Unidad I - Introducción al Aprendizaje Automático

Germán Braun

Facultado de Informática - Universidad Nacional del Comahue

german.braun@fi.uncoma.edu.ar

11 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

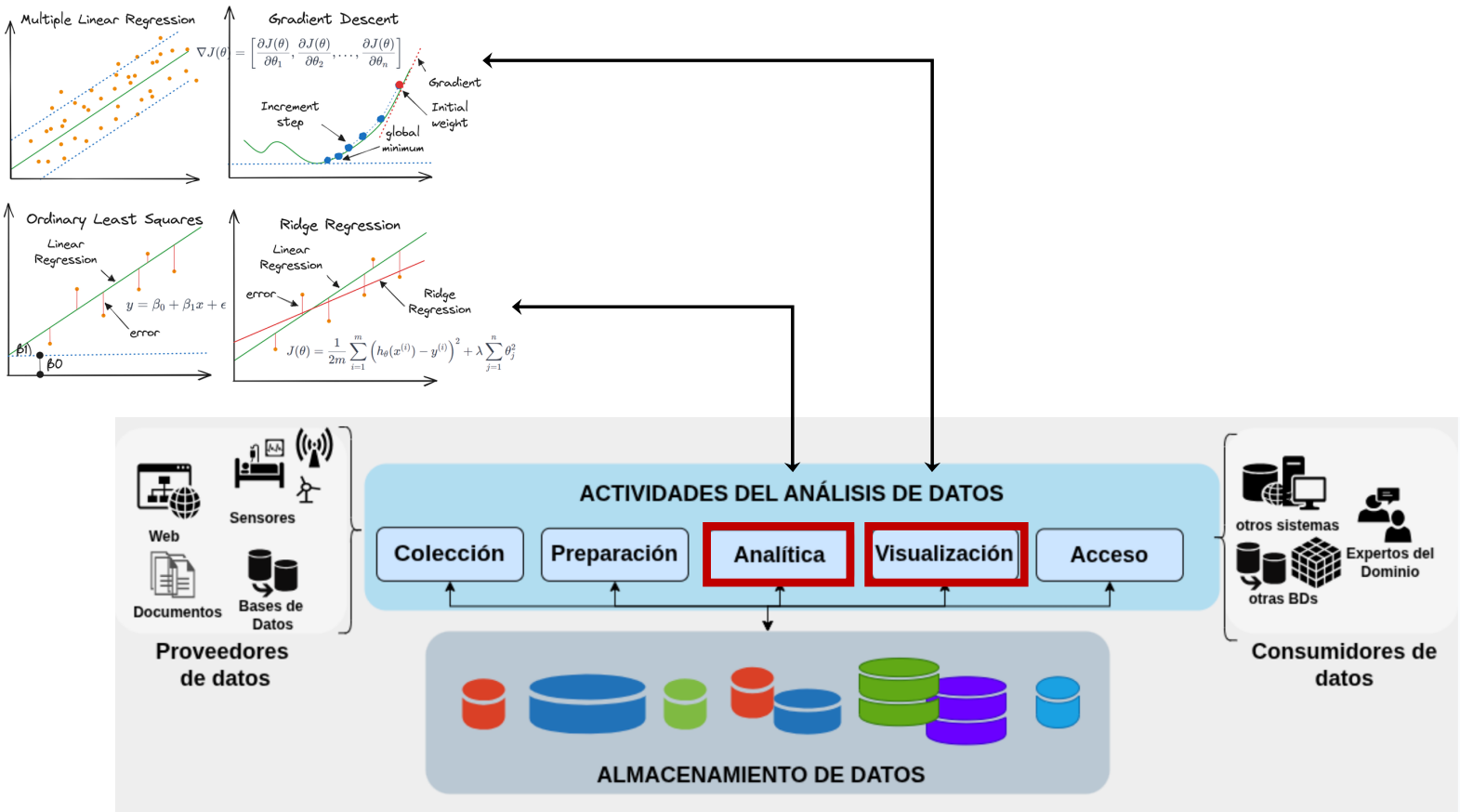


Figura 1.1: Proceso de Análisis de Datos

(*) figuras adaptadas de Buccella y `https://github.com/leehanchung/CSMM-101x-AI`

Inteligencia Artificial

“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

Inteligencia Artificial

"Definición"

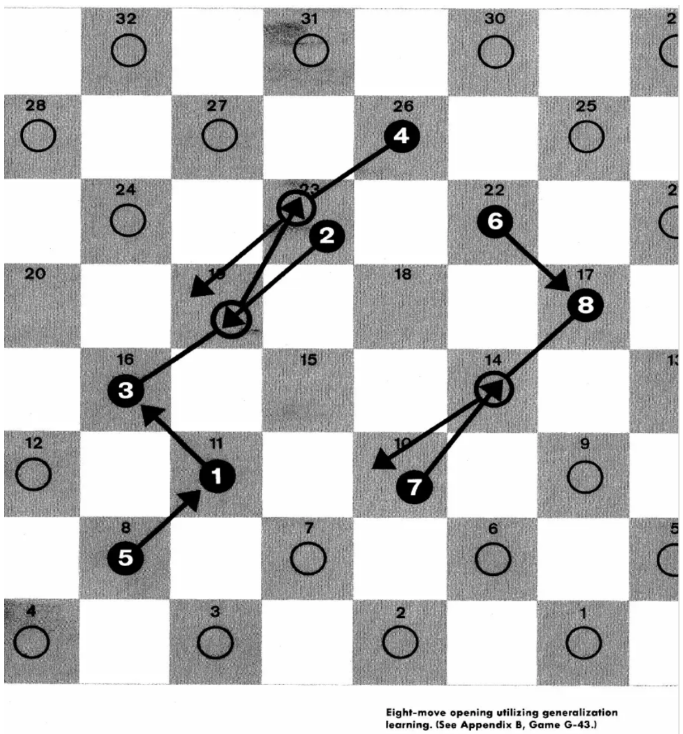
...disciplina que se encarga de comprender y construir entidades artificiales inteligentes que simulan en algún sentido el comportamiento humano [6]

- *knowledge-based* > un programa cuya lógica codifica a gran número de propiedades del mundo y se concluye usando razonamiento lógico deductivo
- *machine learning* > extraer información (patrones) directamente a partir de datos históricos y extrapolar para hacer predicciones

Aprendizaje Automático - Game of Checkers (1956)

Arthur Samuel [10]

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



code[1]

[1] https://github.com/almostimplemented/checkers

Aprendizaje Automático (con't)

Tom Mitchell [8]

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

Aprendizaje Automático (con't)

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.

Ejemplos

Ejemplo (Samuel + Mitchell)

