

# Unidad I - Introducción al Aprendizaje Automático

Germán Braun

Facultado de Informática - Universidad Nacional del Comahue

`german.braun@fi.uncoma.edu.ar`

12 de septiembre de 2025

# Agenda

- 1 Introducción - Orígenes - Usos
- 2 Aprendiendo patrones a partir de datos
- 3 Buenas Prácticas y Evaluación de Performance
- 4 Aspectos Éticos

# Introducción - Orígenes - Usos

## machine learning

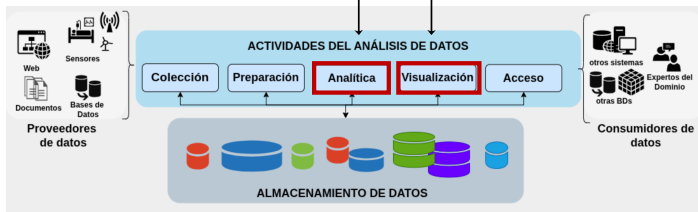
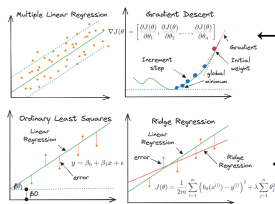


Figura 1.1: Proceso de Análisis de Datos

## “Definición”

...disciplina que se encarga de comprender y construir entidades artificiales inteligentes que simulan en algún sentido el comportamiento humano [6]

## “Definición”

...disciplina que se encarga de comprender y construir entidades artificiales inteligentes que simulan en algún sentido el comportamiento humano [6]

- *knowledge-based* > un programa cuya lógica codifica a gran número de propiedades del mundo y se concluye usando razonamiento lógico deductivo > **IA Simbólica**

## “Definición”

...disciplina que se encarga de comprender y construir entidades artificiales inteligentes que simulan en algún sentido el comportamiento humano [6]

- *knowledge-based* > un programa cuya lógica codifica a gran número de propiedades del mundo y se concluye usando razonamiento lógico deductivo > **IA Simbólica**
- *machine learning* > extraer información (patrones) directamente a partir de datos históricos y extrapolar para hacer predicciones > **IA No Simbólica**

## “Definición”

...disciplina que se encarga de comprender y construir entidades artificiales inteligentes que simulan en algún sentido el comportamiento humano [6]

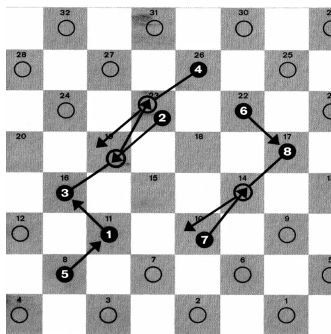
- *knowledge-based* > un programa cuya lógica codifica a gran número de propiedades del mundo y se concluye usando razonamiento lógico deductivo > **IA Simbólica**
- *machine learning* > extraer información (patrones) directamente a partir de datos históricos y extrapolar para hacer predicciones > **IA No Simbólica**
- Ambas > **IA Neuro Simbólica**



# Aprendizaje Automático - Game of Checkers (1956)

## Arthur Samuel [10]

The field of study that gives computers the ability to learn without being explicitly programmed.



Eight-move opening utilizing generalization learning. (See Appendix B, Game G-43.)



code<sup>1</sup>

<sup>1</sup><https://github.com/almostimplemented/checkers>

Tom Mitchell [8]

Machine Learning is the study of computer algorithms that improve automatically through experience.

Tom Mitchell [8]

Machine Learning is the study of computer algorithms that improve automatically through experience.

(also) Tom Mitchell [8]

A computer program is said to learn from experience  $E$  with respect to some task  $T$  and some performance measure  $P$ , if its performance on  $T$ , as measured by  $P$ , improves with experience  $E$ .

## Checkers game (Samuel + Mitchell)

- T: playing checkers
- P: percent of games won against opponent
- E: playing practice games against itself

## Checkers game (Samuel + Mitchell)

- T: playing checkers
- P: percent of games won against opponent
- E: playing practice games against itself

## Clasificador de spam

- T: filtrar correos entrantes
- P: porcentaje de spam removidos
- E: dataset of correos ya clasificados como spam

# Un Agente que aprende

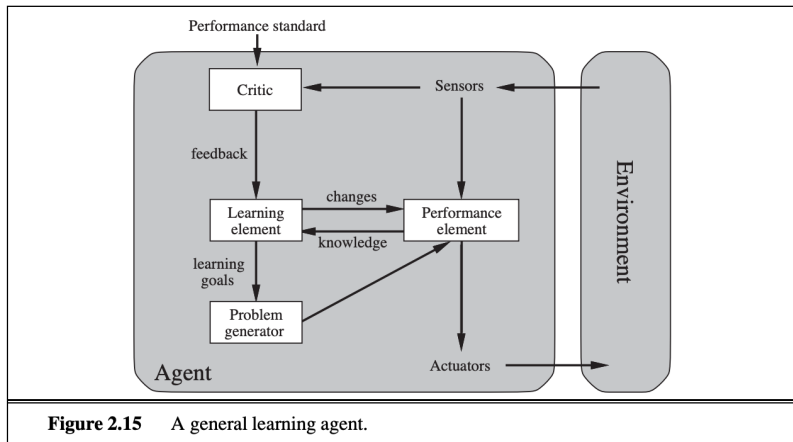
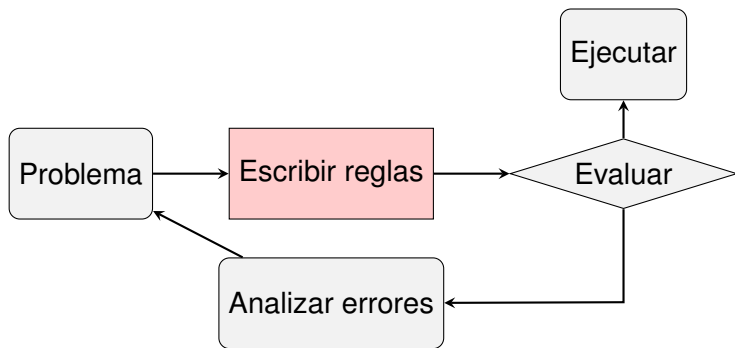


Figura 1.2: Extraída de [9]

# Programación tradicional



Créditos: Farhad Pourkamali Anaraki

## Programa

```
def es_spam(mensaje):
    mensaje = mensaje.lower()

    palabras_sospechosas = [
        "gratis", "gana dinero",
        "urgente", "haz clic",
        "oferta", "compra ahora"
    ]

    for palabra in palabras_sospechosas:
        if palabra in mensaje:
            return True

    if mensaje.count("!") > 3:
        return True

    if mensaje.startswith("!"):
        return True

    if "http://" in mensaje or "https://"
        " in mensaje:
        return True

    return False
```

## Ejemplo de uso

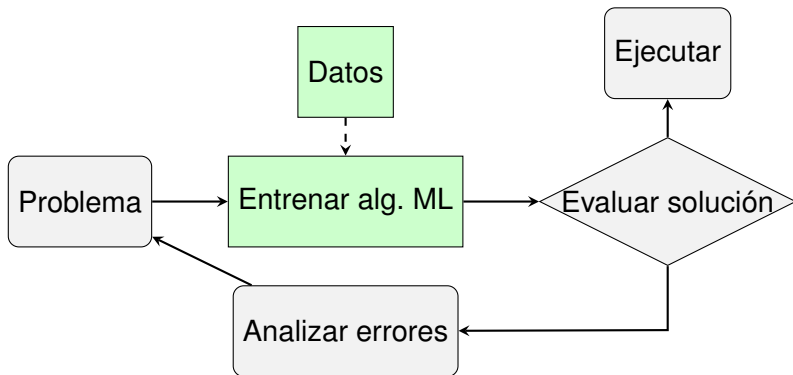
```
mensaje1 = "Gana dinero rapido. Haz clic
           aqui!"
mensaje2 = "Hola, te paso el informe
           adjunto."

if es_spam(mensaje1):
    print("Spam")
else:
    print("No spam")

if es_spam(mensaje2):
    print("Spam")
else:
    print("No spam")
```



Construir modelos predictivos a partir de datos, en vez de programarlos explícitamente



## Ejemplo: Aprendizaje Supervisado

- usamos datos etiquetados para aprender un modelo  $f$
- luego, usamos el modelo  $f$  para predecir  $y$  (datos no etiquetados)



## Análisis/Clasificación de imágenes

- **médicas** <https://entelai.com/en/>
- **identificación de personas/objetos** (vehículos autónomos)

# Aplicaciones (solo algunas!)

## Análisis/Clasificación de imágenes

- **médicas** <https://entel.ai.com/en/>
- **identificación de personas/objetos** (vehículos autónomos)

## Motores de recomendación

- **Amazon** > <https://aws.amazon.com/es/personalize/>
- **Netflix** >  
[https://en.wikipedia.org/wiki/Netflix\\_Prize](https://en.wikipedia.org/wiki/Netflix_Prize)  
(dataset: <https://www.kaggle.com/datasets/netflix-inc/netflix-prize-data/>)

