



Módulo 5 – Aprendizaje de Máquina Supervisado

Métricas de Evaluación

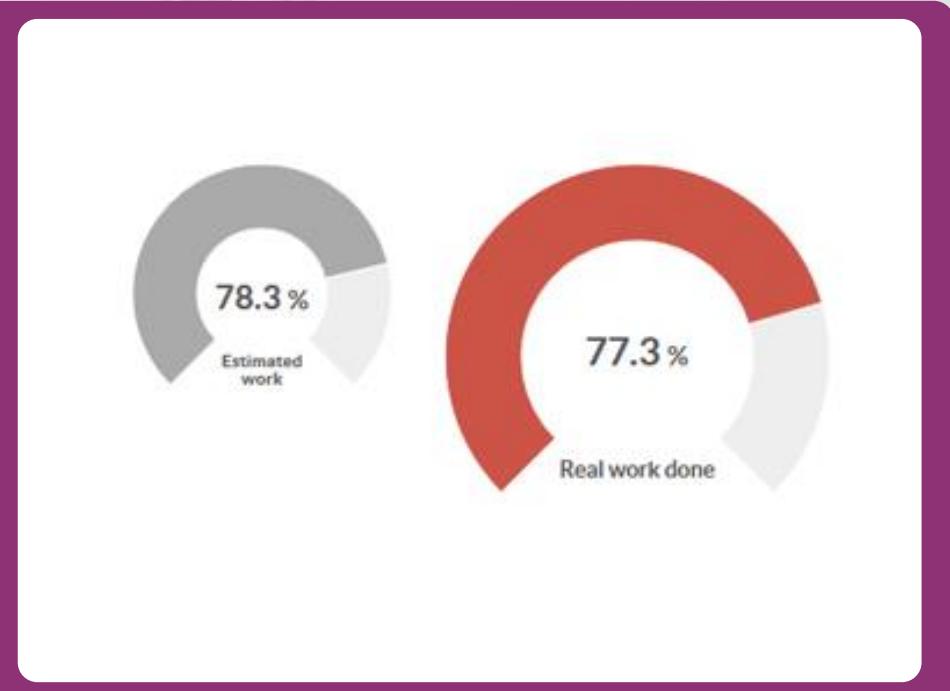
Especialización en Ciencia de Datos

Métricas de desempeño

Una **métrica de clasificación** es simplemente un número que mide el performance del modelo de aprendizaje de máquina que fue entrenado previamente en la labor de asignar clases o categorías a nuevas mediciones.

Típicamente, este valor se encuentra entre 0 y 1, donde un score de 1 corresponde a un modelo perfecto.

Existen varias métricas de desempeño que pueden ser utilizada, cada una de ellas tiene sus ventajas e inconvenientes.



Métricas de Desempeño

“Cada problema es distinto y,
por lo tanto, debe ser evaluado
de forma distinta.”



Accuracy

Es la métrica más utilizada al momento de evaluar un modelo predictivo, simplemente consiste en la proporción de aciertos que tuvo el modelo respecto de la cantidad total de predicciones.

$$Accuracy = \frac{\text{Total de Aciertos}}{\text{Total de Predicciones}}$$

Comente: ¿Es bueno o malo un Accuracy de 60%?

Comente: ¿Es bueno o malo un Accuracy de 60%?

!!!Depende del contexto del problema!!!!

Comente

Algunas ideas a considerar:

- Si se trata de una clasificación binaria, un 60% de accuracy es casi equivalente a lanzar una moneda al aire.
- Si se trata de una clasificación multi-clases, con 17 clases distintas de la variable objetivo, un 60% es bastante aceptable.
- Si el contexto está relacionado con el área de salud, por ejemplo, un test de detección de Covid, un 60% de accuracy es bastante malo.
- Por otro lado, si estamos prediciendo si una persona hará click en una publicidad enviada por mail, posiblemente sea un buen performance.