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

Ejemplos

Ejemplo (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

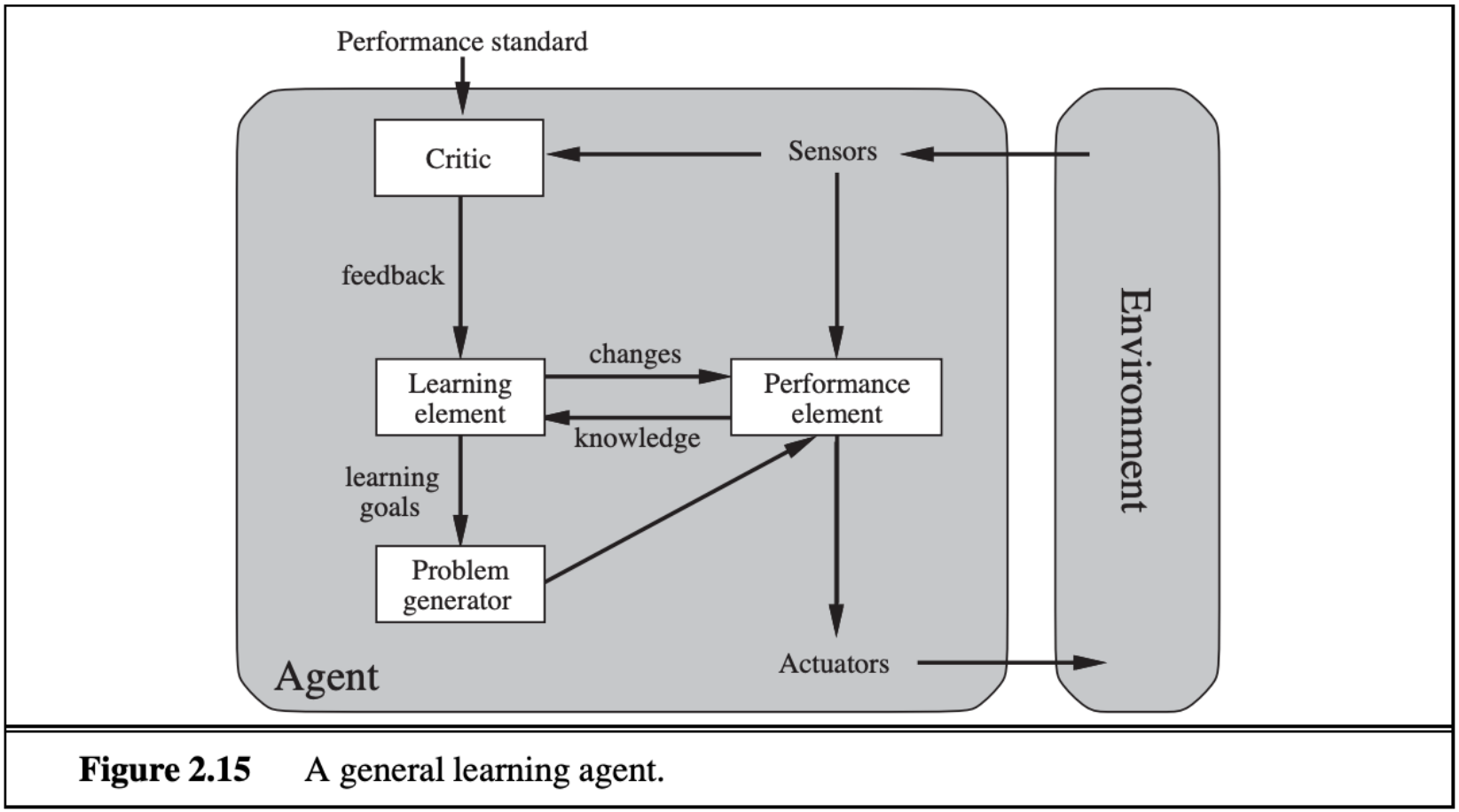
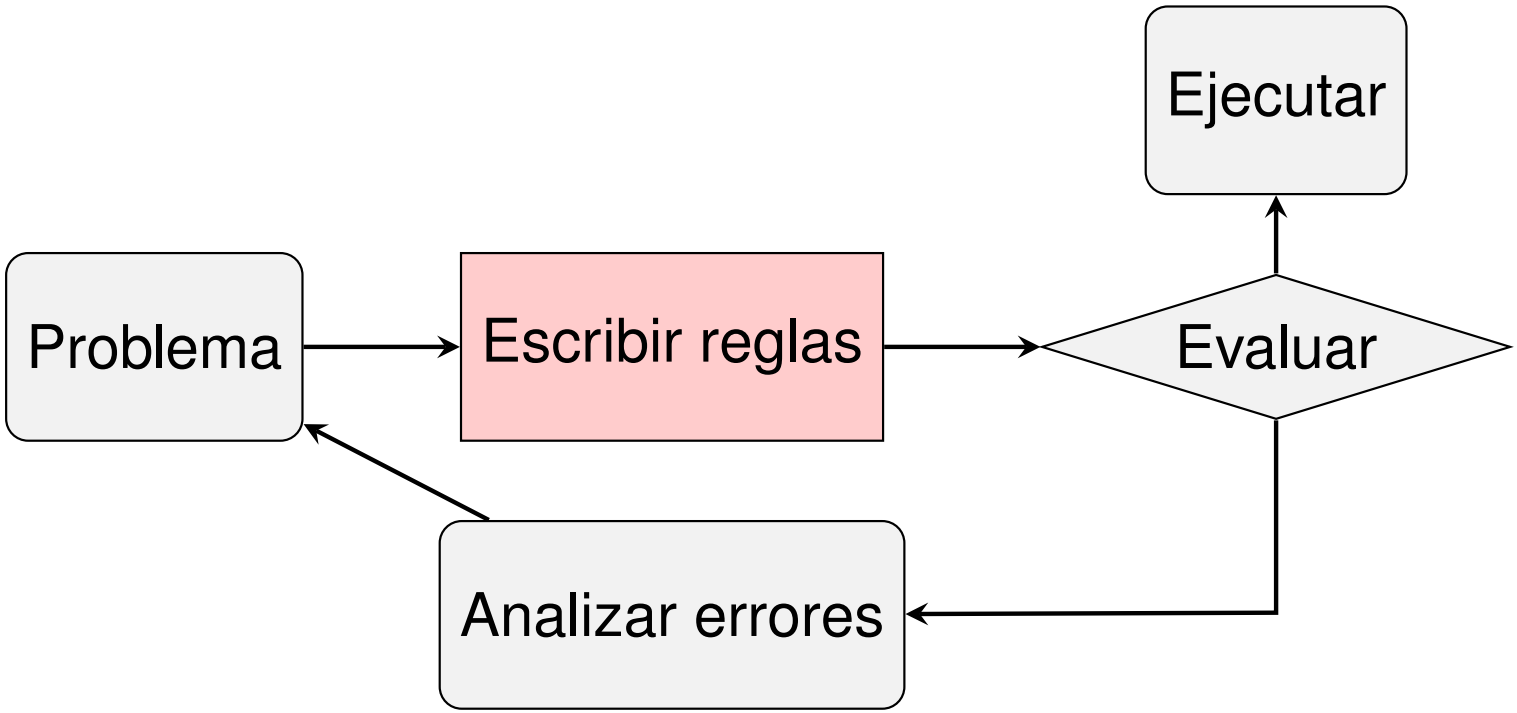


Figure 2.15 A general learning agent.

Figura 1.2: Extraída de [9]

