

ACÀMICA

TEMA DEL DÍA

# Evaluación de modelos

La evaluación de modelos es un aspecto **fundamental** en todo flujo de Data Science. Hoy aprenderemos algunos de los aspectos más importantes de este proceso.



# Agenda

---

Daily

Explicación: métricas de evaluación

**Break**

Hands-on training

Para pensar.... Ejercicio en conjunto.

Cierre



# Daily



Daily



## Sincronizando...

### Toolbox



¿Cómo te ha ido?  
¿Obstáculos?  
¿Cómo seguimos?

### Challenge



¿Cómo te ha ido?  
¿Obstáculos?  
¿Cómo seguimos?

# ¿Cómo evaluar si mi modelo está aprendiendo?



# ¿Cómo podemos evaluar si el modelo está *aprendiendo* o no de nuestros datos?<sup>1</sup>

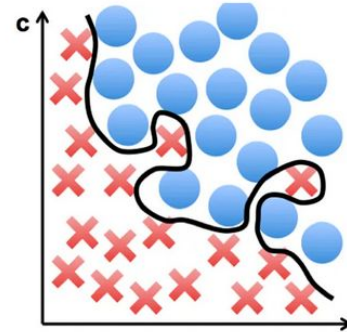
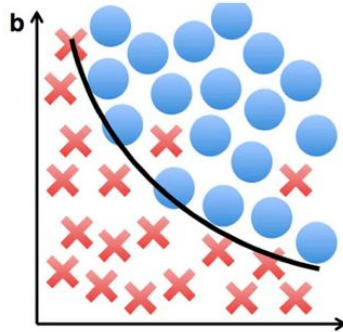
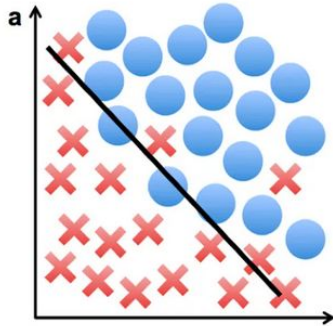
Una forma práctica de evaluar si nuestro modelo aprendió o no de nuestro datos es **observar su desempeño frente a nuevas instancias**.

Pero, ¿por qué necesitamos nuevas instancias y no usamos, simplemente, las instancias que usamos para entrenar?



<sup>1</sup>Podríamos preguntarnos también qué es aprender, pero por ahora vamos a ignorar esa pregunta.

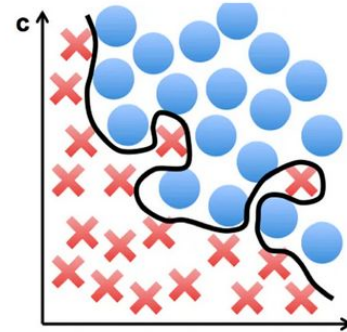
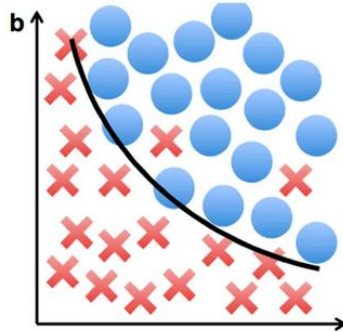
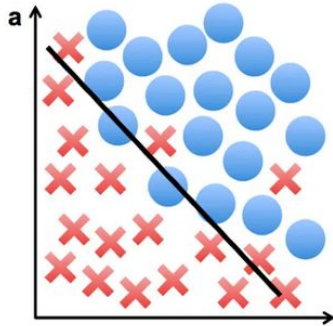
# Consideremos estas 3 situaciones...



¿Qué modelo les parece mejor?

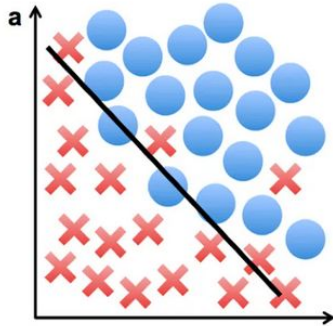


# Consideremos estas 3 situaciones...

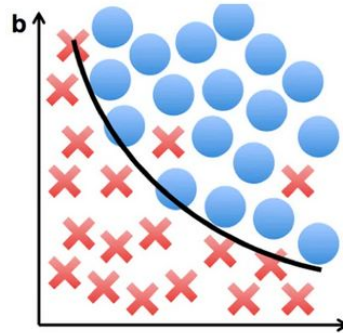


El **modelo a** es muy simple y no reproduce correctamente la frontera entre las clases. Llamaremos **underfitting** a esta situación.