# Aplicaciones (solo algunas!)

## Análisis/Clasificación de imágenes

- **médicas** <https://entel.ai.com/en/>
- **identificación de personas/objetos** (vehículos autónomos)

## Motores de recomendación

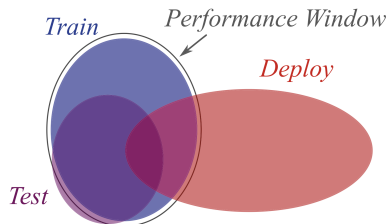
- **Amazon** > <https://aws.amazon.com/es/personalize/>
- **Netflix** >  
[https://en.wikipedia.org/wiki/Netflix\\_Prize](https://en.wikipedia.org/wiki/Netflix_Prize)  
(**dataset:** <https://www.kaggle.com/datasets/netflix-inc/netflix-prize-data/>)

## Procesamiento de language natural

- *flagging* comentarios en plataformas sociales
- reconocimiento de voz (Alex/Siri)
- traductores

## Workflow

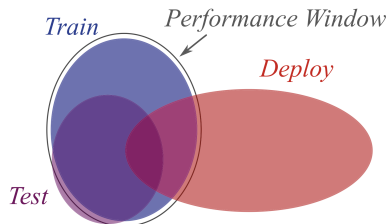
- Entender el dominio, conocimiento previo y metas.
- Pre-procesar datos (integrar, seleccionar, limpiar)
- Entrenar modelos
- Interpretar resultados
- Consolidar y desplegar conocimiento descubierto
- Ciclar sobre estos pasos anteriores



**Figura 1.3:** Un posible escenario de falla en la implementación de ML [5]

## Workflow

- Entender el dominio, conocimiento previo y metas.
- Pre-procesar datos (integrar, seleccionar, limpiar)
- Entrenar modelos
- Interpretar resultados
- Consolidar y desplegar conocimiento descubierto
- Ciclar sobre estos pasos anteriores



**Figura 1.3:** Un posible escenario de falla en la implementación de ML [5]

*El aprendizaje automático es un proceso de prueba y error.*

# Aprendiendo patrones a partir de datos



## Hipótesis del Aprendizaje Inductivo

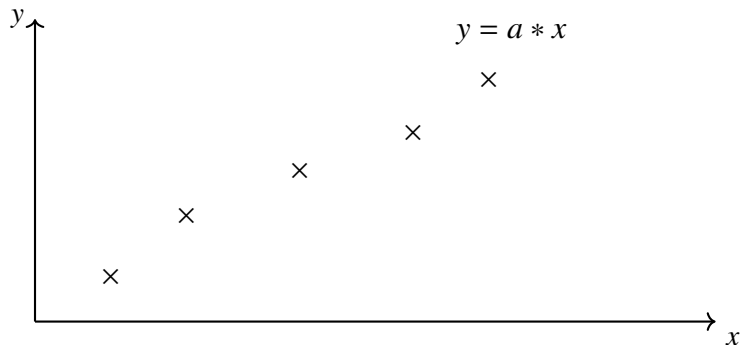
The inductive learning hypothesis. Any hypothesis found to approximate the target function well over a sufficiently large set of training examples will also approximate the target function well over other unobserved examples. [8]

## Hipótesis del Aprendizaje Inductivo

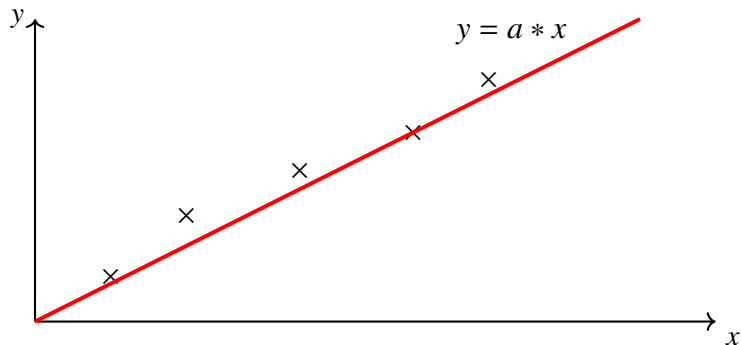
The inductive learning hypothesis. Any hypothesis found to approximate the target function well over a sufficiently large set of training examples will also approximate the target function well over other unobserved examples. [8]

- Given examples of a function  $(x, f(x))$ ,
- Predict function  $f(x)$  for new examples  $x$ 
  - Discrete  $f(x)$ : Classification
  - Continuous  $f(x)$ : Regression
  - $f(x) = \text{Probability}(x)$ : Probability estimation

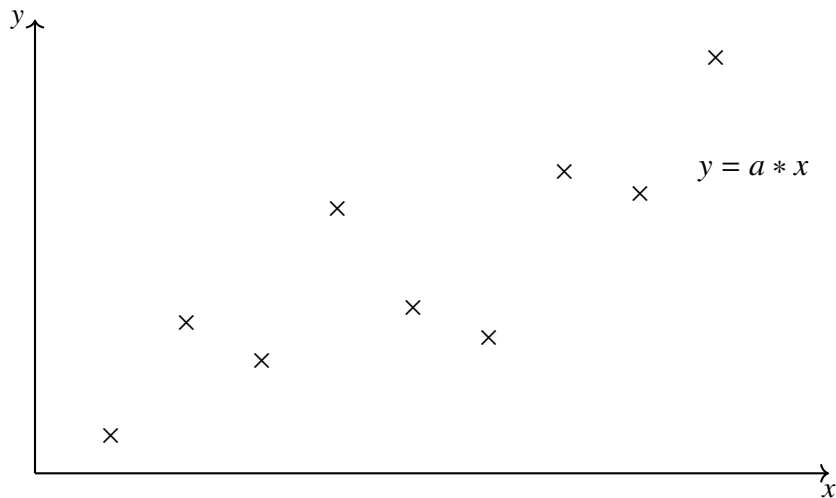
# Inferencia Empírica - Regresión



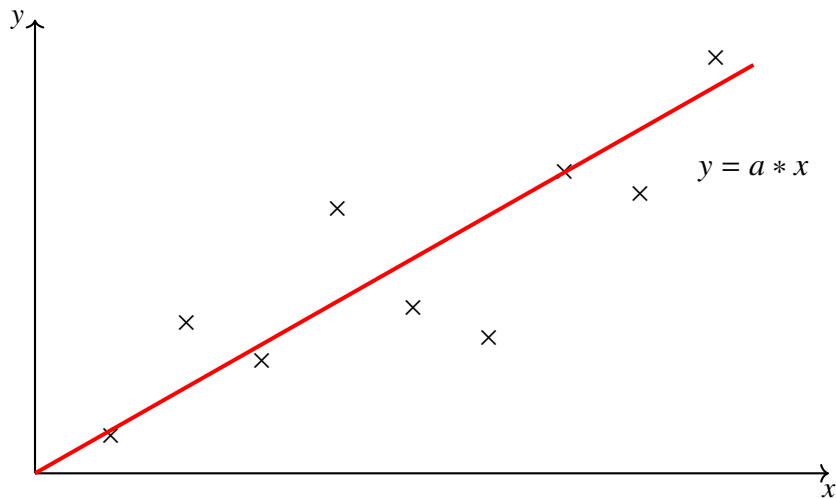
# Inferencia Empírica - Regresión



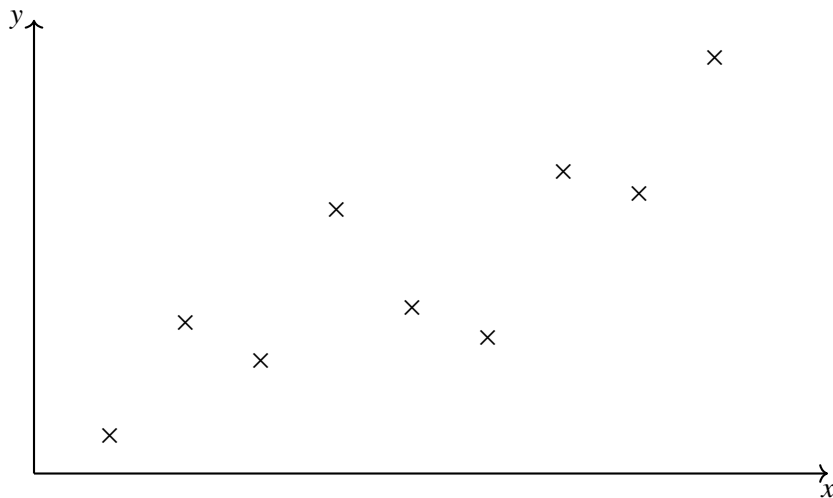
## Inferencia Empírica (cont'd)