Programación tradicional



Créditos: Farhad Pourkamali Anaraki

Programación tradicional - Spam

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

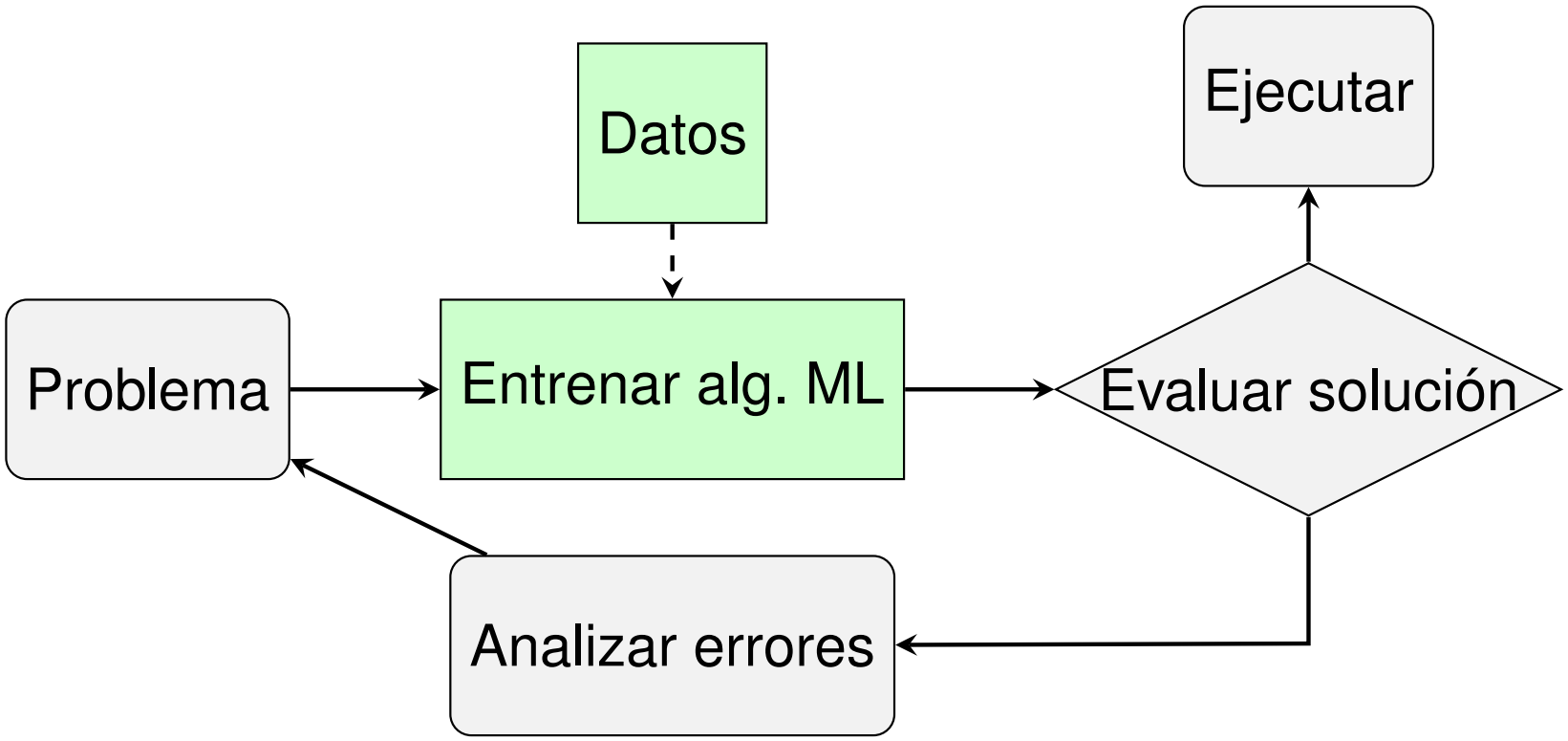
```
mensaje1 = "Gana dinero rapido. Haz clic
       aqui!"
mensaje2 = "Hola, te paso el informe
      adjunto."

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

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

Machine Learning

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



Créditos: Farhad Pourkamali Anaraki

Ejemplo: Aprendizaje Supervisado

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

Créditos: ...

Aplicaciones (solo algunas!)

Análisis/Clasificación de imágenes

- médicas `https://entelai.com/en/`
- identificación de personas/objectos (vehículos autónomos)

Aplicaciones (solo algunas!)

Análisis/Clasificación de imágenes

- médicas `https://entelai.com/en/`
- identificación de personas/objectos (vehículos autónomos)

Motores de recomendación

- Amazon > `https://aws.amazon.com/es/personalize/`
- Netflix > `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://entelai.com/en/`
- identificación de personas/objectos (vehículos autónomos)