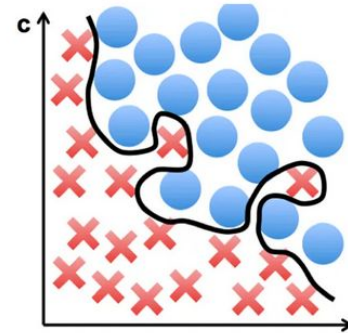
# Consideremos estas 3 situaciones...



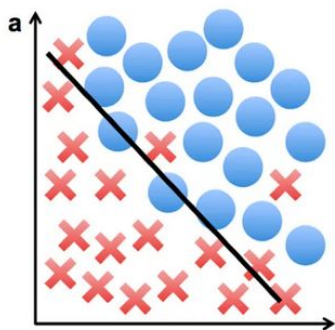
El **modelo a** es muy simple y no reproduce correctamente la frontera entre las clases. Llamaremos **underfitting** a esta situación.



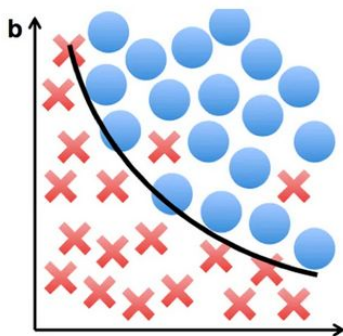
El **modelo b** tiene la complejidad suficiente para encontrar una frontera que parece apropiada para estos datos.



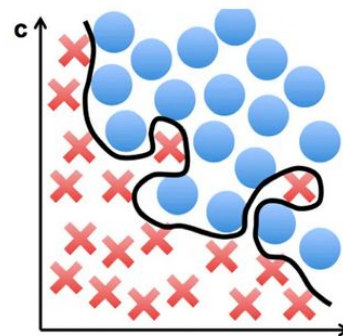
# Consideremos estas 3 situaciones...



El **modelo a** es muy simple y no reproduce correctamente la frontera entre las clases. Llamaremos **underfitting** a esta situación.



El **modelo b** tiene la complejidad suficiente para encontrar una frontera que parece apropiada para estos datos.



El **modelo c** parece muy flexible y se adaptó demasiado a los datos con los que fue entrenado. Llamaremos **overfitting** a esta situación.

WARNING

Entonces, podría ocurrir que el modelo se aprenda “de memoria” los datos con los que fue entrenado, por lo que **es importante evaluarlo con datos que nunca vio**.

Interviewer: What's your biggest strength?

Me: I'm an expert in machine learning.

Interviewer: What's  $6 + 10$ ?

Me: Zero.

Interviewer: Nowhere near, it's 16.

Me: It's 16.

Interviewer: Ok... What's  $10 + 20$ ?

Me: It's 16.

Muchos modelos que utilizaremos son muy flexibles y, de esas dos situaciones, **en general tendremos que preocuparnos más por el Sobreajuste (Overfitting).**\*

\*Más adelante, también veremos técnicas más complejas para evaluar correctamente nuestros modelos y prevenir el Overfitting y el Underfitting

## **Parámetros**

Hacen referencia a las variables que se ajustan automáticamente durante el proceso de entrenamiento.  
Esta es la parte en el que el modelo “aprende solo”.

## **Hiperparámetros**

Son valores que ajustamos y decidimos los/as Data Scientists.



## Para pensar, ¿cómo controlar el overfitting y el underfitting en los árboles de decisión?

El principal parámetros que controla si “overfiteamos” o “underfiteamos” es la profundidad del árbol. Para evitar el overfitting, existen algunos métodos:

- **Criterio de parada:** no construir más allá de cierta profundidad. Ésta es una de las reglas más usadas.
- **Podar:** construir el árbol entero. Podar las ramas cuando ello mejore la performance sobre datos separados
- Y más...