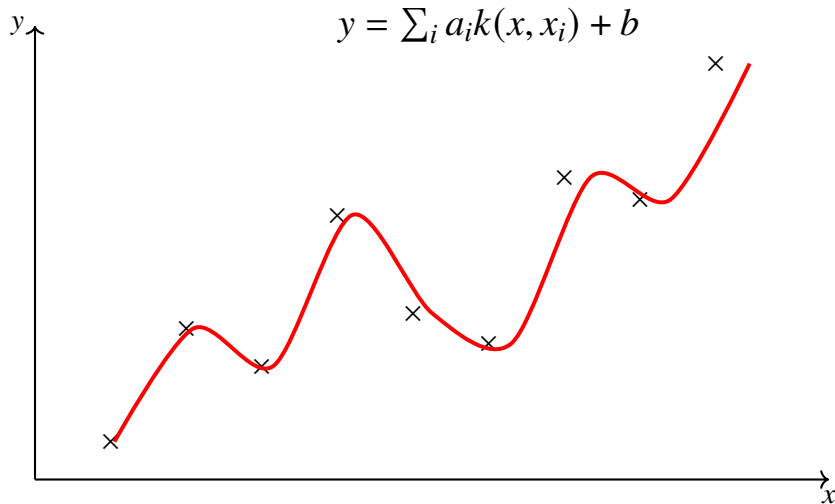


# Inferencia Empírica (cont'd)



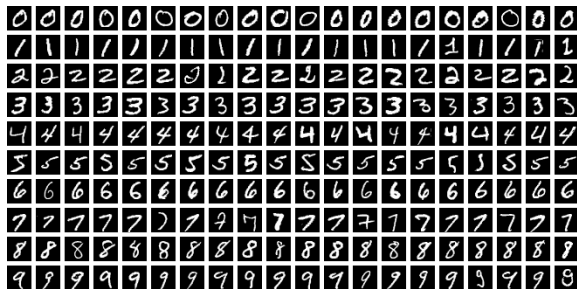
## Inferencia Empírica (cont'd)







# Inferencia Empírica - Clasificación



(\*) figura extraída de [https://en.wikipedia.org/wiki/MNIST\\_database](https://en.wikipedia.org/wiki/MNIST_database)

**From:** promo@regalos-gratis.com  
**To:** usuario@correo.com  
**Subject:** ¡Ganá un iPhone hoy!

¡Felicidades!  
Has sido seleccionado para  
ganar un **iPhone GRATIS**.  
Solo haz clic aquí ahora:  
[www.oferta-falsa.com](http://www.oferta-falsa.com)

¡No pierdas esta oportunidad  
exclusiva!

**From:** promo@regalos-gratis.com  
**To:** usuario@correo.com  
**Subject:** ¡Ganá un iPhone hoy!

¡Felicidades!  
Has sido seleccionado para  
ganar un **iPhone GRATIS**.  
Solo haz clic aquí ahora:  
[www.oferta-falsa.com](http://www.oferta-falsa.com)

¡No pierdas esta oportunidad  
exclusiva!

**SPAM!**

**From:** promo@regalos-gratis.com  
**To:** usuario@correo.com  
**Subject:** ¡Ganá un iPhone hoy!

¡Felicidades!  
Has sido seleccionado para  
ganar un **iPhone GRATIS**.  
Solo haz clic aquí ahora:  
[www.oferta-falsa.com](http://www.oferta-falsa.com)

¡No pierdas esta oportunidad  
exclusiva!

**SPAM! (1)**

**From:** promo@regalos-gratis.com  
**To:** usuario@correo.com  
**Subject:** ¡Ganá un iPhone hoy!

¡Felicidades!  
Has sido seleccionado para  
ganar un **iPhone GRATIS**.  
Solo haz clic aquí ahora:  
[www.oferta-falsa.com](http://www.oferta-falsa.com)

¡No pierdas esta oportunidad  
exclusiva!

**From:** laura@empresa.com  
**To:** juan@correo.com  
**Subject:** Reunión del jueves

Hola Germán,  
Te escribo para confirmar  
nuestra reunión del jueves a  
las 10:00.  
Adjunto encontrarás la agenda.

Saludos,  
Laura

**SPAM! (1)**

**From:** promo@regalos-gratis.com  
**To:** usuario@correo.com  
**Subject:** ¡Ganá un iPhone hoy!

¡Felicidades!  
Has sido seleccionado para  
ganar un **iPhone GRATIS**.  
Solo haz clic aquí ahora:  
[www.oferta-falsa.com](http://www.oferta-falsa.com)

¡No pierdas esta oportunidad  
exclusiva!

**SPAM! (1)**

**From:** laura@empresa.com  
**To:** juan@correo.com  
**Subject:** Reunión del jueves

Hola Germán,  
Te escribo para confirmar  
nuestra reunión del jueves a  
las 10:00.  
Adjunto encontrarás la agenda.

Saludos,  
Laura

**NO SPAM! (0)**

# Construimos el clasificador de spam

Definimos a  $x$  como una característica del correo e  $y$  como su categoría: spam (1) y no spam (0). Por simplicidad, las características  $x$  serán un subconjunto de  $n$  palabras indicativas de spam/no spam.

**From:** promo@regalos-gratis.com

**To:** usuario@correo.com

**Subject:** ¡Ganá un iPhone hoy!

¡Felicidades!

Has sido seleccionado para ganar un

**iPhone GRATIS.**

Solo haz clic aquí ahora:

[www.oferta-falsa.com](http://www.oferta-falsa.com)

¡No pierdas esta oportunidad exclusiva!

# Construimos el clasificador de spam

Definimos a  $x$  como una característica del correo e  $y$  como su categoría: spam (1) y no spam (0). Por simplicidad, las características  $x$  serán un subconjunto de  $n$  palabras indicativas de spam/no spam.

**From:** promo@regalos-gratis.com

**To:** usuario@correo.com

**Subject:** ¡Ganá un iPhone hoy!

$x = \{ \text{ganá, germán, ahora, saludos, reunión, gratis} \}$

¡Felicidades!

Has sido seleccionado para ganar un

**iPhone GRATIS.**

Solo haz clic aquí ahora:

[www.oferta-falsa.com](http://www.oferta-falsa.com)

¡No pierdas esta oportunidad exclusiva!



# Construimos el clasificador de spam

Definimos a  $x$  como una característica del correo e  $y$  como su categoría: spam (1) y no spam (0). Por simplicidad, las características  $x$  serán un subconjunto de  $n$  palabras indicativas de spam/no spam.

**From:** promo@regalos-gratis.com  
**To:** usuario@correo.com  
**Subject:** ¡Ganá un iPhone hoy!

¡Felicidades!  
Has sido seleccionado para ganar un  
**iPhone GRATIS.**  
Solo haz clic aquí ahora:  
www.oferta-falsa.com

¡No pierdas esta oportunidad exclusiva!

$x = \{ \text{ganá, germán, ahora, saludos, reunión, gratis} \}$

$$x_j = \begin{cases} 1 & \text{si la palabra está en el correo} \\ 0 & \text{en otro caso} \end{cases}$$

# Construimos el clasificador de spam

Definimos a  $x$  como una característica del correo y  $y$  como su categoría: spam (1) y no spam (0). Por simplicidad, las características  $x$  serán un subconjunto de  $n$  palabras indicativas de spam/no spam.

**From:** promo@regalos-gratis.com  
**To:** usuario@correo.com  
**Subject:** ¡Ganá un iPhone hoy!

¡Felicidades!  
Has sido seleccionado para ganar un  
**iPhone GRATIS.**  
Solo haz clic aquí ahora:  
www.oferta-falsa.com

¡No pierdas esta oportunidad exclusiva!

$x = \{ \text{ganá, germán, ahora, saludos, reunión, gratis} \}$

$$x_j = \begin{cases} 1 & \text{si la palabra está en el correo} \\ 0 & \text{en otro caso} \end{cases}$$

$$y \in \{0, 1\}$$

dónde:

0 es la "Clase Negativa" (No spam) y  
1 es la "Clase Positiva" (spam)

# Construimos el clasificador de spam

Definimos a  $x$  como una característica del correo y  $y$  como su categoría: spam (1) y no spam (0). Por simplicidad, las características  $x$  serán un subconjunto de  $n$  palabras indicativas de spam/no spam.

**From:** promo@regalos-gratis.com  
**To:** usuario@correo.com  
**Subject:** ¡Ganá un iPhone hoy!

¡Felicidades!  
Has sido seleccionado para ganar un  
**iPhone GRATIS.**  
Solo haz clic aquí ahora:  
www.oferta-falsa.com

¡No pierdas esta oportunidad exclusiva!

$x = \{ \text{ganá, germán, ahora, saludos, reunión, gratis} \}$

$$x_j = \begin{cases} 1 & \text{si la palabra está en el correo} \\ 0 & \text{en otro caso} \end{cases}$$

$$y \in \{0, 1\}$$

dónde:

0 es la "Clase Negativa" (No spam) y  
1 es la "Clase Positiva" (spam)

$X =$

ganá	germán	ahora	saludos	reunión	gratis
1	0	1	0	0	1

# Construimos el clasificador de spam

Definimos a  $x$  como una característca del correo e  $y$  como su categoria: spam (1) y no spam (0). Por simplicidad, las características  $x$  serán un subconjunto de  $n$  palabras indicativas de spam/no spam.