Motores de recomendación

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

Procesamiento de lenguage natural

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

Machine Learning en práctica

Workflow

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

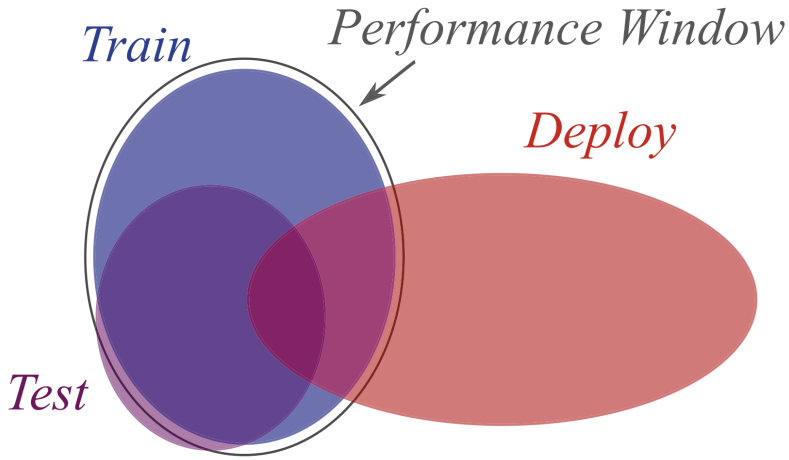


Figura 1.3: Un posible scenario de falla en la implementación de ML [5]

Aprendiendo patrones a partir de datos

Aprendizaje Inductivo

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]

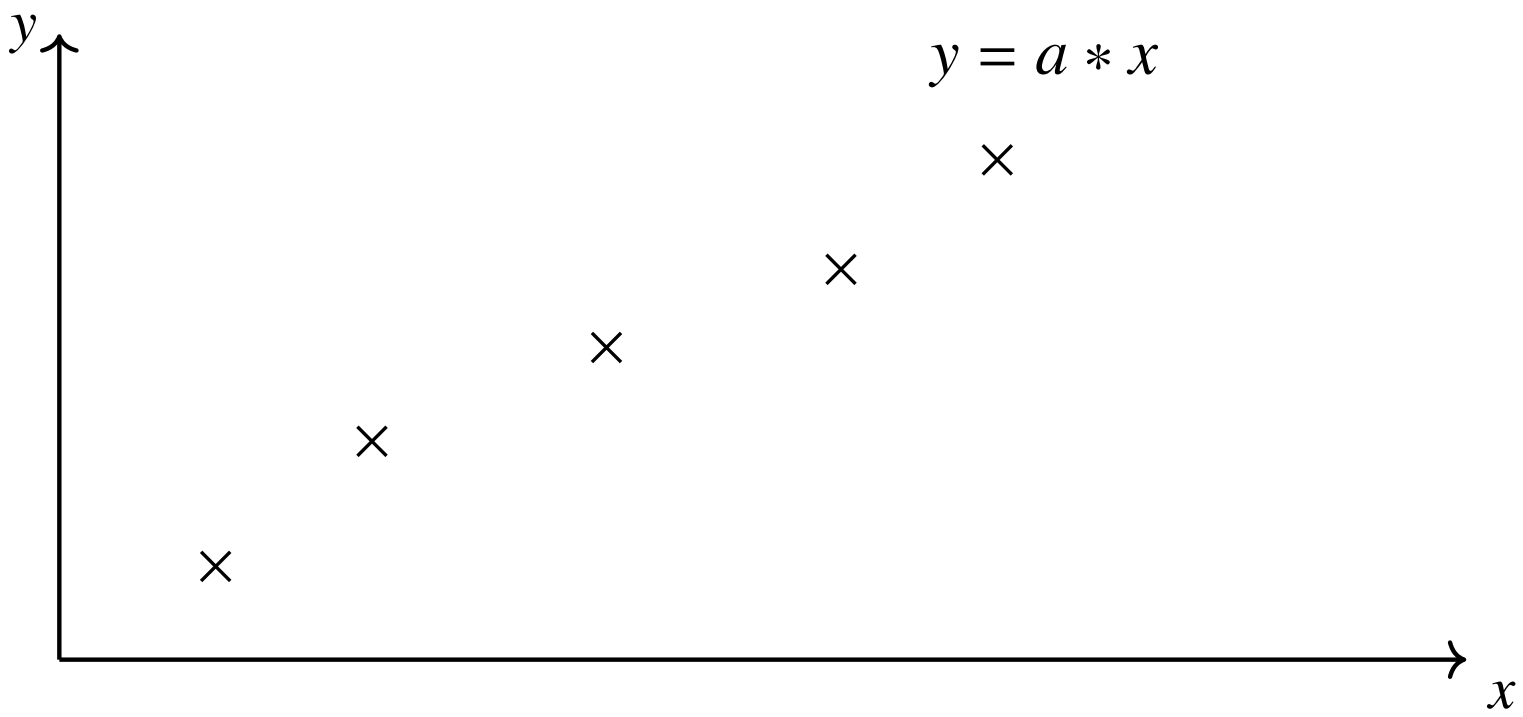
Aprendizaje Inductivo

Hipótesis del Aprendizaje Inductivo

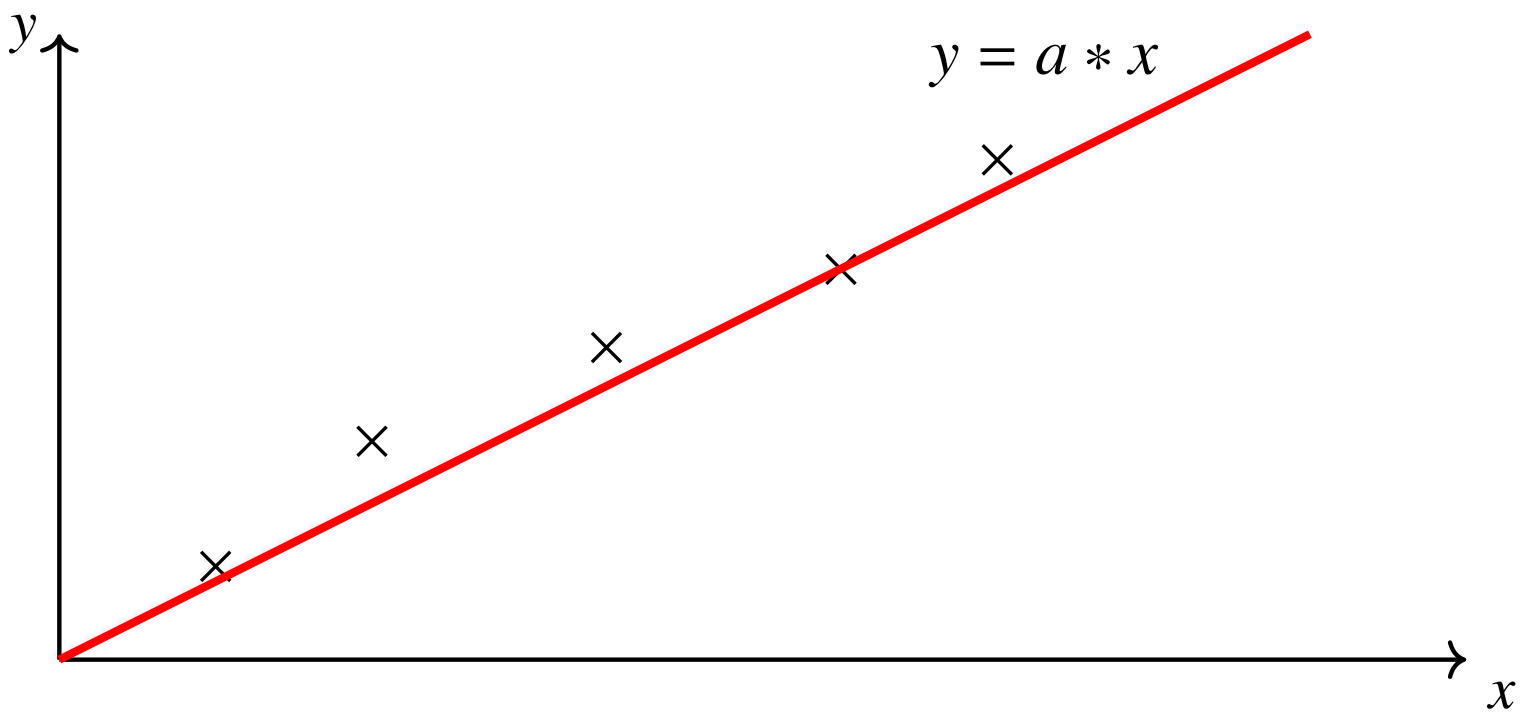
The inductive learning hypothesis. Any hypothesis found to approximate the target function well over a sufficiently large set of training examples will also approximate the target function well over other unobserved examples. [8]

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

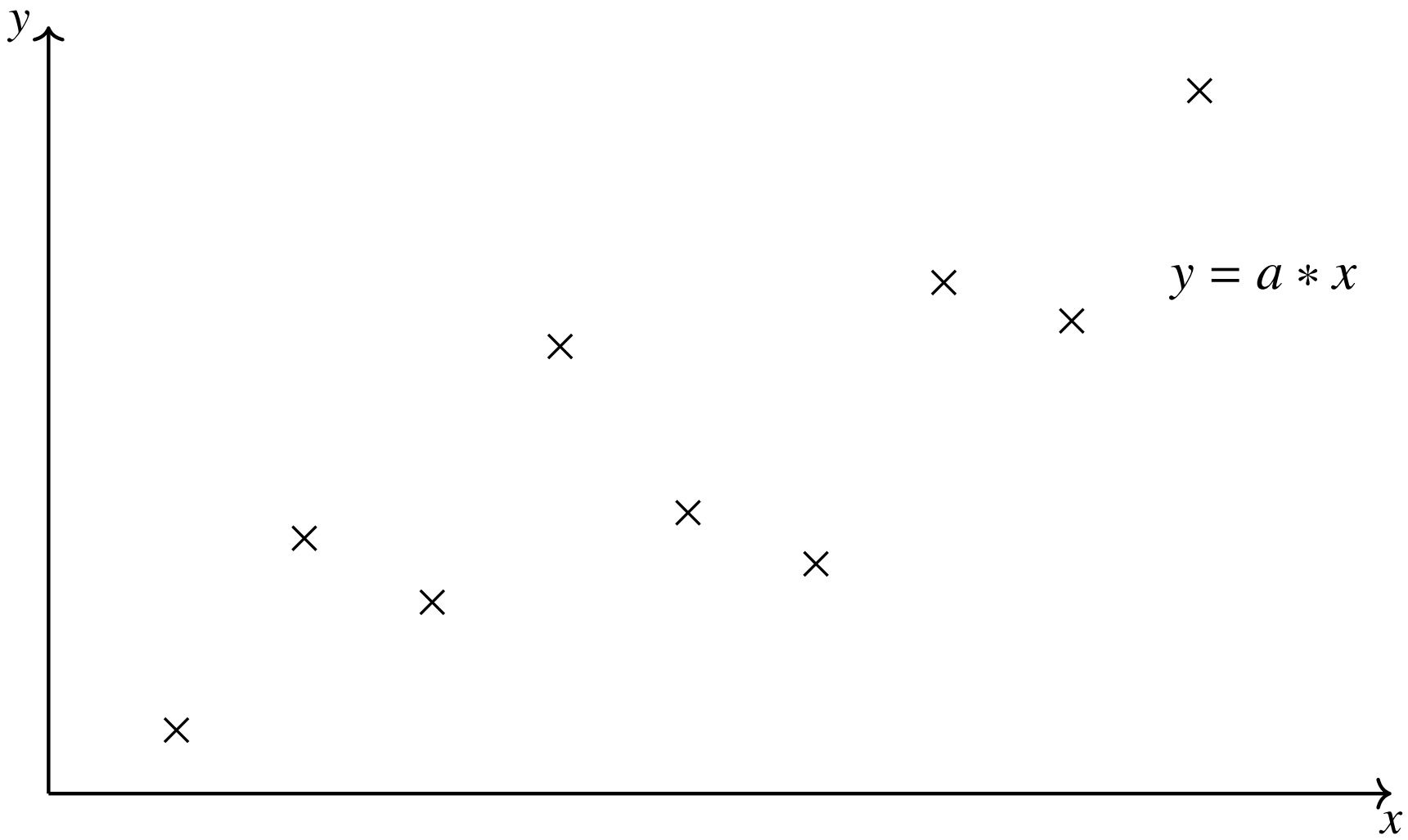
Inferencia Empírica



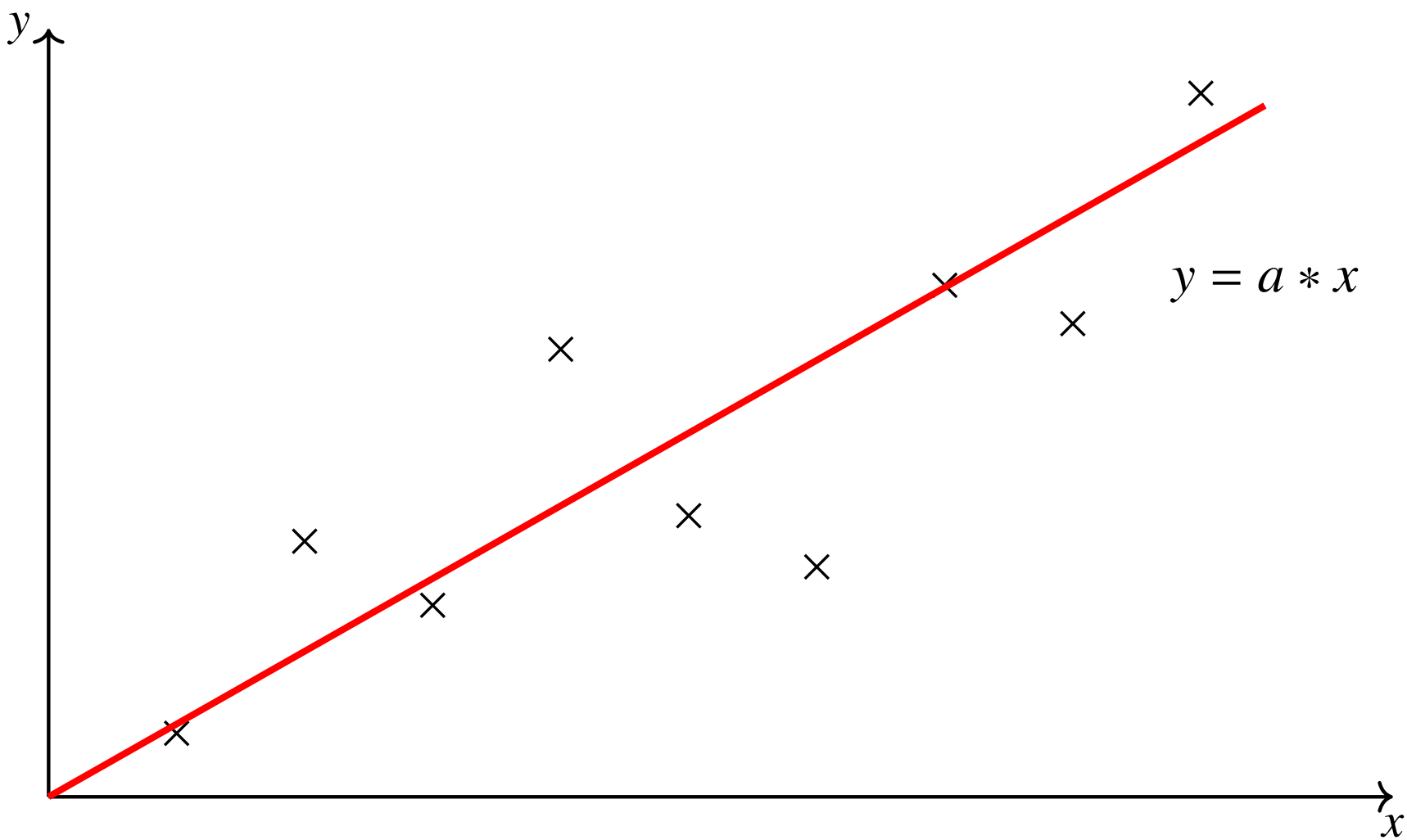
Inferencia Empírica



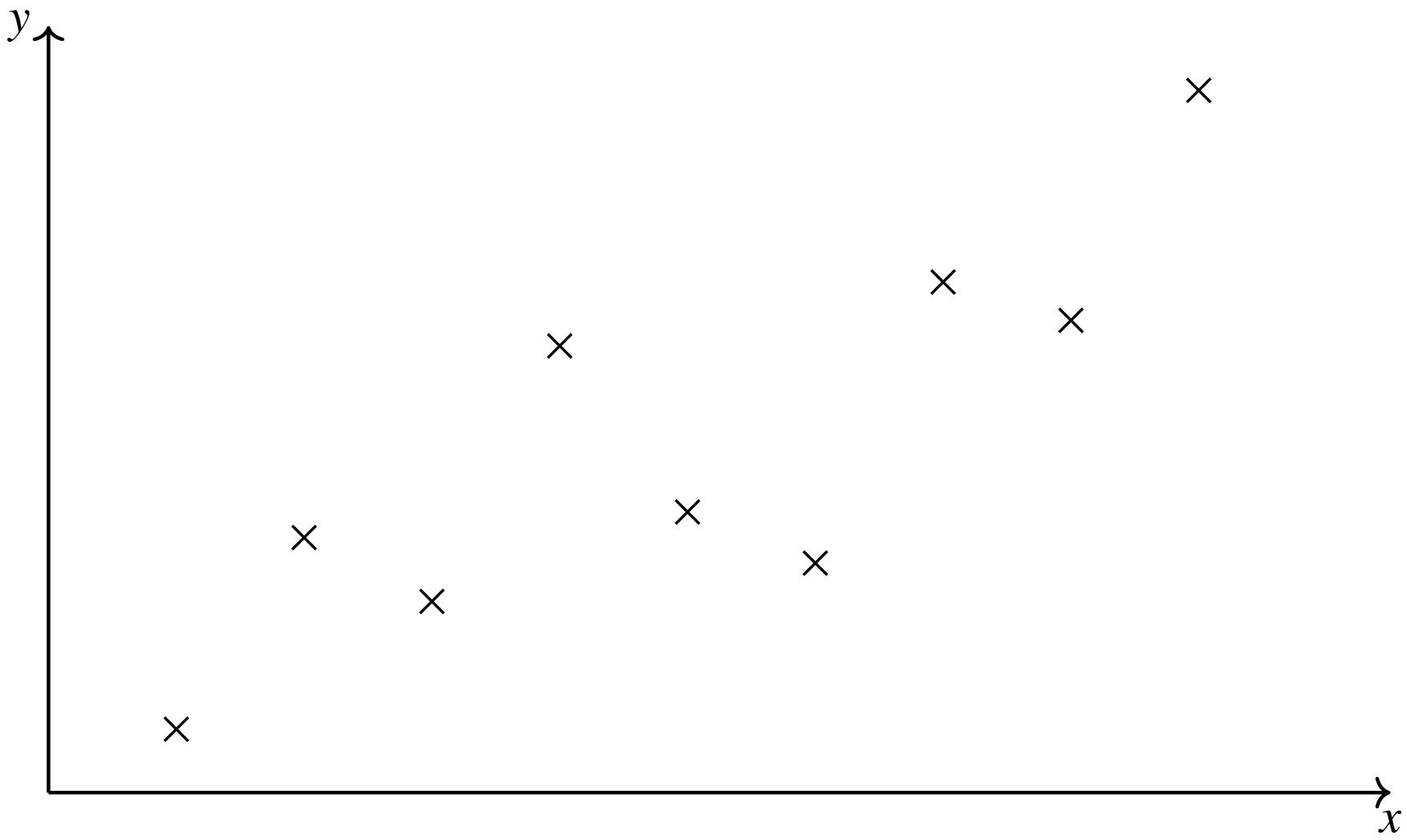
Inferencia Empírica (cont'd)



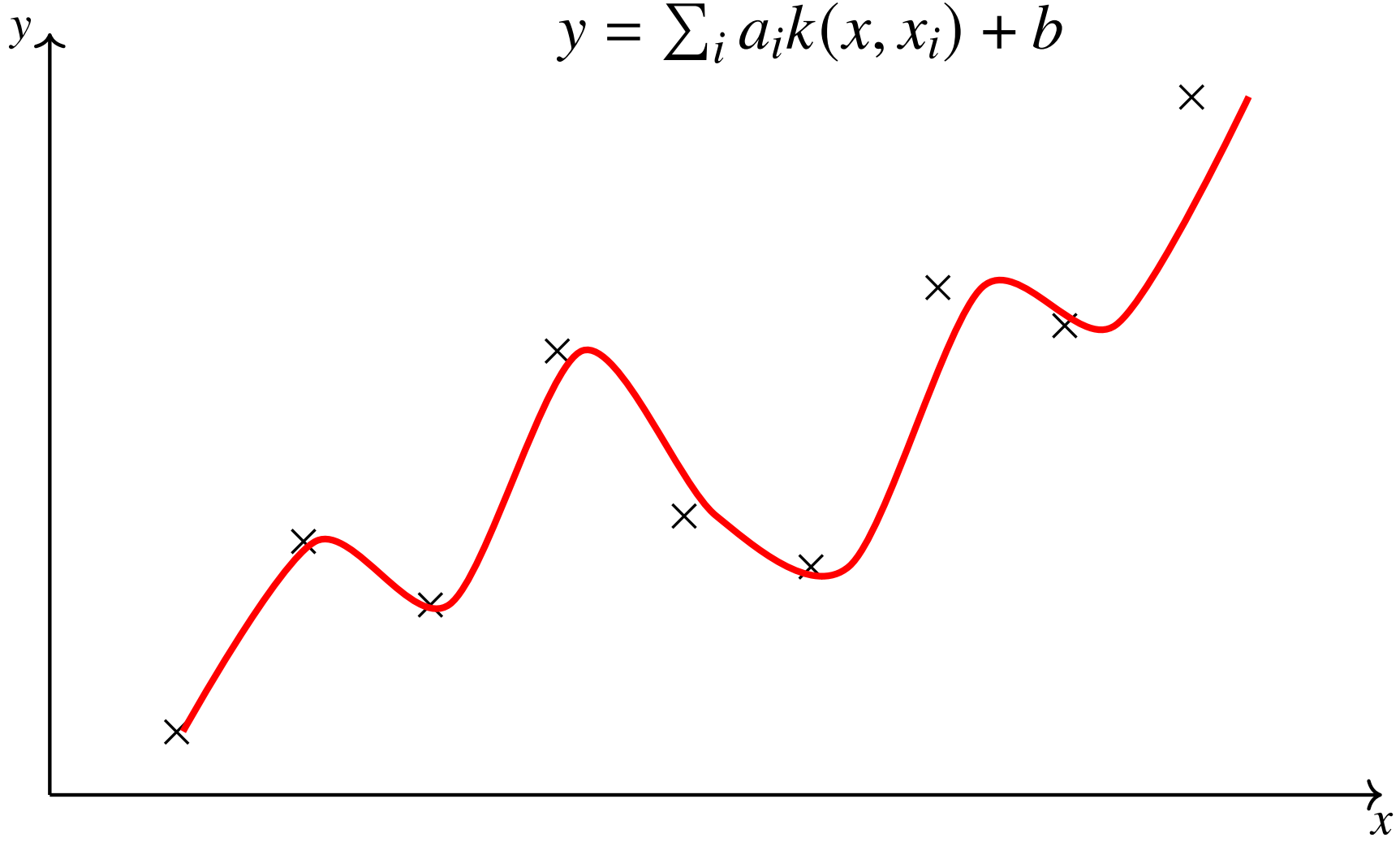