TAKE
NOTES



Matriz de Confusión

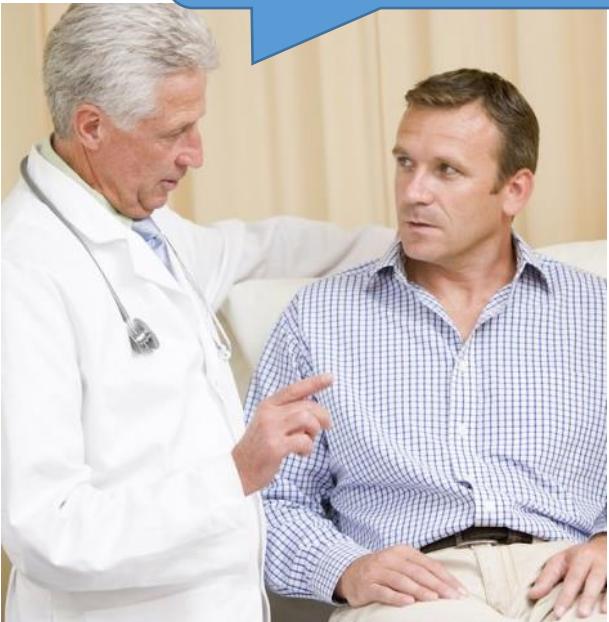
Suponga que está evaluando una clasificación binaria de un modelo predictivo. Una matriz de confusión muestra de forma explícita cuándo una clase es confundida con otra, lo cual nos permite trabajar de forma separada con los distintos tipos de error.

n=165	Predicted:		60
	NO	YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
		55	110

		Predicted class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

Tipos de Error

¡Felicitaciones! Usted
está embarazado



Falso Positivo
(Error tipo I)

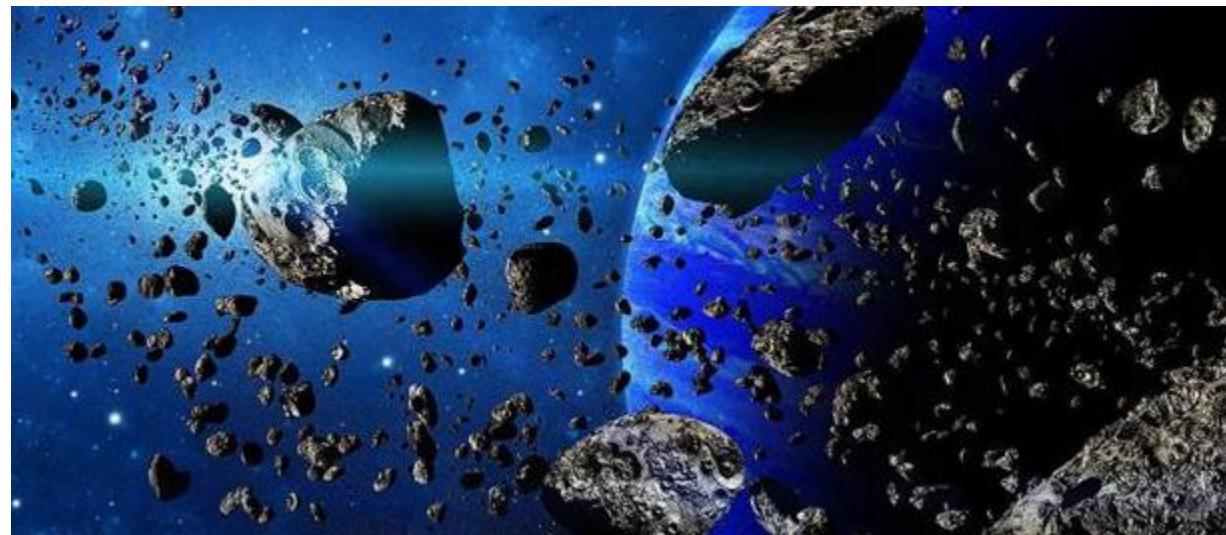
Despreocúpese, el test dice
que no está embarazada



Falso Negativo
(Error tipo II)

¿Por qué es importante el tipo de error?

Supongamos que estamos elaborando un modelo que predice si un asteroide chocará con el planeta tierra dentro de los próximos 3 meses o no, con un accuracy de 99%.



¿Es suficiente esta métrica?

¿Por qué es importante el tipo de error?