**From:** promo@regalos-gratis.com  
**To:** usuario@correo.com  
**Subject:** ¡Ganá un iPhone hoy!

¡Felicidades!  
Has sido seleccionado para ganar un  
**iPhone GRATIS.**  
Solo haz clic aquí ahora:  
www.oferta-falsa.com

¡No pierdas esta oportunidad exclusiva!

$x = \{ \text{ganá, germán, ahora, saludos, reunión, gratis} \}$

$$x_j = \begin{cases} 1 & \text{si la palabra está en el correo} \\ 0 & \text{en otro caso} \end{cases}$$

$$y \in \{0, 1\}$$

dónde:

0 es la “Clase Negativa” (No spam) y  
1 es la “Clase Positiva” (spam)

$X =$

ganá	germán	ahora	saludos	reunión	gratis
1	0	1	0	0	1

(\*) En la práctica, se deberían tomar las  $n$  palabras más frecuentes en el conjunto de entrenamiento

# Construimos el clasificador de spam (cont'd)

```
1 # Ejemplo de datos x_entr, y_entr, x_test, y_test
2 X = [
3     [1, 0, 1, 0, 0, 1],
4     [1, 1, 1, 1, 0, 1],
5     [0, 0, 1, 0, 0, 1]
6 ]
7
8 y = [1, 1, 0]
9
10 # Entrenar modelo (por ejemplo, SVM, etc.)
11 model = EntrenarModelo(x_entr, y_entr)
12 y_pred = Predecir(x_test)
13
14 # Medir la performance
15 cm = matriz_confusion(y_test, y_pred)
16
17 # Predecir si un nuevo mensaje es spam
18 nuevo_correo = [0, 1, 0, 1, 1, 0]
19 prediccion = model.predecir(nuevo_correo)
20
21 print("Spam" if prediccion == 1 else "No spam")
```

# Buenas Prácticas y Evaluación de Performance

¡Recordatorio!

*El aprendizaje automático es un proceso de prueba y error.*

## ¡Recordatorio!

*El aprendizaje automático es un proceso de prueba y error.*

- Entender el dominio, conocimiento previo y metas.
- Pre-procesar datos (integrar, seleccionar, limpiar)
- Entrenar modelos
- Interpretar resultados
- Consolidar y desplegar conocimiento descubierto
- Ciclar sobre estos pasos anteriores



## Problema

- Conjunto de posibles instances  $X$
- Función objetivo desconocida:  $f : X \rightarrow Y$
- Conjunto de posibles hipótesis:  $H = \{h \mid h : X \rightarrow Y\}$

## Input

- Ejemplos de entrenamiento  $\langle x_i, y_i \rangle$ . Para el ejemplo de clasificación de spam,  $x$  es un email e  $y$  es “spam” o “no spam”

## Output

- Hipótesis  $h \in H$  que mejor aproxima la función objetivo  $f$ .

## Problema

- 1 ¿Contamos con un dataset “apropiado”?
- 2 ¿El dataset es representativo de los datos de producción?  
(distribución) > **suposición fundamental del AA**
- 3 ¿Cómo elegimos un espacio de hipótesis?

# “Issues” en el Aprendizaje Automático

## Problema

- 1 ¿Contamos con un dataset “apropiado”?
- 2 ¿El dataset es representativo de los datos de producción?  
(distribución) > **suposición fundamental del AA**
- 3 ¿Cómo elegimos un espacio de hipótesis?

## Input

- 1 ¿Como elegimos el conjunto de entrenamiento?
- 2 ¿Deberíamos entrenar sobre el dataset completo?

# “Issues” en el Aprendizaje Automático

## Problema

- 1 ¿Contamos con un dataset “apropiado”?
- 2 ¿El dataset es representativo de los datos de producción?  
(distribución) > **suposición fundamental del AA**
- 3 ¿Cómo elegimos un espacio de hipótesis?

## Input

- 1 ¿Como elegimos el conjunto de entrenamiento?
- 2 ¿Deberíamos entrenar sobre el dataset completo?

## Output

- 1 ¿Es  $h$  la que “mejor” aproxima la función objetivo  $f$ ?
- 2 ¿Cómo evaluamos el modelo aprendido?

# “Issues” en el Aprendizaje Automático - Problema

- 1 ¿Contamos con un dataset “apropiado”?
- 2 ¿El dataset es representativo de los datos de producción?  
(distribución) > **suposición fundamental del AA**
- 3 ¿Cómo elegimos un espacio de hipótesis?

# “Issues” en el Aprendizaje Automático - Problema

- 1 ¿Contamos con un dataset “apropiado”?
  - 2 ¿El dataset es representativo de los datos de producción?  
(distribución) > **suposición fundamental del AA**
  - 3 ¿Cómo elegimos un espacio de hipótesis?
- En toda aplicación de AA, es esperable poder diferenciar los siguientes tres datasets:

# “Issues” en el Aprendizaje Automático - Problema

- 1 ¿Contamos con un dataset “apropiado”?
- 2 ¿El dataset es representativo de los datos de producción? (distribución) > **suposición fundamental del AA**
- 3 ¿Cómo elegimos un espacio de hipótesis?

- En toda aplicación de AA, es esperable poder diferenciar los siguientes tres datasets:

- 1 **Conjunto de entrenamiento (*training set*)**, usado para estimar los parámetros del modelo

# “Issues” en el Aprendizaje Automático - Problema

- 1 ¿Contamos con un dataset “apropiado”?
- 2 ¿El dataset es representativo de los datos de producción? (distribución) > **suposición fundamental del AA**
- 3 ¿Cómo elegimos un espacio de hipótesis?

- En toda aplicación de AA, es esperable poder diferenciar los siguientes tres datasets:

- 1 **Conjunto de entrenamiento (*training set*)**, usado para estimar los parámetros del modelo
- 2 **Conjunto de desarrollo o validación (*development set*)**, usado para el diseño (hiperparámetros, arquitectura, características)



# “Issues” en el Aprendizaje Automático - Problema

- 1 ¿Contamos con un dataset “apropiado”?
- 2 ¿El dataset es representativo de los datos de producción? (distribución) > **suposición fundamental del AA**
- 3 ¿Cómo elegimos un espacio de hipótesis?

- En toda aplicación de AA, es esperable poder diferenciar los siguientes tres datasets:

- 1 **Conjunto de entrenamiento (*training set*)**, usado para estimar los parámetros del modelo
- 2 **Conjunto de desarrollo o validación (*development set*)**, usado para el diseño (hiperparámetros, arquitectura, características)
- 3 **Conjunto de evaluación (*evaluation set*)**, usado para reportar la performance final del model aprendido.

# “Issues” en el AA - Problema (cont'd)

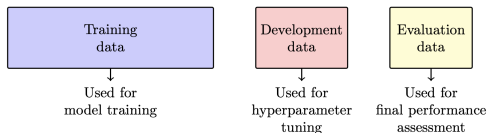


Figura 3.1: Datasets en AA [7]

- (Idealmente) desarrollo y evaluación deberían ser consistentes en términos de su distribución, reflejando los datos de producción.

# “Issues” en el AA - Problema (cont'd)

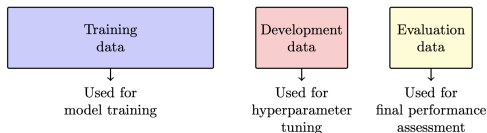


Figura 3.1: Datasets en AA [7]

- (Idealmente) desarrollo y evaluación deberían ser consistentes en términos de su distribución, reflejando los datos de producción.
- El objetivo del conjunto de desarrollo es detectar cambios en la performance del modelo, i.e. minimizar el error de la generalización. Por ej,

*modelo = SVC(kernel = 'rbf', C = C, gamma = gamma)*

# “Issues” en el AA - Problema (cont'd)

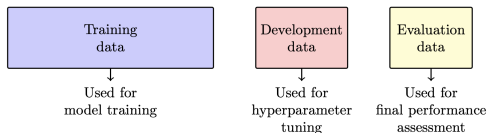


Figura 3.1: Datasets en AA [7]

- (Idealmente) desarrollo y evaluación deberían ser consistentes en términos de su distribución, reflejando los datos de producción.
- El objetivo del conjunto de desarrollo es detectar cambios en la performance del modelo, i.e. minimizar el error de la generalización. Por ej,

*modelo = SVC(kernel = 'rbf', C = C, gamma = gamma)*

- El objetivo del conjunto de evaluación es estimar la performance en el mundo real

# “Issues” en el AA - Problema (cont’d)

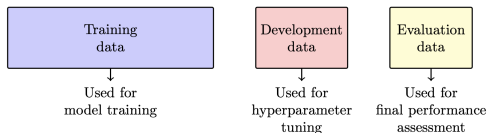


Figura 3.1: Datasets en AA [7]

- (Idealmente) desarrollo y evaluación deberían ser consistentes en términos de su distribución, reflejando los datos de producción.
- El objetivo del conjunto de desarrollo es detectar cambios en la performance del modelo, i.e. minimizar el error de la generalización. Por ej,

*modelo = SVC(kernel = 'rbf', C = C, gamma = gamma)*

- El objetivo del conjunto de evaluación es estimar la performance en el mundo real
- En general, el 30 % del dataset es destinado a desarrollo+evaluación - **para dataset pequeños!**

## “Issues” en el AA - Problema (cont'd)

¿Qué sucede cuando el dataset no es lo suficientemente grande como para dividirlo en conjuntos desarrollo/entrenamiento/evaluación?