# ¿Cómo entreno un modelo?





# ¿Cómo entreno un modelo?

En nuestro flujo de trabajo, tendremos que emular una situación donde el modelo es entrenado con ciertos datos y luego es evaluado con datos nuevos.

¡Hacerlo es muy sencillo!

Antes de entrenar un modelo...

- 1. Separo una porción de los datos**
- 2. Evalúo el desempeño del modelo sobre esos datos.**

---

En general, los datos se separan al azar para evitar cualquier orden o estructura subyacente en los datos<sup>1</sup>.

<sup>1</sup> En algunos problemas, por ejemplo cuando queremos entrenar un modelo que haga predicciones a futuro, esto no es válido.

## En Scikit-Learn

¡Esta función es tan importante que viene en todos los entornos de desarrollo de Machine Learning!

En Scikit-Learn, la función se llama [train\\_test\\_split](#).

```
sklearn.model_selection.train_test_split
```

```
sklearn.model_selection. train_test_split (*arrays, **options)
```

[\[source\]](#)

# Métricas de evaluación para Clasificación



¿Cómo evaluamos los resultados de una clasificación?

# ¿Cómo evaluamos los resultados de una **clasificación**?

Exactitud

Matriz de Confusión

Precisión/Exhaustividad

F-Score

# Clasificación Binaria

Problema general: separar los elementos de un conjunto en **dos grupos** bajo cierta/s regla/s de clasificación.

# Clasificación Binaria

Problema general: separar los elementos de un conjunto en **dos grupos** bajo cierta/s regla/s de clasificación.

Ejemplos:



## Examen médico

Enfermo / No enfermo  
Test de embarazo



## Educación

Aprobado / No  
aprobado



## Bromatología

Apto / No apto para  
consumo



## Control calidad

Seguro / No seguro

## Clasificación Binaria

En general, nos interesa un grupo en particular.



## Clasificación Binaria

¿Qué puede pasar con **un** test?  
(ejemplo: test de embarazo)

# Clasificación Binaria

¿Qué puede pasar con **un** test?  
(ejemplo: test de embarazo)

**Verdadero positivo (acierto)**

test positivo, paciente embarazada

**Verdadero negativo (acierto)**

test negativo, paciente **no** embarazada

# Clasificación Binaria

¿Qué puede pasar con **un** test?  
(ejemplo: test de embarazo)