Como se puede observar, la métrica de accuracy no es suficiente, puesto que el algoritmo no tiene el poder predictivo esperado para detectar los eventos de asteroides que llegan a la tierra.

Predicciones

		Predicciones		
		Sí	No	
Actual	Sí	TP 3	FN 24	27
	No	FP 12	TN 326	338
		15	350	

	Predicted Negative	Predicted Positive
Actual Negative	True Negatives (TN)	False Positives (FP)
Actual Positive	False Negatives (FN)	True Positives (TP)

La **Exactitud o “Accuracy”** (AC) se refiere a la dispersión del conjunto de valores obtenidos a partir de mediciones repetidas de una magnitud. Cuanto menor es la dispersión mayor la precisión. Se representa por la proporción entre el número de predicciones correctas (tanto positivas como negativas) y el total de predicciones, y se calcula mediante la ecuación:



Accuracy



Exactitud

Predicciones

		Sí	No	
		TP	FN	
Actual	Sí	3	24	27
	No	FP	TN	338
		12	326	338
		15	350	

$$AC = \frac{TP+TN}{TOTAL} = \frac{3+326}{365} = 0.9013$$



KIBERNUM

Precision

La **Precisión** o, en inglés, “Precision” se refiere a lo cerca que está el resultado de una medición del valor verdadero. En términos estadísticos, la exactitud está relacionada con el sesgo de una estimación. También, se conoce como **Verdadero Positivo** (o “**True positive rate**”). Se representa por la proporción entre los positivos reales predichos por el algoritmo y todos los casos positivos. Es decir, cuántas ha predicho correctamente el algoritmo que lo son.



Precision



Precisión

Predicciones

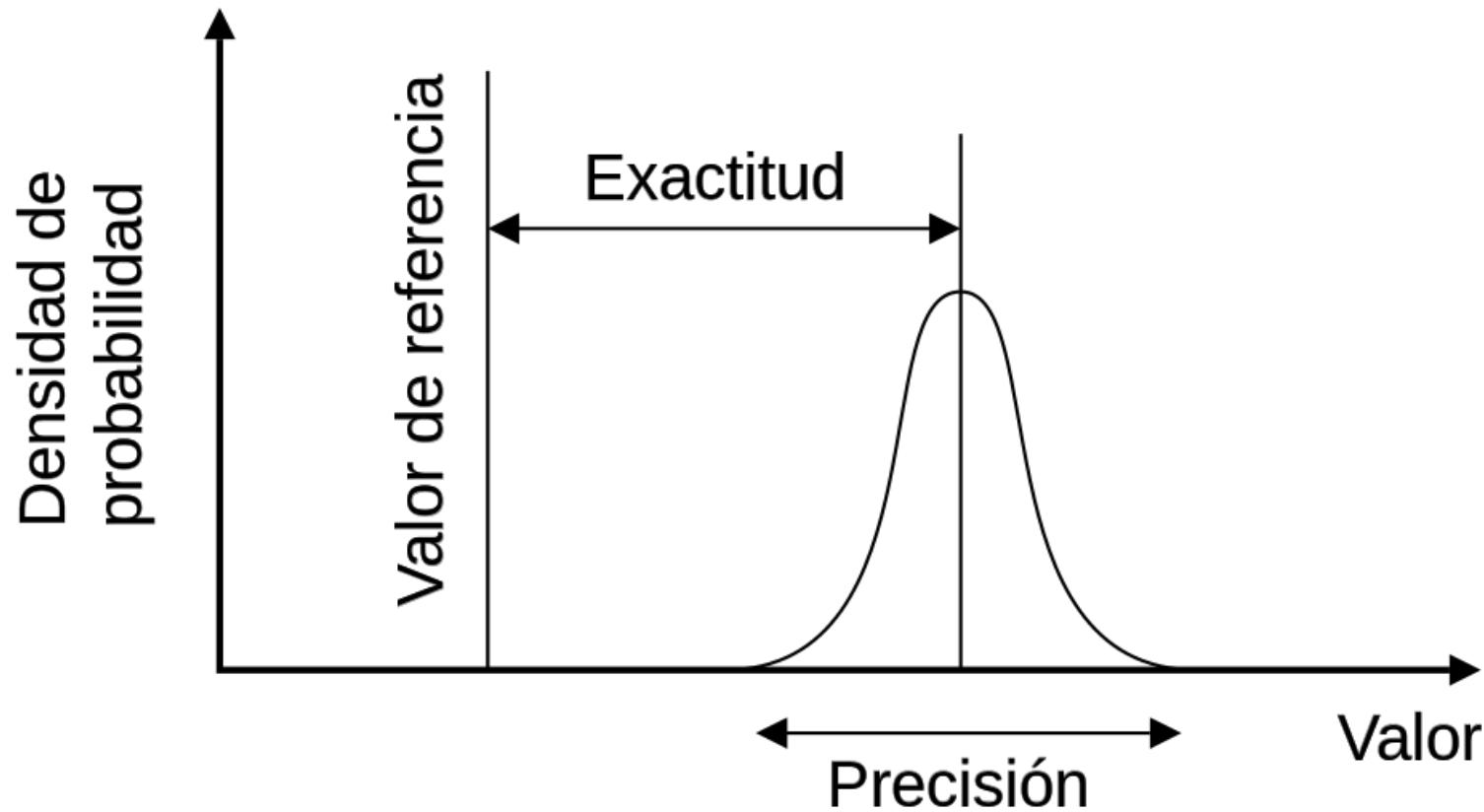
		Predicciones		
		Sí	No	
Actual	N=365			
	Sí	TP 3	FN 24	27
	No	FP 12	TN 326	338
		15	350	