# “Issues” en el AA - Problema (cont’d)

¿Qué sucede cuando el dataset no es lo suficientemente grande como para dividirlo en conjuntos desarrollo/entrenamiento/evaluación?

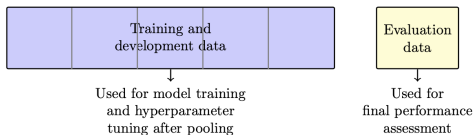


Figura 3.2: Cross-validation [7]

# “Issues” en el AA - Problema (cont’d)

¿Qué sucede cuando el dataset no es lo suficientemente grande como para dividirlo en conjuntos desarrollo/entrenamiento/evaluación?

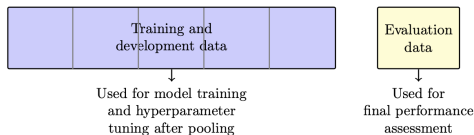


Figura 3.2: Cross-validation [7]

- Para cada partición, las restantes son usadas para entrenar el modelo



# “Issues” en el AA - Problema (cont’d)

¿Qué sucede cuando el dataset no es lo suficientemente grande como para dividirlo en conjuntos desarrollo/entrenamiento/evaluación?

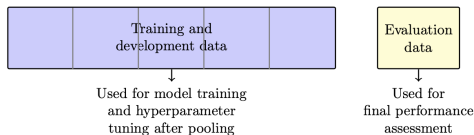


Figura 3.2: Cross-validation [7]

- Para cada partición, las restantes son usadas para entrenar el modelo
- Se evalúa el modelo usando la partición restante

# “Issues” en el AA - Problema (cont’d)

¿Qué sucede cuando el dataset no es lo suficientemente grande como para dividirlo en conjuntos desarrollo/entrenamiento/evaluación?

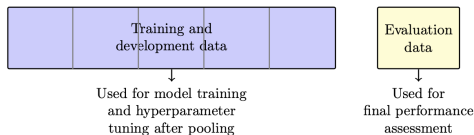


Figura 3.2: Cross-validation [7]

- Para cada partición, las restantes son usadas para entrenar el modelo
- Se evalúa el modelo usando la partición restante
- Se repite tantas veces como particiones haya en el dataset

# “Issues” en el AA - Problema (cont’d)

¿Qué sucede cuando el dataset no es lo suficientemente grande como para dividirlo en conjuntos desarrollo/entrenamiento/evaluación?

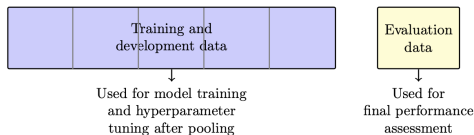


Figura 3.2: Cross-validation [7]

- Para cada partición, las restantes son usadas para entrenar el modelo
- Se evalúa el modelo usando la partición restante
- Se repite tantas veces como particiones haya en el dataset
- Si el tamaño dataset lo permite, un conjunto de evaluación debería ser usado luego de la *cross-validation* para una evaluación final y así tomar decisiones de diseño

# “Issues” en el AA - Reglas de Oro!

El conjunto de evaluación nunca debe usarse para revisar el modelo o hacer “tuning” de los parámetros

# “Issues” en el AA - Reglas de Oro!

El conjunto de evaluación nunca debe usarse para revisar el modelo o hacer “tuning” de los parámetros

Los datos usados para entrenamiento o desarrollo no deben usarse para evaluar la performance final del modelo. Con dichos datos, los resultados serán “optimistas”

- 1 ¿Como elegimos el conjunto de entrenamiento?

## 1 ¿Como elegimos el conjunto de entrenamiento?

- El conjunto de entrenamiento podría tener una distribución diferente a la de los conjuntos restantes.

## 1 ¿Como elegimos el conjunto de entrenamiento?

- El conjunto de entrenamiento podría tener una distribución diferente a la de los conjuntos restantes.
  - ej si estamos clasificando imagenes de perros, entonces el conjunto de entrenamiento podría incluir otros tipos de animales
- Si un modelo tiene una muy buena performance sobre el conjunto de entrenamiento pero no así sobre el de evaluación, podría haber *overfitting*



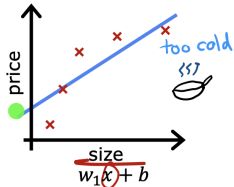
## 1 ¿Como elegimos el conjunto de entrenamiento?

- El conjunto de entrenamiento podría tener una distribución diferente a la de los conjuntos restantes.
  - ej si estamos clasificando imagenes de perros, entonces el conjunto de entrenamiento podría incluir otros tipos de animales
- Si un modelo tiene una muy buena performance sobre el conjunto de entrenamiento pero no así sobre el de evaluación, podría haber *overfitting*
  - ej nuestro clasificador de spam lo hace perfectamente sobre emails en entrenamiento, pero no sobre emails nuevos (sin clasificar)

## 1 ¿Como elegimos el conjunto de entrenamiento?

- El conjunto de entrenamiento podría tener una distribución diferente a la de los conjuntos restantes.
  - ej si estamos clasificando imagenes de perros, entonces el conjunto de entrenamiento podría incluir otros tipos de animales
- Si un modelo tiene una muy buena performance sobre el conjunto de entrenamiento pero no así sobre el de evaluación, podría haber *overfitting*
  - ej nuestro clasificador de spam lo hace perfectamente sobre emails en entrenamiento, pero no sobre emails nuevos (sin clasificar)
- Alternativas para prevenir *overfitting* (algunas): *cross-validation*, simplificación del modelo, ampliación del conjunto de entrenamiento!

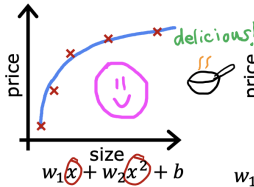
# “Issues” en el AA - Overfitting



underfit

- Does not fit the training set well

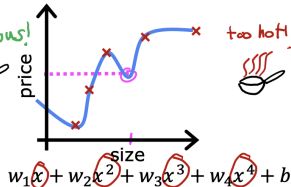
high bias



just right

- Fits training set pretty well

generalization

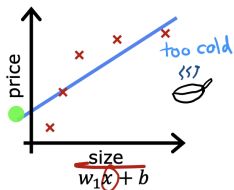


overfit

- Fits the training set extremely well

high variance

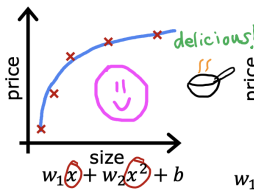
# “Issues” en el AA - Overfitting



underfit

- Does not fit the training set well

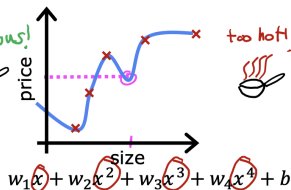
high bias



just right

- Fits training set pretty well

generalization



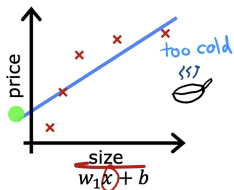
overfit

- Fits the training set extremely well

high variance

- **Sesgo (bias):** modelo “*demasiado simple*” para capturar la complejidad de los datos.

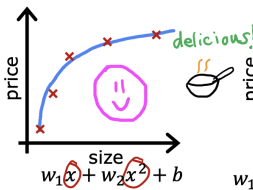
# “Issues” en el AA - Overfitting



underfit

- Does not fit the training set well

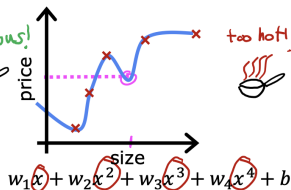
high bias



just right

- Fits training set pretty well

generalization



overfit

- Fits the training set extremely well

high variance

- **Sesgo (bias):** modelo “*demasiado simple*” para capturar la complejidad de los datos.
- **Varianza (Variance):** Aprende “*demasiado bien*” los datos de entrenamiento, incluido el ruido.