**Verdadero positivo (acierto)**

test positivo, paciente embarazada

**Verdadero negativo (acierto)**

test negativo, paciente no embarazada

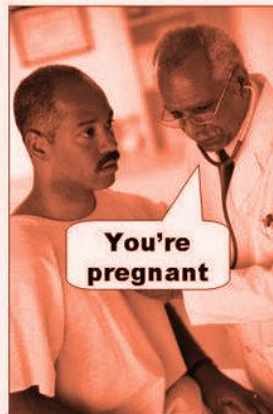
**Falso positivo (error tipo 1)**

test positivo, paciente **no** embarazada

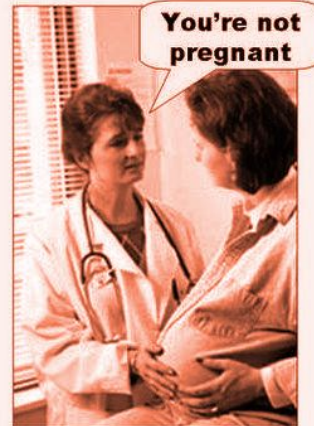
**Falso negativo (error tipo 2)**

test negativo, paciente embarazada

**Type I error**  
(false positive)

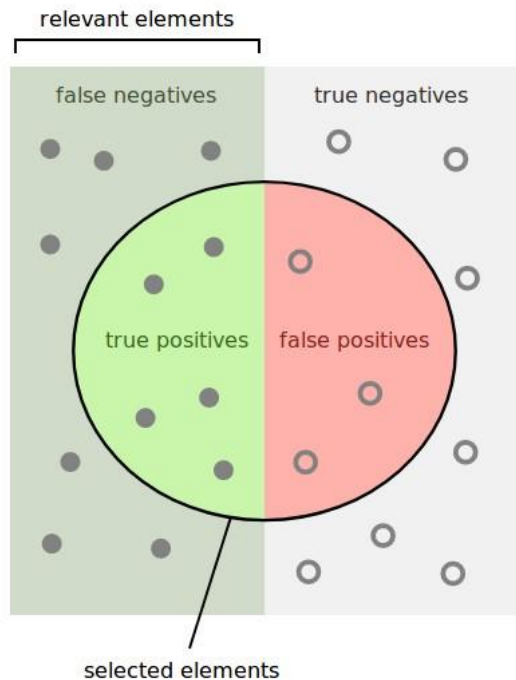


**Type II error**  
(false negative)



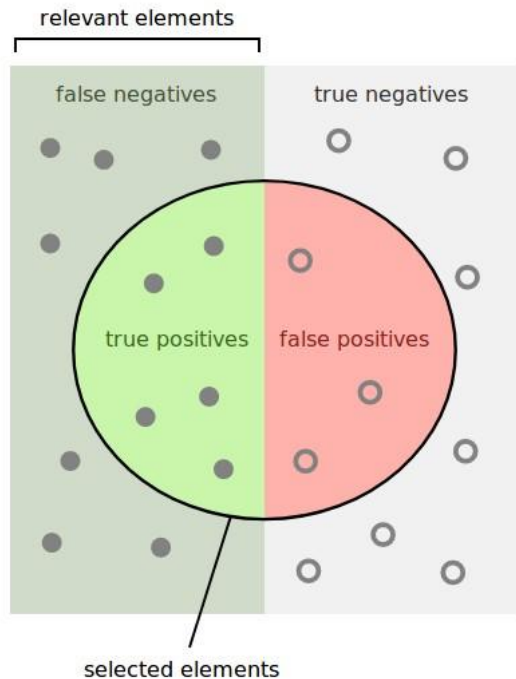
# Clasificación Binaria

Cuando hacemos muchos tests...



# Clasificación Binaria

## Cuando hacemos muchos tests...



How many selected items are relevant?

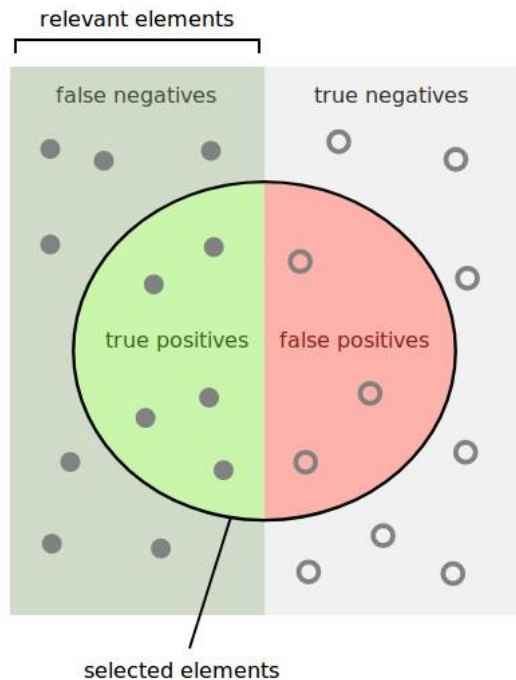
$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?

$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

# Clasificación Binaria

## Cuando hacemos muchos tests...



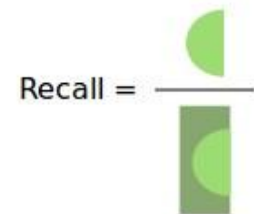
Fuente: Wikipedia

How many selected items are relevant?



$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$

How many relevant items are selected?



