

# Apuntes de clase 14 Marzo

Inteligencia Artificial - IC-6200  
Randy Fonseca  
Profesor: Msc. Steven Pacheco Portugu  z

Github: stevenpach10

## I. M  TRICAS

Las m  tricas sirven como punto de comparaci  n para determinar si un modelo es mejor que otro. Demuestran cuantitativamente la calidad y rendimiento de nuestro modelo. Pueden existir m  tricas en testing y en training.

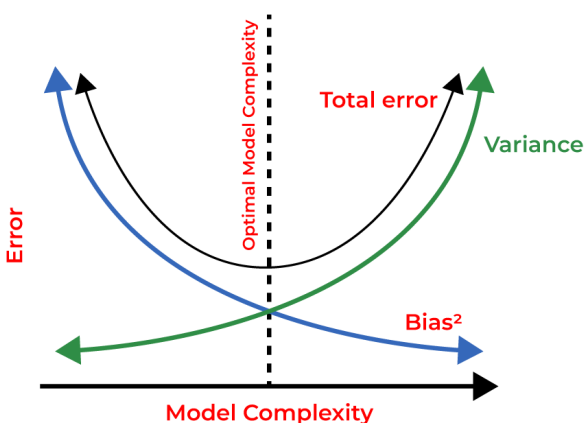


Fig. 1. Bias y variance pueden servir para comparar dos modelos

### A. Matriz de confusi  n

Determina el   xito de un modelo en la clasificaci  n de un problema en categor  as. El caso m  s b  sico ser   el de clasificaci  n binaria:

**Target class:** Es la etiqueta real de la muestra observada.

**Predicted class:** Es la etiqueta predicha por el modelo

Pueden darse 4 escenarios:

- True positive: Valor predicho positivo, valor objetivo positivo
- False positive: Valor predicho positivo, valor objetivo negativo
- True negative: Valor predicho negativo, valor objetivo negativo
- False negative: Valor predicho negativo, valor objetivo positivo

## Actual Values

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

### B. Accuracy

Traducido como exactitud, es el resultado de dividir las predicciones correctas entre la cantidad de predicciones totales.

$$\frac{TP + TN}{TP + TN + FP + FN}$$

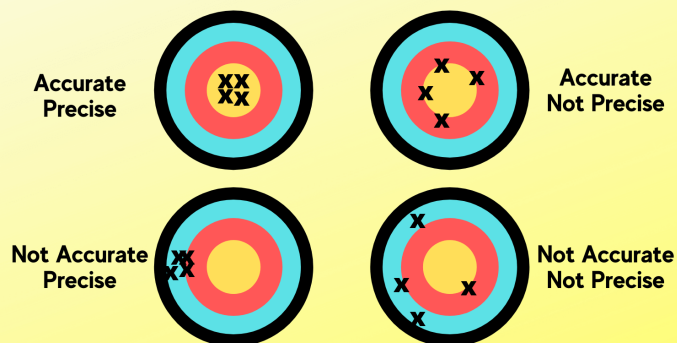
Es   til si las clases est  n balanceadas, y se les da la misma importancia.

### C. Precision

Mide la tasa de predicciones positivas correctas entre las predicciones positivas.

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

## Accuracy and Precision

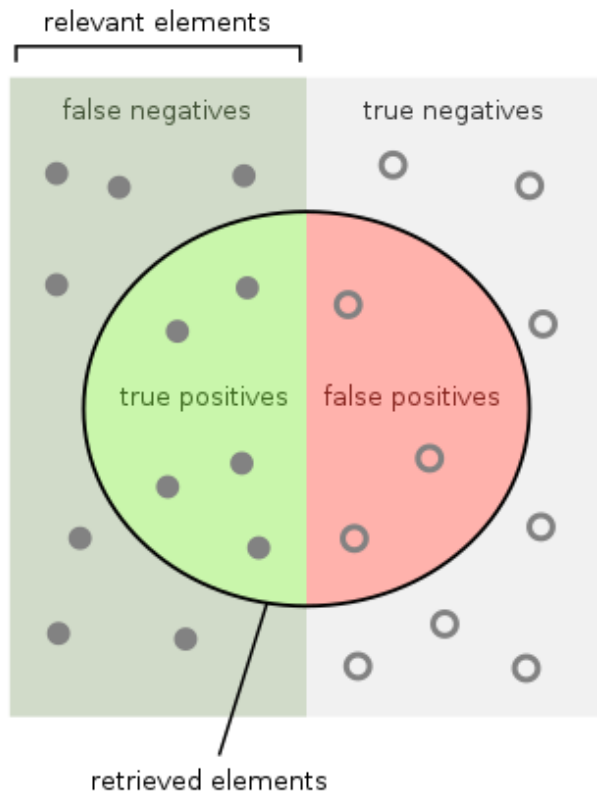


sciencenotes.org

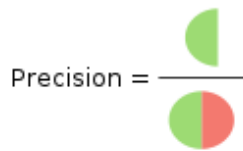
#### D. Recall (Sensitivity)

Mide el error de tipo II, predicciones correctas entre los ejemplos positivos

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

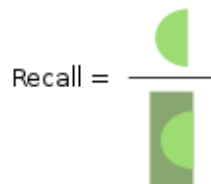


How many retrieved items are relevant?