Inferencia Empírica (cont'd)



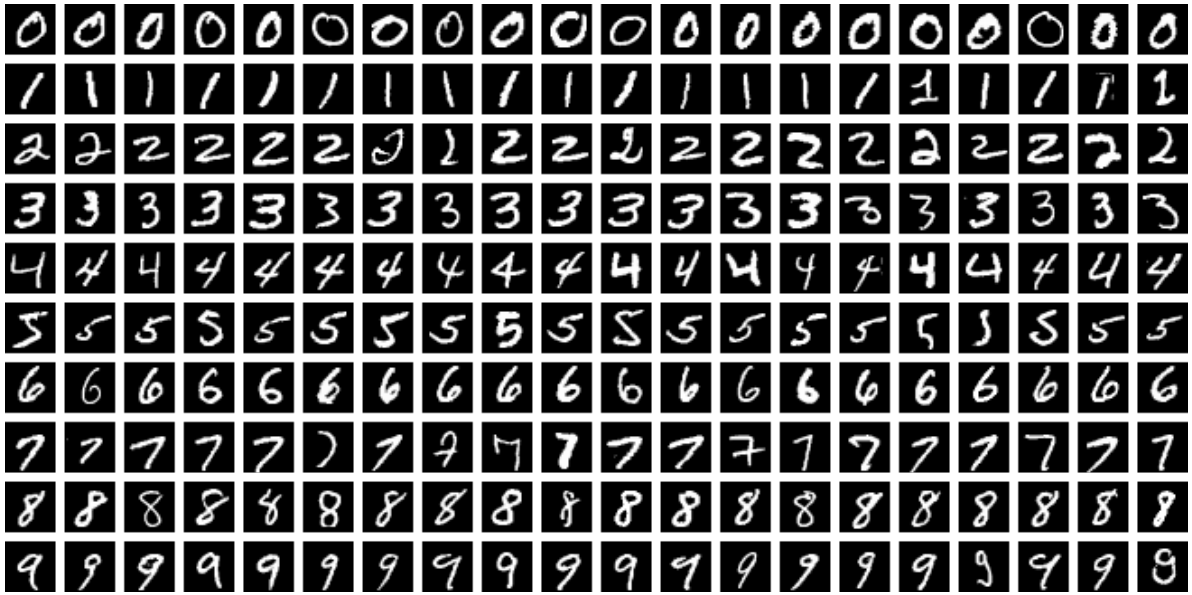
Inferencia Empírica (cont'd)



Inferencia Empírica (cont'd)



Inferencia Empírica - Percepción



(*) figura extraída de `https://en.wikipedia.org/wiki/MNIST_database`

Clasificador de 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!

Clasificador de 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!

SPAM!

Clasificador de 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!

SPAM! **(1)**

Clasificador de 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!

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)**

Clasificador de 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!

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 caracterísitca del correo e y como su categoria: spam (1) y no spam (0). Por simplicidad, las característcas 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!

Construimos el clasificador de spam

Definimos a x como una caracterísitca del correo e y como su categoria: spam (1) y no spam (0). Por simplicidad, las característcas 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!

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

Construimos el clasificador de spam

Definimos a x como una caracterísitca del correo e y como su categoria: spam (1) y no spam (0). Por simplicidad, las característcas 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!

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

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

Construimos el clasificador de spam

Definimos a x como una caracterísitca del correo e y como su categoria: spam (1) y no spam (0). Por simplicidad, las característcas 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!

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

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

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

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 caracterísitca del correo e y como su categoria: spam (1) y no spam (0). Por simplicidad, las característcas 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 caracterísitca del correo e y como su categoria: spam (1) y no spam (0). Por simplicidad, las característcas 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 frequentes en el conjunto de entrenamiento

Construimos el clasificador de spam (cont'd)

