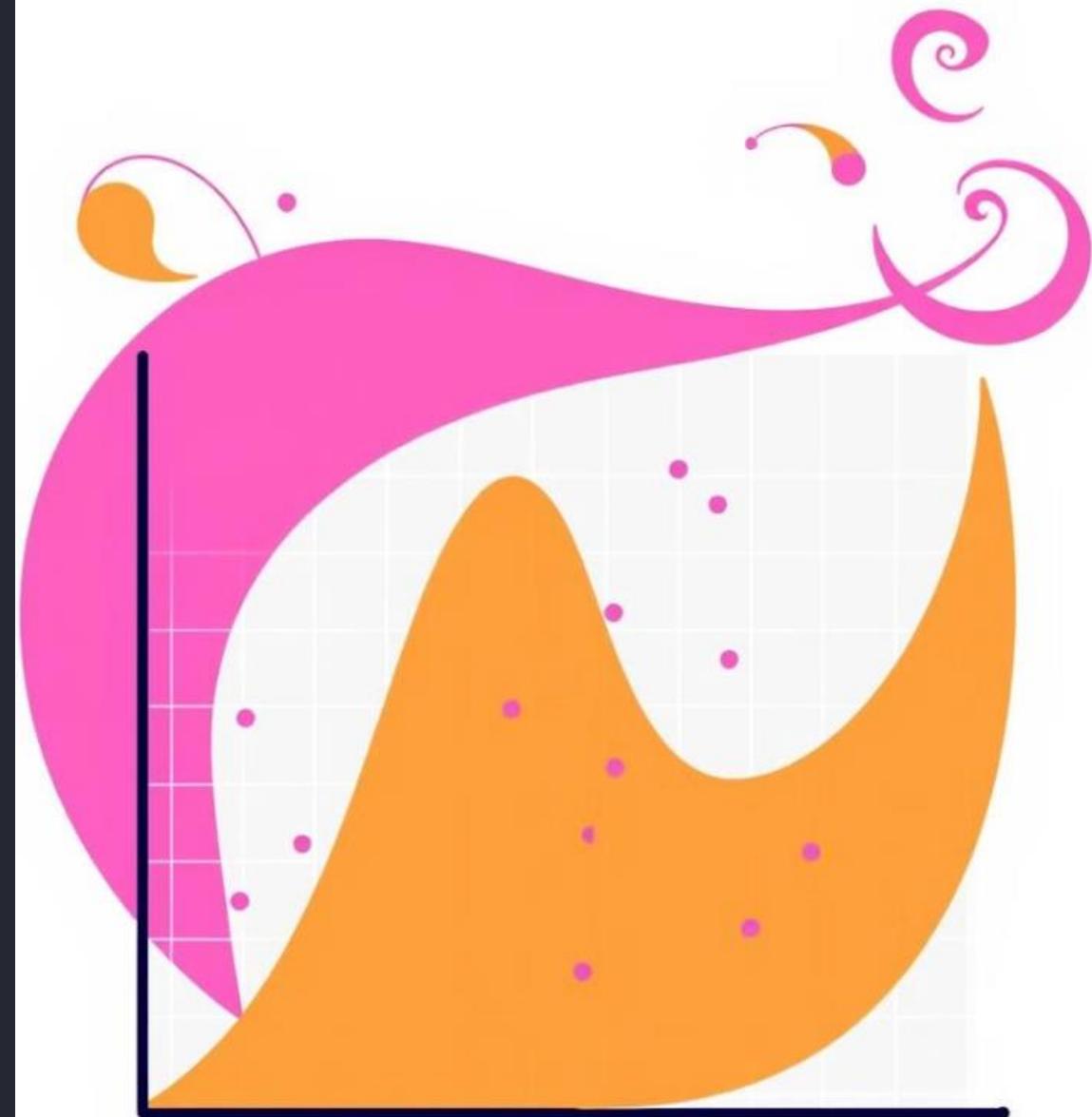


Componentes de la matriz de confusión

- Precisión: La precisión (precision en inglés) es una métrica de evaluación fundamental en problemas de clasificación que mide la proporción de ejemplos positivos correctamente identificados por el modelo entre todos los ejemplos que el modelo predijo como positivos. La precisión se calcula utilizando la siguiente fórmula::

$$Precision = \frac{TP}{TP + FP}$$

La precisión responde a la pregunta: “De todas las instancias que el modelo predijo como positivas, ¿cuántas realmente lo son?”. En otras palabras, mide la exactitud de las predicciones positivas del modelo.