$$P = \frac{TP}{FP+TP} = \frac{3}{12+3} = 0.2$$

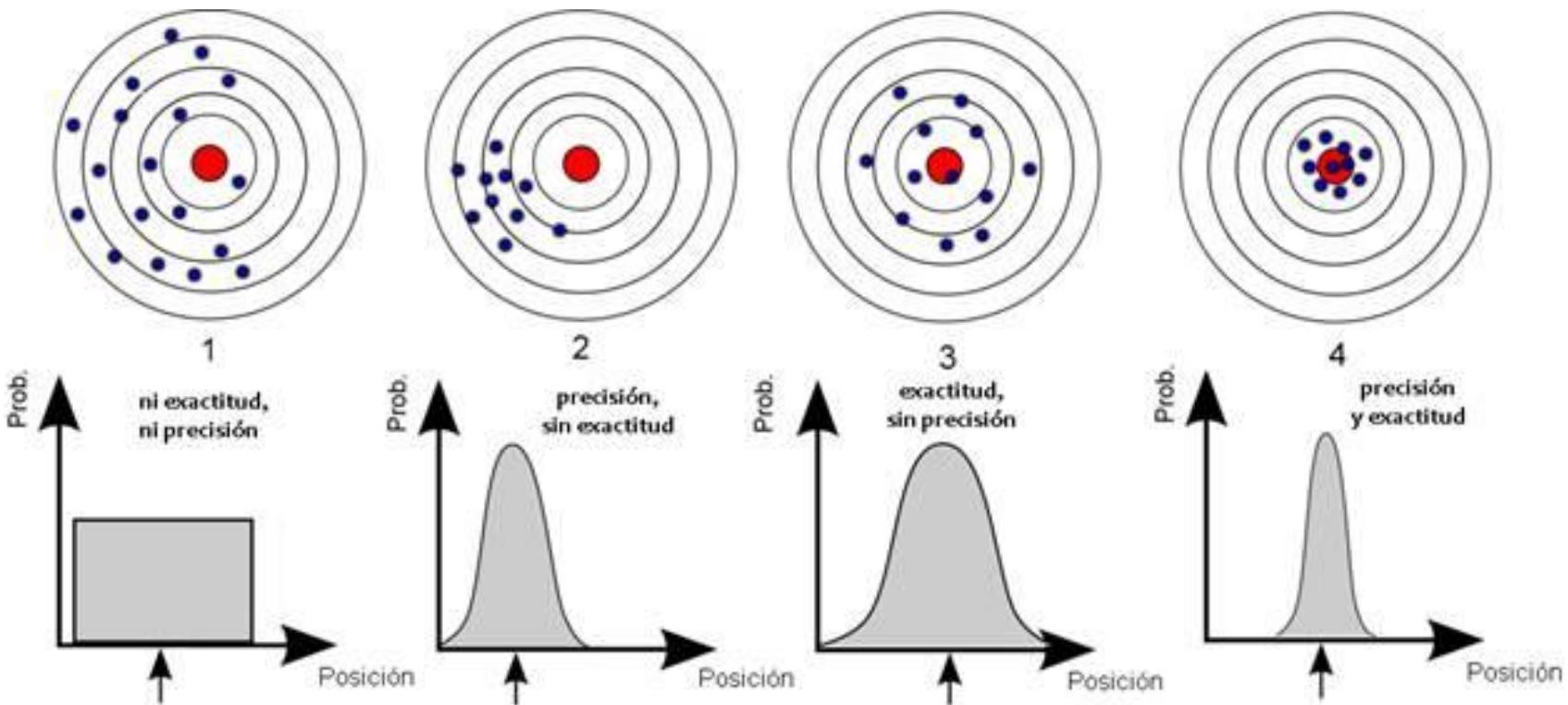


KIBERNUM

Exactitud y Precisión



Exactitud y Precisión



Recall

La **Sensibilidad** (“Recall”), también se conoce como **Tasa de Verdaderos Positivos (True Positive Rate)** (TP). Es la proporción de casos positivos que fueron correctamente identificadas por el algoritmo.



Recall



Sensibilidad

		Predicciones		
		Sí	No	
Actual	N=365			
	Sí	TP 3	FN 24	27
	No	FP 12	TN 326	338
		15	350	