```
# Ejemplo de datos
X = [
    [1, 0, 1, 0, 0, 1],
    [1, 1, 1, 1, 0, 1],
    [0, 0, 1, 0, 0, 1]
]

y = [1, 1, 0]

# Entrenar modelo (por ejemplo, SVM, etc.)
model = EntrenarModelo(X, y)

# Predecir si un nuevo mensaje es spam
nuevo_correo = [0, 1, 0, 1, 1, 0]
prediccion = model.predecir(nuevo_correo)

print("Spam" if prediccion == 1 else "No spam")
```

Buenas Prácticas y Evaluación de Performance

Machine Learning en Práctica (cont'd)

¡Recordatorio!

El aprendizaje automático es un proceso de prueba y error.

Machine Learning en Práctica (cont'd)

¡Recordatorio!

El aprendizaje automático es un proceso de prueba y error.

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

Función de Aproximación

Problema

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

Input

- Ejemplos de entrenamiento $< x_i, y_i >$. 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.

“Issues” en el Aprendizaje Automático

Problema

1. ¿Contamos con un dataset “apropiado”?
2. ¿El dataset es representativo de los datos de producción? (distribución)
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)
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)
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)
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)
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)
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)
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, arquitecutura, características)

“Issues” en el Aprendizaje Automático - Problema

1. ¿Contamos con un dataset “apropiado”?
2. ¿El dataset es representativo de los datos de producción? (distribución)
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, arquitecutura, características)
  3. **Conjunto de evaluación (*evaluation set*)**, usado para reportar la performance final del model aprendido.

“Issues” en el AA - Problema (cont’d)

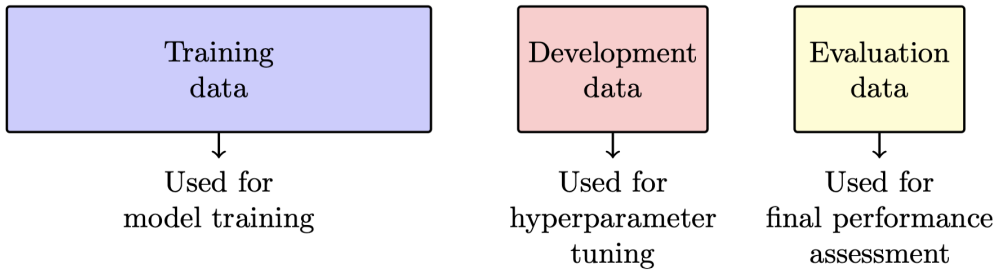


Figura 3.1: Datasets en AA [7]

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

“Issues” en el AA - Problema (cont’d)

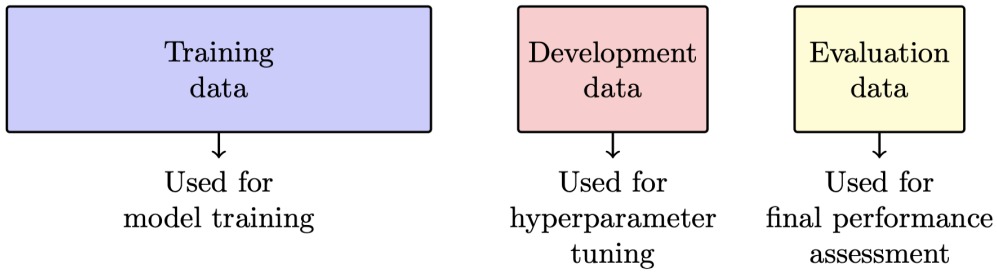


Figura 3.1: Datasets en AA [7]

- (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 perfomance 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)

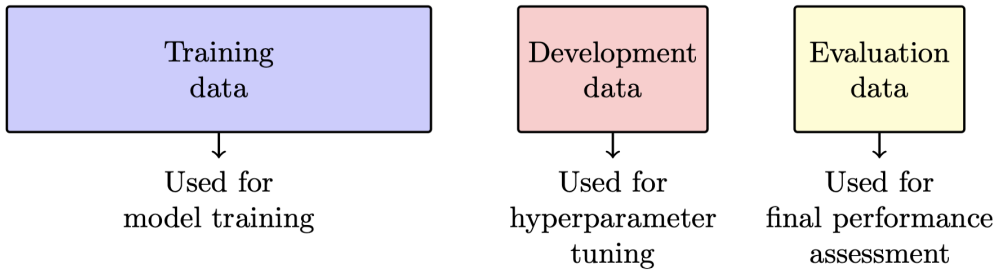


Figura 3.1: Datasets en AA [7]

- (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 perfomance 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)

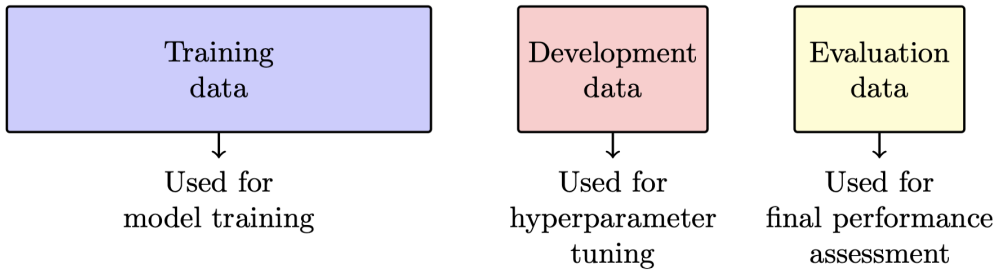


Figura 3.1: Datasets en AA [7]

- (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 perfomance 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 divirlo en conjuntos desarrollo/entrenamiento/evaluación?

“Issues” en el AA - Problema (cont’d)

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

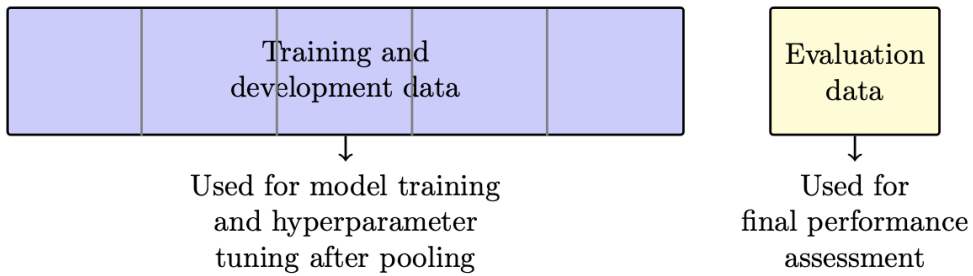


Figura 3.2: Cross-validation [7]

“Issues” en el AA - Problema (cont’d)

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

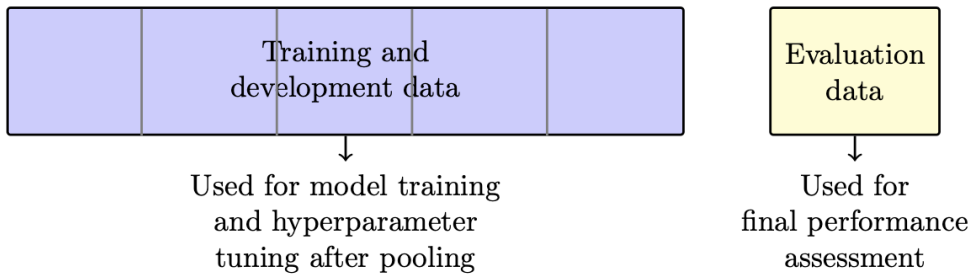


Figura 3.2: Cross-validation [7]

- 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 divirlo en conjuntos desarrollo/entrenamiento/evaluación?

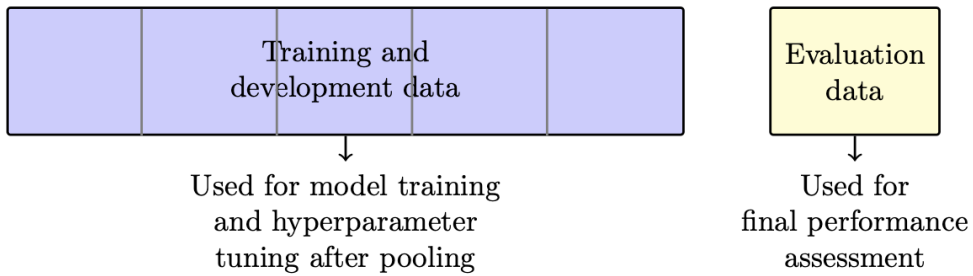


Figura 3.2: Cross-validation [7]

- 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 divirlo en conjuntos desarrollo/entrenamiento/evaluación?

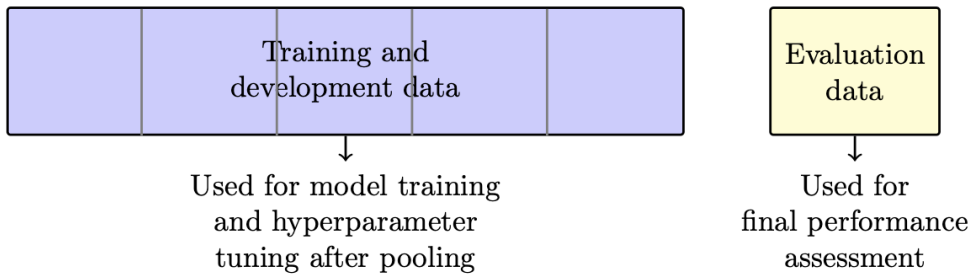


Figura 3.2: Cross-validation [7]

- 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 divirlo en conjuntos desarrollo/entrenamiento/evaluación?

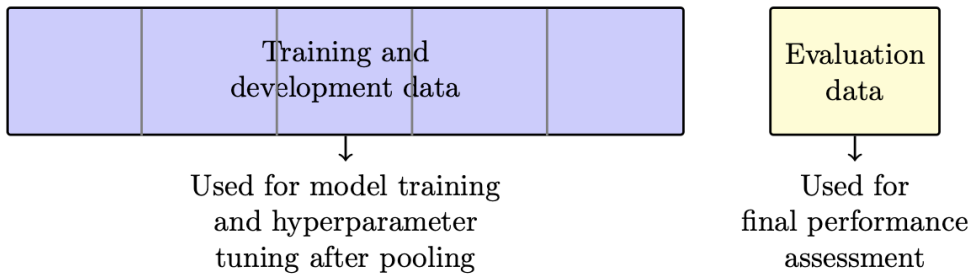


Figura 3.2: Cross-validation [7]

- 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” del los parámetros

“Issues” en el AA - Reglas de Oro!

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

“Issues” en el Aprendizaje Automático - Input

1. ¿Como elegimos el conjunto de entrenamiento?

“Issues” en el Aprendizaje Automático - Input

1. ¿Como elegimos el conjunto de entrenamiento?

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

“Issues” en el Aprendizaje Automático - Input

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*

“Issues” en el Aprendizaje Automático - Input

1. ¿Como elegimos el conjunto de entrenamiento?

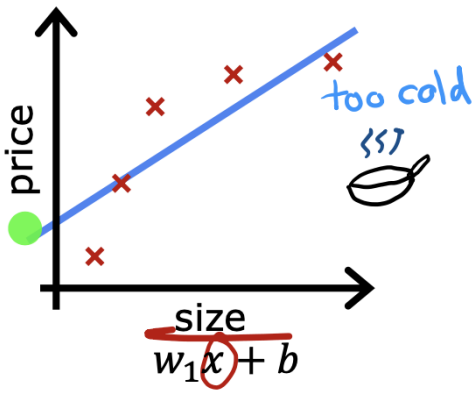
- El conjunto de entrenamiento podría tener una distribución diferente a la de los conjuntos restantes.
  - ej si estamos clasificando imagenes de perros, entonces el conjunto de entrenamiento podría incluir otros tipos de animales
- Si un modelo tiene una muy buena performance sobre el conjunto de entrenamiento 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)

“Issues” en el Aprendizaje Automático - Input

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!

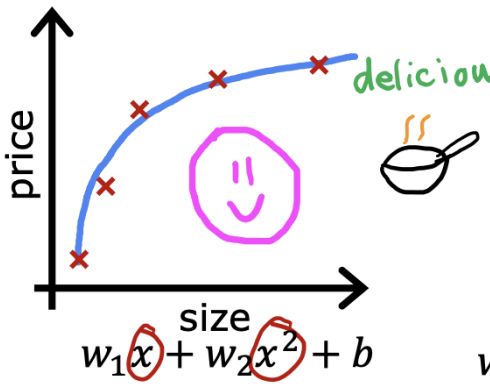