## “Issues” en el AA - Overfitting (cont’d)

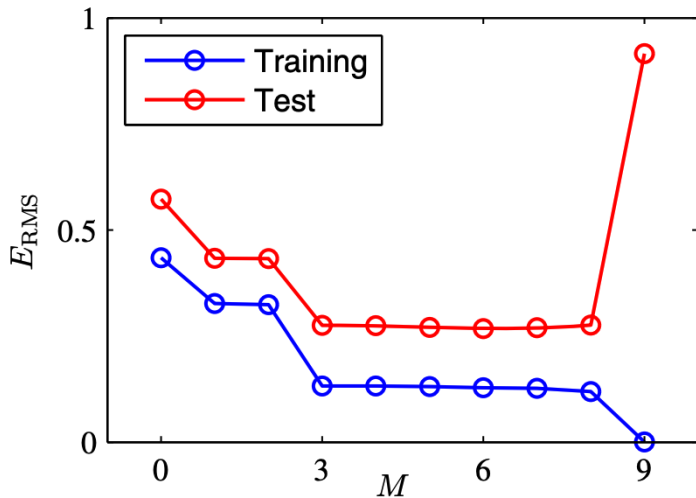


Figura 3.3: Error en training vs test (Bishop [3])

## Output

- 1 ¿Es mi hipótesis  $h$  la que “mejor” aproxima la función objetivo  $f$ ?
  - 2 ¿Cómo podemos evaluar la precisión de una hipótesis sobre datos no etiquetados?
- El algoritmo de aprendizaje toma el “mejor elemento” del espacio de hipótesis, i.e. la función que mejor aproxima el conjunto de entrenamiento.

## Output

- 1 ¿Es mi hipótesis  $h$  la que “mejor” aproxima la función objetivo  $f$ ?
  - 2 ¿Cómo podemos evaluar la precisión de una hipótesis sobre datos no etiquetados?
- El algoritmo de aprendizaje toma el “mejor elemento” del espacio de hipótesis, i.e. la función que mejor aproxima el conjunto de entrenamiento.
  - Los modelos son completamente definidos por sus parámetros, por lo tanto, encontrar un modelo óptimo es similar a encontrar un conjunto óptimo de parámetros.



## Output

- 1 ¿Es mi hipótesis  $h$  la que “mejor” aproxima la función objetivo  $f$ ?
  - 2 ¿Cómo podemos evaluar la precisión de una hipótesis sobre datos no etiquetados?
- El algoritmo de aprendizaje toma el “mejor elemento” del espacio de hipótesis, i.e. la función que mejor aproxima el conjunto de entrenamiento.
  - Los modelos son completamente definidos por sus parámetros, por lo tanto, encontrar un modelo óptimo es similar a encontrar un conjunto óptimo de parámetros.
    - ej para una función lineal,  $y = \theta_0 + \theta_1 x$ , los parámetros  $\theta_0$  y  $\theta_1$  determinan la relación entre  $y$  y  $x$ .

- 1 ¿Es mi hipótesis  $h$  la que “mejor” aproxima la función objetivo  $f$ ?
  - 2 ¿Cómo podemos evaluar la precisión de una hipótesis sobre datos no etiquetados?
- Una opción para guiar la selección del espacio de hipótesis es usar *prior knowledge*, i.e. conocimiento previo o asunciones pre-existentes.

- 1 ¿Es mi hipótesis  $h$  la que “mejor” aproxima la función objetivo  $f$ ?
  - 2 ¿Cómo podemos evaluar la precisión de una hipótesis sobre datos no etiquetados?
- Una opción para guiar la selección del espacio de hipótesis es usar *prior knowledge*, i.e. conocimiento previo o asunciones pre-existentes.
    - ej Emails con “GANASTE EL PREMIO!” en el asunto son probablemente spams

- 1 ¿Es mi hipótesis  $h$  la que “mejor” aproxima la función objetivo  $f$ ?
  - 2 ¿Cómo podemos evaluar la precisión de una hipótesis sobre datos no etiquetados?
- Una opción para guiar la selección del espacio de hipótesis es usar *prior knowledge*, i.e. conocimiento previo o asunciones pre-existentes.
    - ej Emails con “GANASTE EL PREMIO!” en el asunto son probablemente spams
  - Es recomendable usar la hipótesis “más simple” consistente con los datos > ayuda a evitar *overfitting*!

- La elección de métricas es esencial en la evaluación de los modelos de AA

- La elección de métricas es esencial en la evaluación de los modelos de AA
- Las métricas debería depender de la tarea que se está tratando de resolver

- La elección de métricas es esencial en la evaluación de los modelos de AA
- Las métricas debería depender de la tarea que se está tratando de resolver
  - ej La precisión (*accuracy*) puede no ser útil si existen diferentes costos de una clasificación errónea. En un dominio médico, un falso negativo (no se “predice” enfermedad pero la hay!) es más costoso que un falso positivo.

## Matriz de confusión

- Usaremos una matriz de confusión para evaluar la performance de un clasificador



## Matriz de confusión

- Usaremos una matriz de confusión para evaluar la performance de un clasificador
- Suponemos que nuestro clasificador trata de predecir si un email es spam o no. Entonces, dada una entrada  $x$ :

## Matriz de confusión

- Usaremos una matriz de confusión para evaluar la performance de un clasificador
- Suponemos que nuestro clasificador trata de predecir si un email es spam o no. Entonces, dada una entrada  $x$ :
  - **True positive (TP):**  $x$  puede ser un spam y nuestro modelo lo predice como tal

## Matriz de confusión

- Usaremos una matriz de confusión para evaluar la performance de un clasificador
- Suponemos que nuestro clasificador trata de predecir si un email es spam o no. Entonces, dada una entrada  $x$ :
  - **True positive (TP):**  $x$  puede ser un spam y nuestro modelo lo predice como tal
  - **False negative (FN):**  $x$  es un spam y nuestro modelo lo predice como no spam

## Matriz de confusión

- Usaremos una matriz de confusión para evaluar la performance de un clasificador
- Suponemos que nuestro clasificador trata de predecir si un email es spam o no. Entonces, dada una entrada  $x$ :
  - **True positive (TP)**:  $x$  puede ser un spam y nuestro modelo lo predice como tal
  - **False negative (FN)**:  $x$  es un spam y nuestro modelo lo predice como no spam
  - **False positive (FP)**:  $x$  no es un spam y nuestro modelo lo predice como spam

## Matriz de confusión

- Usaremos una matriz de confusión para evaluar la performance de un clasificador
- Suponemos que nuestro clasificador trata de predecir si un email es spam o no. Entonces, dada una entrada  $x$ :
  - **True positive (TP)**:  $x$  puede ser un spam y nuestro modelo lo predice como tal
  - **False negative (FN)**:  $x$  es un spam y nuestro modelo lo predice como no spam
  - **False positive (FP)**:  $x$  no es un spam y nuestro modelo lo predice como spam
  - **True negative (TN)**:  $x$  no es un spam y nuestro modelo lo predice como tal

# Matriz de confusión / Precision / Recall

Predicción	Actual	
	1	0
1	TP	FP
0	FN	TN

## ■ Precisión

(sobre todos los emails donde  $y = 1$  (spam), cuantos son efectivamente spam)

Predicción	Actual	
	1	0
1	TP	FP
0	FN	TN

$$\#TP / \#Positivos_{pred}$$

$$= \#TP / (\#TP + \#FP)$$

predicción	Actual	
	1	0
1	85	10
0	15	90

Predicción	Actual	
	1	0
1	TP	FP
0	FN	TN

predicción	Actual	
	1	0
1	85	10
0	15	90

## ■ Precisión

(sobre todos los emails donde  $y = 1$  (spam), cuantos son efectivamente spam)

$$\#TP / \#Positivos_{pred}$$

$$= \#TP / (\#TP + \#FP)$$

## ■ Recall

(sobre todos los emails que son spam, que fracción fue detectada como spam)

$$\#TP / \#Positivos_{actual}$$

$$= \#TP / (\#TP + \#FN)$$



# Matriz de confusión / Precision / Recall (cont'd)

## ■ Matriz

Predicción	Actual	
	1	0
1	TP	FP
0	FN	TN

## ■ Precisión

(sobre todos los emails donde  $y = 1$  (spam), cuantos son efectivamente spam)

$$\begin{aligned} & \#TP / \#Positivos_{pred} \\ &= \#TP / (\#TP + \#FP) \end{aligned}$$

## ■ Recall

(sobre todos los emails que son spam, que fracción fue detectada como spam)

$$\begin{aligned} & \#TP / \#Positivos_{actual} \\ &= \#TP / (\#TP + \#FN) \end{aligned}$$

# Matriz de confusión / Precision / Recall (cont'd)

## ■ Matriz

Predicción	Actual	
	1	0
1	TP	FP
0	FN	TN

- Asumimos que  $0 \leq h(x) \leq 1$  y definimos un umbral (*threshold*) para la predicción.

## ■ Precisión

(sobre todos los emails donde  $y = 1$  (spam), cuantos son efectivamente spam)

$$\begin{aligned} & \#TP / \#Positivos_{pred} \\ &= \#TP / (\#TP + \#FP) \end{aligned}$$

## ■ Recall

(sobre todos los emails que son spam, que fracción fue detectada como spam)

$$\begin{aligned} & \#TP / \#Positivos_{actual} \\ &= \#TP / (\#TP + \#FN) \end{aligned}$$

# Matriz de confusión / Precision / Recall (cont'd)

## ■ Matriz

Predicción	Actual	
	1	0
1	TP	FP
0	FN	TN

## ■ Precisión

(sobre todos los emails donde  $y = 1$  (spam), cuantos son efectivamente spam)

$$\begin{aligned} & \#TP / \#Positivos_{pred} \\ &= \#TP / (\#TP + \#FP) \end{aligned}$$

## ■ Recall

(sobre todos los emails que son spam, que fracción fue detectada como spam)

$$\begin{aligned} & \#TP / \#Positivos_{actual} \\ &= \#TP / (\#TP + \#FN) \end{aligned}$$

- Asumimos que  $0 \leq h(x) \leq 1$  y definimos un umbral (*threshold*) para la predicción.