$$TP = \frac{TP}{FN+TP} = \frac{3}{24+3} = 0.11$$

Specificity

La **Especificidad**, por otra parte, es la **Tasa de Verdaderos Negativos**, (“**true negative rate**” o **TN**). Se trata de los casos negativos que el algoritmo ha clasificado correctamente.



Specificity



Especificidad

Predicciones

		Predicciones		
		Sí	No	
Actual	N=365	TP	FN	
	Sí	3	24	27
	No	FP	TN	338
		12	326	
		15	350	

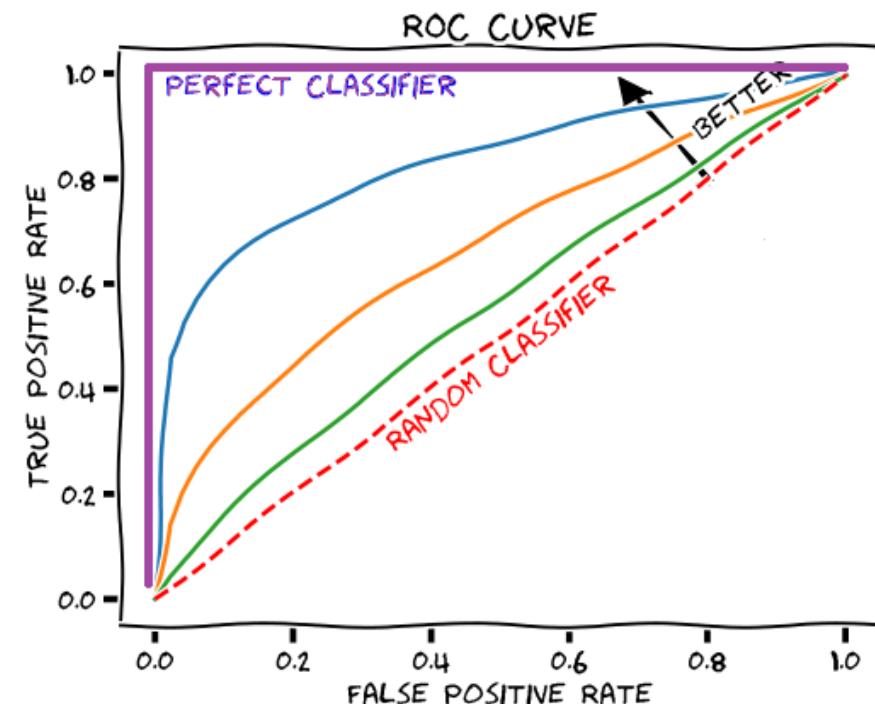
$$TN = \frac{TN}{TN+FP} = \frac{326}{326+12} = 0.96$$