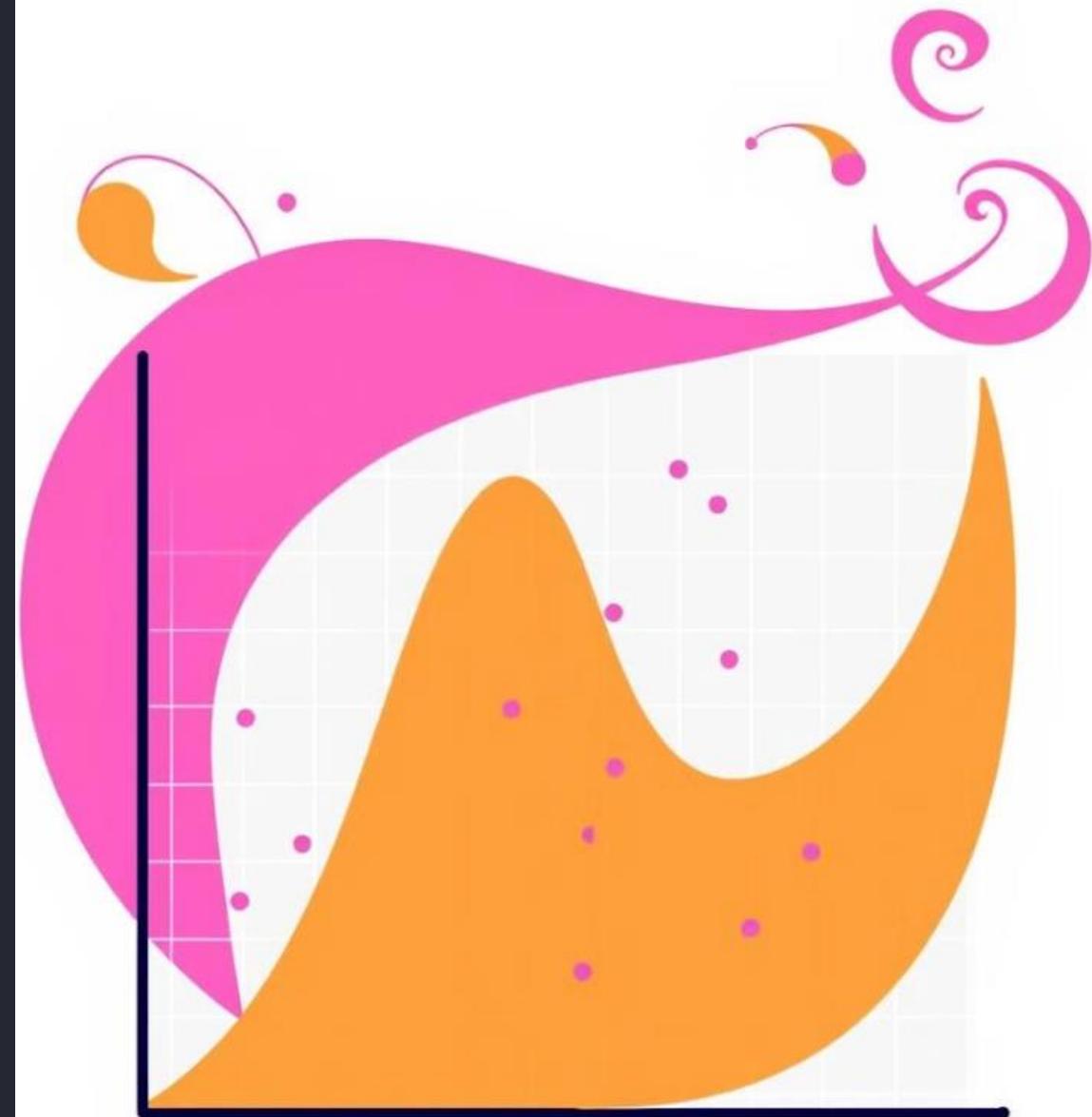


Componentes de la matriz de confusión

- Sensibilidad o Recall: El Recall, también conocido como sensibilidad o tasa de verdaderos positivos, es una métrica importante en problemas de clasificación que mide la proporción de ejemplos positivos que el modelo predijo correctamente entre todos los ejemplos que realmente son positivos. El recall se calcula utilizando la siguiente fórmula:

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

La sensibilidad responde a la pregunta: “De todas las instancias que realmente son positivas, ¿cuantas el modelo predijo correctamente como positivas?”. En otras palabras, mide la capacidad del modelo para identificar correctamente los casos positivos.