**umbral** = 0,5

- Predecimos  $\hat{y} = 1$  si  $h(x) \geq 0,5$
- Predecimos  $\hat{y} = 0$  si  $h(x) < 0,5$

# Matriz de confusión / Precision / Recall (cont'd)

## ■ Matriz

Predicción	Actual	
	1	0
1	TP	FP
0	FN	TN

## ■ Precisión

(sobre todos los emails donde  $y = 1$  (spam), cuantos son efectivamente spam)

$$\begin{aligned} & \#TP / \#Positivos_{pred} \\ &= \#TP / (\#TP + \#FP) \end{aligned}$$

## ■ Recall

(sobre todos los emails que son spam, que fracción fue detectada como spam)

$$\begin{aligned} & \#TP / \#Positivos_{actual} \\ &= \#TP / (\#TP + \#FN) \end{aligned}$$

- Asumimos que  $0 \leq h(x) \leq 1$  y definimos un umbral (*threshold*) para la predicción.

**umbral = 0,5**

- Predecimos  $\hat{y} = 1$  si  $h(x) \geq 0,5$
- Predecimos  $\hat{y} = 0$  si  $h(x) < 0,5$

**umbral = 0,9**

- $\hat{y} = 1$  si  $h(x) \geq 0,9$
- $\hat{y} = 0$  si  $h(x) < 0,9$

⇒ alta precisión, bajo recall  
(\*) mayormente TP y menor número de FP

# Matriz de confusión / Precision / Recall (cont'd)

## ■ Matriz

Predicción	Actual	
	1	0
1	TP	FP
0	FN	TN

## ■ Precisión

(sobre todos los emails donde  $y = 1$  (spam), cuantos son efectivamente spam)

$$\#TP / \#Positivos_{pred}$$

$$= \#TP / (\#TP + \#FP)$$

## ■ Recall

(sobre todos los emails que son spam, que fracción fue detectada como spam)

$$\#TP / \#Positivos_{actual}$$

$$= \#TP / (\#TP + \#FN)$$

- Asumimos que  $0 \leq h(x) \leq 1$  y definimos un umbral (*threshold*) para la predicción.

**umbral = 0,5**

- Predecimos  $\hat{y} = 1$  si  $h(x) \geq 0,5$
- Predecimos  $\hat{y} = 0$  si  $h(x) < 0,5$

**umbral = 0,9**

- $\hat{y} = 1$  si  $h(x) \geq 0,9$
- $\hat{y} = 0$  si  $h(x) < 0,9$

⇒ alta precisión, bajo recall  
(\*) mayormente TP y menor número de FP

**umbral = 0,3**

- $\hat{y} = 1$  si  $h(x) \geq 0,3$
- $\hat{y} = 0$  si  $h(x) < 0,3$

⇒ baja precisión, alto recall  
(\*) menor número TP y mayormente FP

Si tenemos diferentes algoritmos para el mismo problema con diferentes valores de Precisión y Recall, ¿cómo podemos compararlos para elegir la mejor solución?

Algoritmo	Precisión (P)	Recall (R)	F-score
Algoritmo 1	0.5	0.4	0.444
Algoritmo 2	0.7	0.1	0.175
Algoritmo 3	0.02	1.0 (*)	0.0392

Si tenemos diferentes algoritmos para el mismo problema con diferentes valores de Precisión y Recall, ¿cómo podemos compararlos para elegir la mejor solución?

Algoritmo	Precisión (P)	Recall (R)	F-score
Algoritmo 1	0.5	0.4	0.444
Algoritmo 2	0.7	0.1	0.175
Algoritmo 3	0.02	1.0 (*)	0.0392

$$2 \frac{PR}{P+R}$$

■ Si  $P = 0$  o  $R = 0 \implies \text{F-score} = 0$

Si tenemos diferentes algoritmos para el mismo problema con diferentes valores de Precisión y Recall, ¿cómo podemos compararlos para elegir la mejor solución?

Algoritmo	Precisión (P)	Recall (R)	F-score
Algoritmo 1	0.5	0.4	0.444
Algoritmo 2	0.7	0.1	0.175
Algoritmo 3	0.02	1.0 (*)	0.0392

$$2 \frac{PR}{P+R}$$

■ Si  $P = 0$  o  $R = 0 \implies$  F-score = 0

■ Si  $P = 1$  o  $R = 1 \implies$  F-score = 1



# Matriz de confusión / Precision / Recall / F-score

Si tenemos diferentes algoritmos para el mismo problema con diferentes valores de Precisión y Recall, ¿cómo podemos compararlos para elegir la mejor solución?

Algoritmo	Precisión (P)	Recall (R)	F-score
Algoritmo 1	0.5	0.4	0.444
Algoritmo 2	0.7	0.1	0.175
Algoritmo 3	0.02	1.0 (*)	0.0392

$$2 \frac{PR}{P+R}$$

■ Si  $P = 0$  o  $R = 0 \implies$  F-score = 0

■ Si  $P = 1$  o  $R = 1 \implies$  F-score = 1

- Métricas apropiadas para casos donde hay una distribución *skewed* en un dataset, i.e. una clase aparece más que otra (clases desbalanceadas) – 95 % no spam y 5 % spam.

Si tenemos diferentes algoritmos para el mismo problema con diferentes valores de Precisión y Recall, ¿cómo podemos compararlos para elegir la mejor solución?

Algoritmo	Precisión (P)	Recall (R)	F-score
Algoritmo 1	0.5	0.4	0.444
Algoritmo 2	0.7	0.1	0.175
Algoritmo 3	0.02	1.0 (*)	0.0392

$$2 \frac{PR}{P+R}$$

■ Si  $P = 0$  o  $R = 0 \implies \text{F-score} = 0$

■ Si  $P = 1$  o  $R = 1 \implies \text{F-score} = 1$

- Métricas apropiadas para casos donde hay una distribución *skewed* en un dataset, i.e. una clase aparece más que otra (clases desbalanceadas) – 95 % no spam y 5 % spam.
- F-score *penaliza* los valores desbalanceados

En casos en los cuales valores negativos son también importantes (por ej, ausencia de una enfermedad), la alternativa a la precision es **especificidad**

Predicción	Actual	
	1	0
1	TP	FP
0	FN	TN

En casos en los cuales valores negativos son también importantes (por ej, ausencia de una enfermedad), la alternativa a la precision es **especificidad**

Predicción	Actual	
	1	0
1	TP	FP
0	FN	TN

## ■ Especificidad

(sobre todos los emails donde  $y = 0$  (no spam), cuantos son efectivamente no spam)

$$\begin{aligned} & \#TN / \#Negativos_{actual} \\ &= \#TN / (\#TN + \#FP) \end{aligned}$$

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

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

- Sin embargo, esta métrica puede NO ser útil en caso de **clases desbalanceadas**

# Matriz de confusión / Exactitud (*accuracy*)

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

- Sin embargo, esta métrica puede NO ser útil en caso de **clases desbalanceadas**

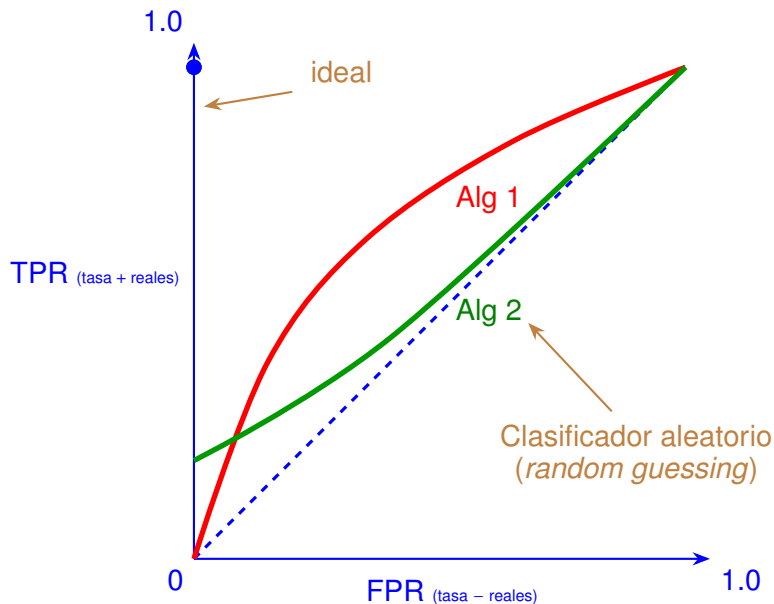
Por ejemplo, para un modelo de detección de enfermedades extrañas se obtiene la siguiente matriz de confusión:

Predicción	Actual	
	1	0
1	0	0
0	10	990

$$Recall = \frac{TP}{TP + FN} = \frac{0}{0 + 10} = 0$$

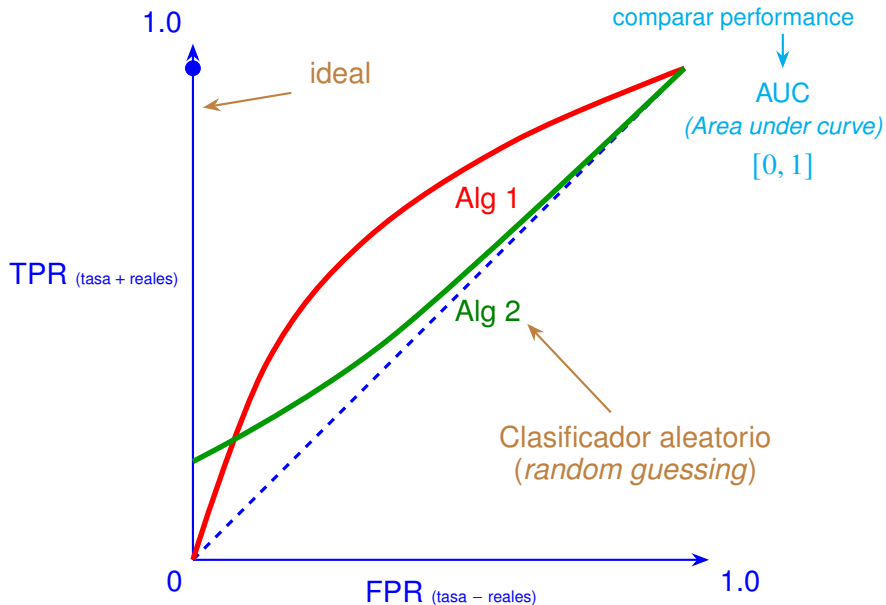
$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{0 + 990}{1000} = 99\%$$

# Métricas - Curvas ROC





# Métricas - Curvas ROC



# Aspectos Éticos

*La ética y la ciencia no son disciplinas separadas, sino interconectadas. Es esencial que los desarrollos científicos y técnicos estuvieran siempre alineados con valores éticos que promuevan el bienestar humano. (Mario Bunge [4])*

La IA necesita datos  $\Rightarrow$  *privacidad / governance*

# Ética en IA (y ML) [2]

La IA necesita datos  $\Rightarrow$  *privacidad / governance*

La IA es, con frecuencia, una caja negra  $\Rightarrow$  *explicabilidad / transparencia*

# Ética en IA (y ML) [2]

La IA necesita datos  $\Rightarrow$  *privacidad / governance*

La IA es, con frecuencia, una caja negra  $\Rightarrow$  *explicabilidad / transparencia*

La IA puede recomendar o tomar decisiones  $\Rightarrow$  *fairness* (equidad)

# Ética en IA (y ML) [2]

La IA necesita datos  $\Rightarrow$  *privacidad / governance*

La IA es, con frecuencia, una caja negra  $\Rightarrow$  *explicabilidad / transparencia*

La IA puede recomendar o tomar decisiones  $\Rightarrow$  *fairness* (equidad)

La IA tiene errores  $\Rightarrow$  *accountability* (responsabilidad)

# Ética en IA (y ML) [2]

La IA necesita datos  $\Rightarrow$  *privacidad / governance*

La IA es, con frecuencia, una caja negra  $\Rightarrow$  *explicabilidad / transparencia*

La IA puede recomendar o tomar decisiones  $\Rightarrow$  *fairness* (equidad)

La IA tiene errores  $\Rightarrow$  *accountability* (responsabilidad)

La IA puede generar perfiles de personas y manipular sus preferencias  $\Rightarrow$  *moralidad*



# Ética en IA (y ML) [2]

La IA necesita datos  $\Rightarrow$  *privacidad / governance*

La IA es, con frecuencia, una caja negra  $\Rightarrow$  *explicabilidad / transparencia*

La IA puede recomendar o tomar decisiones  $\Rightarrow$  *fairness* (equidad)

La IA tiene errores  $\Rightarrow$  *accountability* (responsabilidad)

La IA puede generar perfiles de personas y manipular sus preferencias  $\Rightarrow$  *moralidad*

La IA es penetrante y dinámica  $\Rightarrow$  *grandes daños por mal uso / capacidad para transformaciones*

# Ética en IA (y ML) [2]

La IA necesita datos  $\Rightarrow$  *privacidad / governance*

La IA es, con frecuencia, una caja negra  $\Rightarrow$  *explicabilidad / transparencia*

La IA puede recomendar o tomar decisiones  $\Rightarrow$  *fairness* (equidad)

La IA tiene errores  $\Rightarrow$  *accountability* (responsabilidad)

La IA puede generar perfiles de personas y manipular sus preferencias  $\Rightarrow$  *moralidad*

La IA es penetrante y dinámica  $\Rightarrow$  *grandes daños por mal uso / capacidad para transformaciones*

Buen o mal uso de la tecnología  $\Rightarrow$  *desarrollo sostenible vs. armas autónomas*

## Why Amazon's Automated Hiring Tool Discriminated Against Women

[https://www.aclu.org/news/womens-rights/  
why-amazons-automated-hiring-tool-discriminated-against](https://www.aclu.org/news/womens-rights/why-amazons-automated-hiring-tool-discriminated-against)

## Why Amazon's Automated Hiring Tool Discriminated Against Women

[https://www.aclu.org/news/womens-rights/  
why-amazons-automated-hiring-tool-discriminated-against](https://www.aclu.org/news/womens-rights/why-amazons-automated-hiring-tool-discriminated-against)

## Facebook enables gender discrimination in job ads

[https://edition.cnn.com/2025/02/28/tech/  
facebook-gender-discrimination-europe-ruling-asequals-intl](https://edition.cnn.com/2025/02/28/tech/facebook-gender-discrimination-europe-ruling-asequals-intl)

## Why Amazon's Automated Hiring Tool Discriminated Against Women

<https://www.aclu.org/news/womens-rights/why-amazons-automated-hiring-tool-discriminated-against>

## Facebook enables gender discrimination in job ads

<https://edition.cnn.com/2025/02/28/tech/facebook-gender-discrimination-europe-ruling-asequals-intl>

## Twitter investigates racial bias in image previews

<https://www.bbc.com/news/technology-54234822>

## Why Amazon's Automated Hiring Tool Discriminated Against Women

<https://www.aclu.org/news/womens-rights/why-amazons-automated-hiring-tool-discriminated-against>

## Facebook enables gender discrimination in job ads

<https://edition.cnn.com/2025/02/28/tech/facebook-gender-discrimination-europe-ruling-asequals-intl>

## Twitter investigates racial bias in image previews

<https://www.bbc.com/news/technology-54234822>

## AI360

- <https://github.com/Trusted-AI/AIF360>
- AI360 [1]

- Clasificación del Aprendizaje Automático y algoritmos
- Regresión
- Máquinas de soporte vectorial
- Redes Neuronales
- Aprendizaje no supervisado

¡Gracias!





BELLAMY, R. K. E., DEY, K., HIND, M., HOFFMAN, S. C., HOUDE, S., KANNAN, K., LOHIA, P., MARTINO, J., MEHTA, S., MOJSILOVIC, A., NAGAR, S., RAMAMURTHY, K. N., RICHARDS, J., SAHA, D., SATTIGERI, P., SINGH, M., VARSHNEY, K. R., AND ZHANG, Y.

AI Fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias, Oct. 2018.



BERGER, S. E., AND ROSSI, F.

Ai and neurotechnology: Learning from ai ethics to address an expanded ethics landscape.

*Communications of the ACM* 66, 3 (2023), 58–68.



BISHOP, C. M.

*Pattern Recognition and Machine Learning*, 1 ed.

Information Science and Statistics. Springer, New York, 2006.



BUNGE, M.

*Ética, ciencia y técnica.*

Sudamericana, Buenos Aires, 1997.



CARBONE, M. R.

When not to use machine learning: A perspective on potential and limitations.

*MRS Bulletin* 47, 9 (2022), 968–974.



FERRANTE, E., ALONSO ALEMANY, L.,

FERNANDEZ SLEZAK, D., FERRER, L., MILONE, D., AND

STEGMAYER, G.

*¿Aprendizaje automático?: Un viaje al corazón de la inteligencia artificial contemporánea.*

Ediciones UNL / Vera Editorial Cartonera, Argentina, 2022.

CC BY-NC-SA 4.0; disponible en Internet Archive.



FERRER, L., SCHARENBERG, O., AND BÄCKSTRÖM, T.

Good practices for evaluation of machine learning systems,  
2024.



MITCHELL, T. M.

*Machine Learning.*

McGraw-Hill, New York, 1997.



RUSSELL, S., AND NORVIG, P.

*Artificial Intelligence: A Modern Approach*, 3 ed.

Prentice Hall, 2010.



SAMUEL, A. L.

Some studies in machine learning using the game of  
checkers.

*IBM Journal of Research and Development* 3, 3 (1959),  
210–229.