$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

$$\text{Precisión} = \frac{\text{Aciertos}}{\text{Aciertos} + \text{Falsos Positivos}}$$

$$\text{Exhaustividad} = \frac{\text{Aciertos}}{\text{Aciertos} + \text{Falsos Negativos}}$$

# Clasificación Binaria • Matriz de confusión

Ejemplo: Titanic

		Clase Predicha	
		Clase 1	Clase 2
Clase Verdadera	Clase 1	Elementos de la clase 1 correctamente identificados	Elementos de la clase 1 identificados como clase 2
	Clase 2	Elementos de la clase 2 identificados como clase 1	Elementos de la clase 2 correctamente identificados

# Clasificación Binaria • Matriz de confusión

Ejemplo: Titanic

		Clase Predicha	
		Clase 1	Clase 2
Clase Verdadera	Clase 1	Elementos de la clase 1 correctamente identificados	Elementos de la clase 1 identificados como clase 2
	Clase 2	Elementos de la clase 2 identificados como clase 1	Elementos de la clase 2 correctamente identificados

**¡Tiene toda la información que necesitamos!**



# Clasificación Binaria • Matriz de confusión

Ejemplo: Titanic

		Clase Predicha	
		No Sobrevivieron	Sobrevivieron
Clase Verdadera	No Sobrevivieron	513	110
	Sobrevivieron	103	283

**Ejercicio:** para la clase “sobrevivieron”, indicar Aciertos, Falsos Positivos, Falsos Negativos. Calcular Precisión y Exhaustividad.

# Clasificación Binaria • Matriz de confusión

Ejemplo: Titanic

		Clase Predicha	
		No Sobrevivieron	Sobrevivieron
Clase Verdadera	No Sobrevivieron	513	110
	Sobrevivieron	103	283

Falsos positivos

Falsos negativos

Aciertos

Precisión:  $TP/(TP + FP) = 283/(283+110) = 0.72$

**Exhaustividad:  $TP/(TP + FN) = 283/(283+103) = 0.73$**

# Clasificación Binaria • Matriz de confusión

Ejemplo: Titanic

		Clase Predicha		
		No Sobrevivieron	Sobrevivieron	
Clase Verdadera	No Sobrevivieron	513	110	No aciertos
	Sobrevivieron	103	283	Aciertos

$$\text{Exactitud} = \text{Aciertos} / \text{Total} = (513 + 283) / (513 + 283 + 110 + 103) = 0.789$$

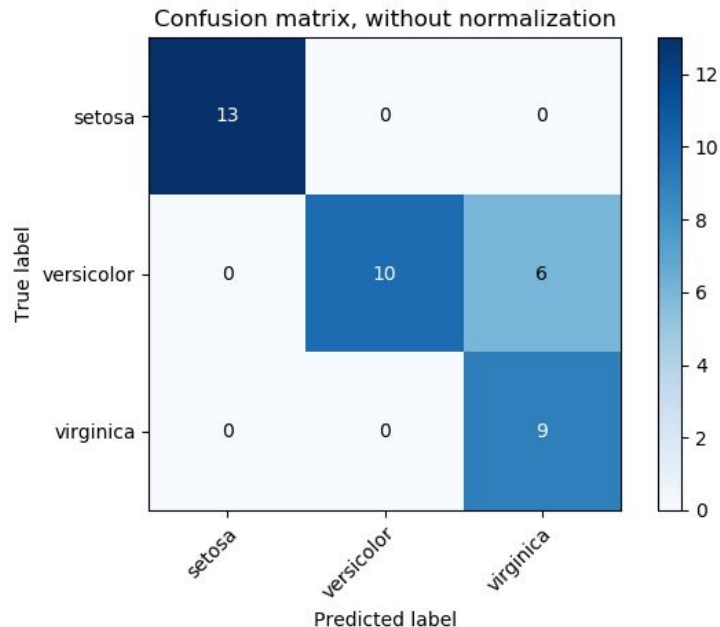
# Clasificación Multiclase

¿Cómo se generalizan los conceptos?