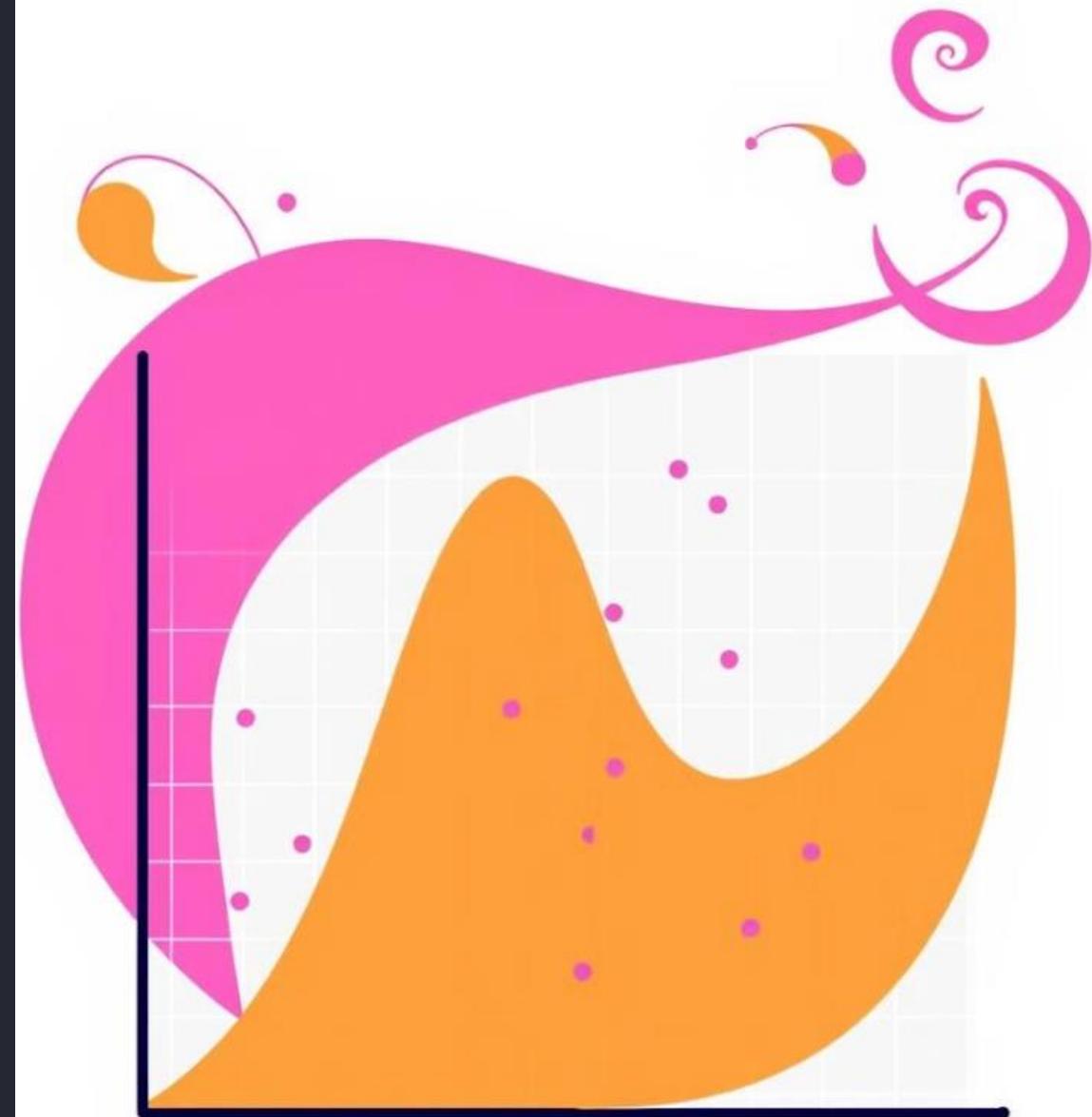


Componentes de la matriz de confusión

- F1 score: El F1-score es una métrica de evaluación que combina la precisión (precision) y el recall (sensibilidad) en una sola métrica para proporcionar una medida más equilibrada del desempeño de un modelo de clasificación. Es especialmente útil cuando hay clases desbalanceadas en los datos. El F1-score se calcula utilizando la siguiente fórmula:

$$F1 = \frac{2}{\frac{1}{Precision} + \frac{1}{Sensibilidad}}$$

El F1-score es especialmente útil cuando las clases en los datos de prueba están desbalanceadas. En tales casos, una alta precisión por sí sola podría no ser suficiente si el recall es bajo y viceversa.





La curva ROC (receiver operating characteristic).

La Curva ROC es una gráfica que muestra el rendimiento de un clasificador binario a diferentes umbrales de decisión. Es útil para analizar el equilibrio entre sensibilidad y especificidad.

En el eje X se representa la Tasa de Falsos Positivos (FPR):

$$FPR = \frac{FP}{FP + TN}$$

En el eje Y se representa la Tasa de Verdaderos Positivos (Recall o Sensibilidad, TPR):

$$TPR = \frac{TP}{TP + FN}$$

La curva ROC (receiver operating characteristic).