Precision =

How many relevant items are retrieved?



Recall =

#### E. F1-Score

Permite determinar la calidad del modelo relacionando precision y recall. El puntaje ideal debería aproximarse a 1.

$$\frac{2 * Precision * Recall}{Precision + Recall}$$

#### F. Caso de estudio

El problema consiste en diagnosticar correctamente pacientes que tengan cáncer, la clase positiva será tener cáncer

y la clase negativa lo contrario; se tienen los siguientes datos de 1000 pacientes:

- Clase positiva: 30
- Clase negativa: 970

Y se obtuvieron los siguientes valores de salida del algoritmo:

- True positive: 25
- False negative: 5
- True negative: 950
- False positive: 20

Veamos las métricas obtenidas:

1) Accuracy::

$$\frac{TP + TN}{TP + TN + FP + FN} = \frac{25 + 950}{1000} = 0.975 = 97.5\%$$

Es un buen resultado, sin embargo la tarea es diagnosticar pacientes con cáncer, cómo lo hizo?

2) Recall:

$$\frac{TP}{TP + FN} = \frac{25}{25 + 5} = 0.833 = 83.3\%$$

El porcentaje disminuyó, según esta métrica hay más espacio de mejora.

3) Precision:

$$\frac{TP}{TP + FP} = \frac{25}{25 + 20} = 0.55 = 55\%$$

En este caso el porcentaje ha sido disminuido drásticamente, lo que nos dicen los datos aquí es que el modelo no está siendo muy certero, está diagnosticando muchos pacientes con cáncer que en realidad no tienen.

Aquí surge la pregunta de si es tolerable? Habría que entender que es mucho peor diagnosticar a un paciente con cáncer como sano, que diagnosticar una persona sana como enferma. Por lo tanto, el determinar cual métrica mejorar o no está estrechamente ligado al contexto del problema.

4) F1-Score:

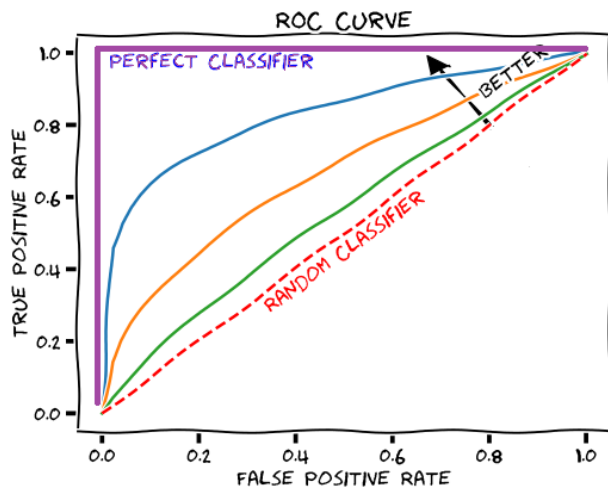
$$\frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * 0.55 * 0.833}{0.55 + 0.833} = 0.662 = 66.2\%$$

Teniendo en cuenta estas dos características, se obtiene un mejor panorama de la valoración del modelo. Hay que notar que un accuracy alto no significa que el modelo sea bueno, en general hay que identificar cuál es la característica más relevante a nuestro problema.

#### G. ROC y AUC

Receiver operating characteristic y area under the curve Imaginando una regresión logística, se debe evaluar la tasa de false positivos y true positive en función de un umbral. La idea detrás de eso es que debe existir una relación entre el umbral y esta tasa, cuanto más varíe el umbral más variarán los falsos positivos. El objetivo sería de que el área bajo la

curva se aproxime a 1, esto significaría que el modelo está abarcando los casos de manera correcta, que los positivos que seleccione realmente sean positivos; entre menos área bajo la curva exista más inexacto será.



#### REFERENCES

- [1] Bajaj, A. (2023) Performance metrics in machine learning [complete guide], neptune.ai. Available at: <https://neptune.ai/blog/performance-metrics-in-machine-learning-complete-guide> (Accessed: 16 March 2024).
- [2] Shah, D. (2023) Top performance metrics in machine learning: A comprehensive guide, V7. Available at: <https://www.v7labs.com/blog/performance-metrics-in-machine-learning> (Accessed: 16 March 2024).