**Acierto, Verdaderos Negativo, Falsos Positivos y Negativos:** tengo que elegir una clase de interés.

**Precisión y Exhaustividad:** ídem, por clase.

**Exactitud:** Sigue valiendo

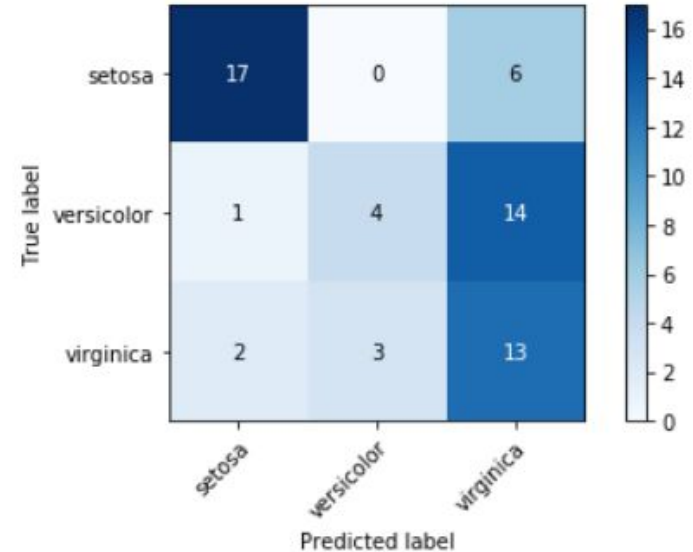


# Clasificación Multiclase

## Ejercicio

Dada la siguiente Matriz de Confusión:

- Elegir una clase e indicar Aciertos, Falsos Positivos, Falsos Negativos para esa clase.
- Calcular Precisión y Exhaustividad para la clase elegida.
- Calcular F-Score para la clase elegida
- Calcular la exactitud del modelo.



# Clasificación Multiclase • Comentarios

Elegir la métrica correcta para nuestro problema es parte del trabajo que un Data Scientist tiene que hacer. A veces queremos favorecer precisión, otras veces exhaustividad.

**¿En qué circunstancias les parece que preferimos una sobre otra?**

Más adelante veremos:

- Curva ROC y área bajo la curva
- Funciones de costo

En scikit-learn:

<https://scikit-learn.org/stable/modules/classes.html#sklearn-metrics-metrics>

A close-up photograph of a white ceramic cup filled with a latte. The surface of the milk is decorated with intricate latte art, featuring a central heart shape surrounded by concentric, wavy lines. The cup is placed on a matching white saucer. In the background, a white napkin and a silver fork are visible, though they are out of focus. The overall lighting is soft and even, highlighting the textures of the coffee and the smooth surface of the cup.

**¡BREAK!**

---



# Hands-on training







DS\_Toolbox\_13\_Evaluación.ipynb



**Para pensar...**

**proponer un test  
100% exhaustivo y  
otro 100% preciso.**

**¿Son útiles estos tests?**

## Para pensar...

**...proponer un test 100% exhaustivo y otro 100% preciso. ¿Son útiles estos tests?**

Problema: ninguna por sí sola alcanza para evaluar el desempeño del test, ya que precisión y exhaustividad compiten entre sí.

Objetivo: encontrar un compromiso entre ambas métricas.

## Para pensar...

...proponer un test 100% exhaustivo y otro 100% preciso. ¿Son útiles estos tests?

Problema: ninguna por sí sola alcanza para evaluar el desempeño del test, ya que precisión y exhaustividad compiten entre sí.


Objetivo: encontrar un compromiso entre ambas métricas.

$$\text{F-SCORE} \longrightarrow F = 2 \times \frac{\text{precisión} \times \text{exhaustividad}}{\text{precisión} + \text{exhaustividad}}$$

# Recursos



# Evaluación de modelos

- [Python Data Science Handbook](#) - Capítulo 5, “Machine Learning”, sección “[Hyperparameters and Model Validation](#)”. Adelanta algunos conceptos que veremos en el segundo Sprint, como Validación Cruzada, y utiliza una terminología levemente distinta, pero es una introducción muy comprensible. La recomendamos mucho.
-  [The Data Science Design Manual](#), de Steven Skienna.

# Estado de tu proyecto

## Primer Modelo de Machine Learning





# Estado del proyecto

Identifica el estado de avance de tu proyecto. Te será muy útil saberlo para poder esquematizar tu trabajo en las próximas semanas y entregar a tiempo.

Puedes estimar tu grado de avance de la siguiente manera:

**Grupo 25%:** Recién estoy comenzando

**Grupo 50%:** He avanzado pero aún me falta

**Grupo 75%:** Estoy muy avanzado

**Grupo 100%:** He finalizado el proyecto



# Para la próxima

---

- Termina el notebook de hoy.
- Lee la Toolbox 14 y carga las dudas que tengas al Trello.

ACÀMICA