Curva ROC

ROC: **R**eceiver **O**perating **C**haracteristic (Característica Operativa del Receptor). Es una gráfica que enfrenta el **ratio de falsos positivos (eje x)** con el **ratio de verdaderos positivos (eje y)**. Estos ratios los va obteniendo en función de una serie de umbrales definidos entre 0 y 1. En palabras comunes, enfrenta la «falsa alarma» vs la tasa de éxito.

- Permite comparar diferentes modelos para identificar cual otorga mejor rendimiento como clasificador.
- El **área debajo de la curva (AUC)** puede ser utilizado como resumen de la calidad del modelo.
- Valores pequeños en el eje X indican pocos falsos positivos y muchos verdaderos negativos
- Valores grandes en el eje Y indican elevados verdaderos positivos y pocos falsos negativos.

Más información: https://es.wikipedia.org/wiki/Curva_ROC

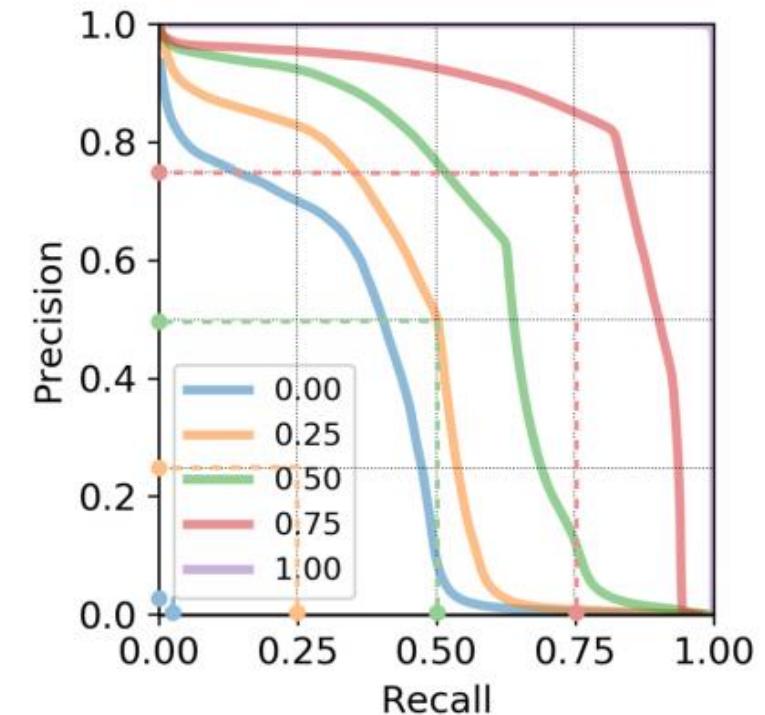


Curva Exactitud-Sensibilidad

La exactitud se calcula como el **número de verdaderos positivos entre la suma de verdaderos positivos y de falsos positivos**. Describe cómo de bueno es el modelo a la hora de predecir las salidas de la clase positiva. Otra forma de llamar a la precisión también se le llama poder predictivo positivo.

La **sensibilidad (recall)**, verdaderos positivos divididos entre la suma de verdaderos positivos y de falsos positivos.

Como resumen, la curva de precisión-sensibilidad enfrenta la precisión (eje y) con la sensibilidad (eje x) para diferentes umbrales.



Métricas en Python

```
from sklearn.metrics import confusion_matrix  
  
confusion_matrix(predictions, y_test)  
  
array([[12,  7],  
       [11, 37]], dtype=int64)
```

```
from sklearn.metrics import classification_report  
  
print(classification_report(predictions, y_test))
```

	precision	recall	f1-score	support
0	0.52	0.63	0.57	19
1	0.84	0.77	0.80	48
avg / total	0.75	0.73	0.74	67



Dudas y consultas



Fin Presentación



KIBERNUM