“Issues” en el AA - Overfitting



underfit

- Does not fit the training set well

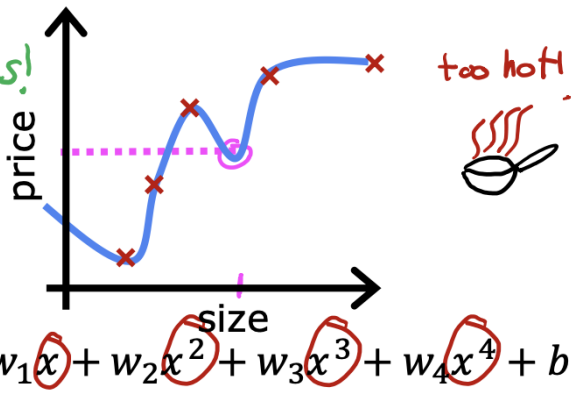
high bias



just right

- Fits training set pretty well

generalization



overfit

- Fits the training set extremely well

high variance

Créditos: Andrew Ng (Standford course)

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

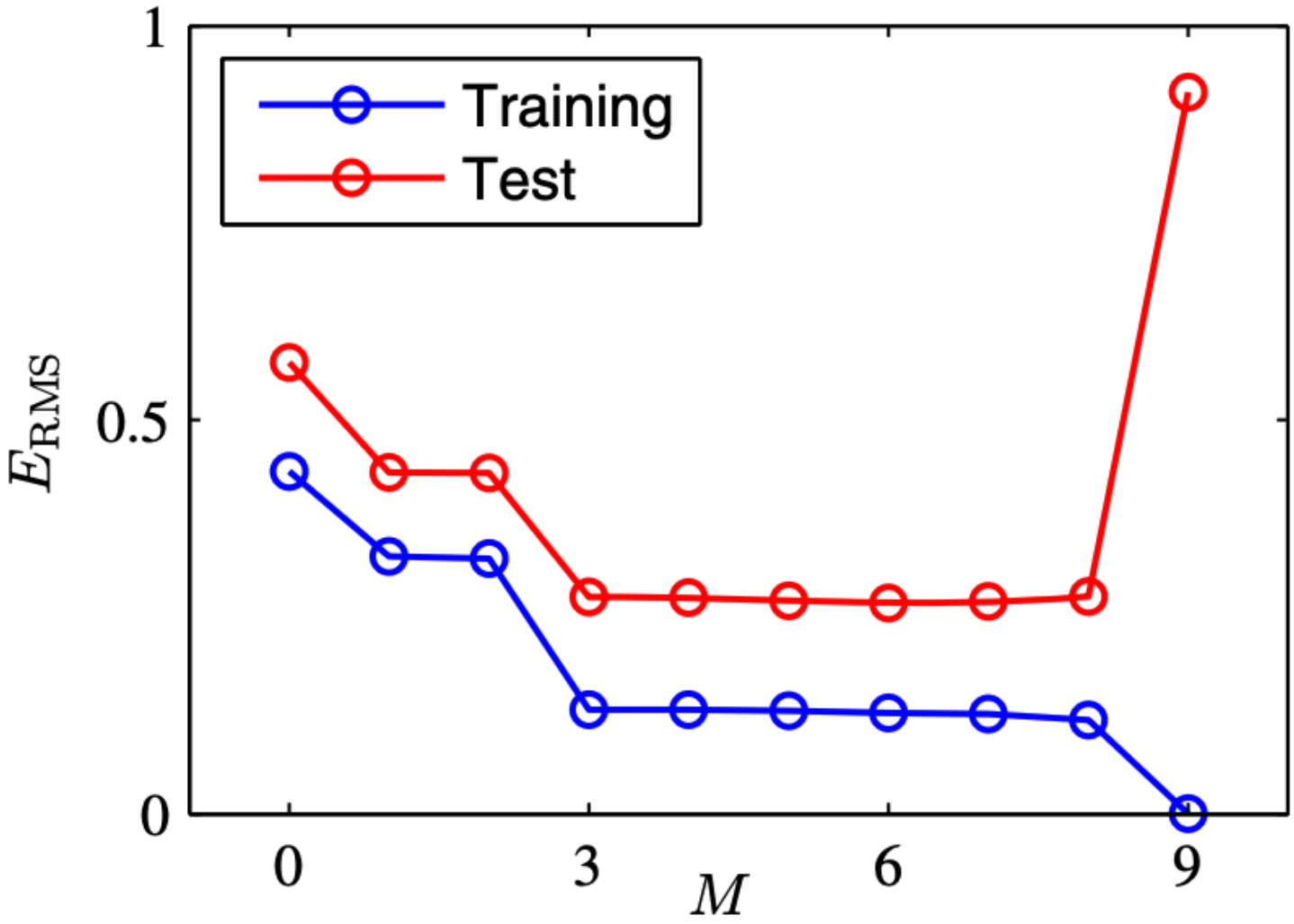


Figura 3.3: Error en training vs test (Bishop [3])

“Issues” en el Aprendizaje Automático - Output

Output

1. ¿Es mi hipótesis h la que “mejor” aproxima la función objectivo 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 objectivo 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 objectivo 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 θ_0 y θ_1 determinan la relación entre y y x.

“Issues” en el Aprendizaje Automático - Output

1. ¿Es mi hipótesis h la que “mejor” aproxima la función objectivo 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 asumciones pre-existentes.

“Issues” en el Aprendizaje Automático - Output

1. ¿Es mi hipótesis h la que “mejor” aproxima la función objectivo 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 asumciones pre-existentes.
  - ej Emails con “GANASTE EL PREMIO!” en el asunto son probablemente spams

“Issues” en el Aprendizaje Automático - Output

1. ¿Es mi hipótesis h la que “mejor” aproxima la función objectivo 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 asumciones 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*!

“Issues” en el AA - Métricas

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

“Issues” en el AA - Métricas

- 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

“Issues” en el AA - Métricas

- 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 erronea. 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

	Actual	
Predicción	**1**	**0**
1	**TP**	**FP**
0	**FN**	**TN**

Matriz de confusión / Precision / Recall

	Actual	
Predicción	**1**	**0**
1	**TP**	**FP**
0	**FN**	**TN**

	Actual	
predicción	**1**	**0**
1	**85**	**10**
0	**15**	**90**

- **Precisión**

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

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

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

Matriz de confusión / Precision / Recall

	Actual	
Predicción	**1**	**0**
1	**TP**	**FP**
0	**FN**	**TN**

	Actual	
predicción	**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**

	Actual	
Predicción	**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)

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

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

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

- **Matriz**

	Actual	
Predicción	**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.

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

- **Matriz**

	Actual	
Predicción	**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 $h(x) < 0{,}5$

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

- **Matriz**

	Actual	
Predicción	**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 $h(x) < 0{,}9$

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

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

- **Matriz**

	Actual	
Predicción	**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 $h(x) < 0{,}9$

$\Longrightarrow$ 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 $h(x) < 0{,}3$

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

Matriz de confusión / Precision / Recall / F-score

Si tenemos difrentes 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**

Matriz de confusión / Precision / Recall / F-score

Si tenemos difrentes 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 $= 0$

Matriz de confusión / Precision / Recall / F-score

Si tenemos difrentes 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 difrentes 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 desbalancedas) – 95 % no spam y 5 % spam.

Matriz de confusión / Precision / Recall / F-score

Si tenemos difrentes 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 desbalancedas) – 95 % no spam y 5 % spam.
- F-score *penaliza* los valores desbalanceados

Matriz de confusión / Especificidad

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

	Actual	
Predicción	**1**	**0**
1	**TP**	**FP**
0	**FN**	**TN**

Matriz de confusión / Especificidad

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

	Actual	
Predicción	**1**	**0**
1	**TP**	**FP**
0	**FN**	**TN**

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

$$\#TN / \#Negativos_{actual}$$
$$= \#TN / (\#TN + \#FP)$$

Matriz de confusión / Exactitud (*accuracy*)

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

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**

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**

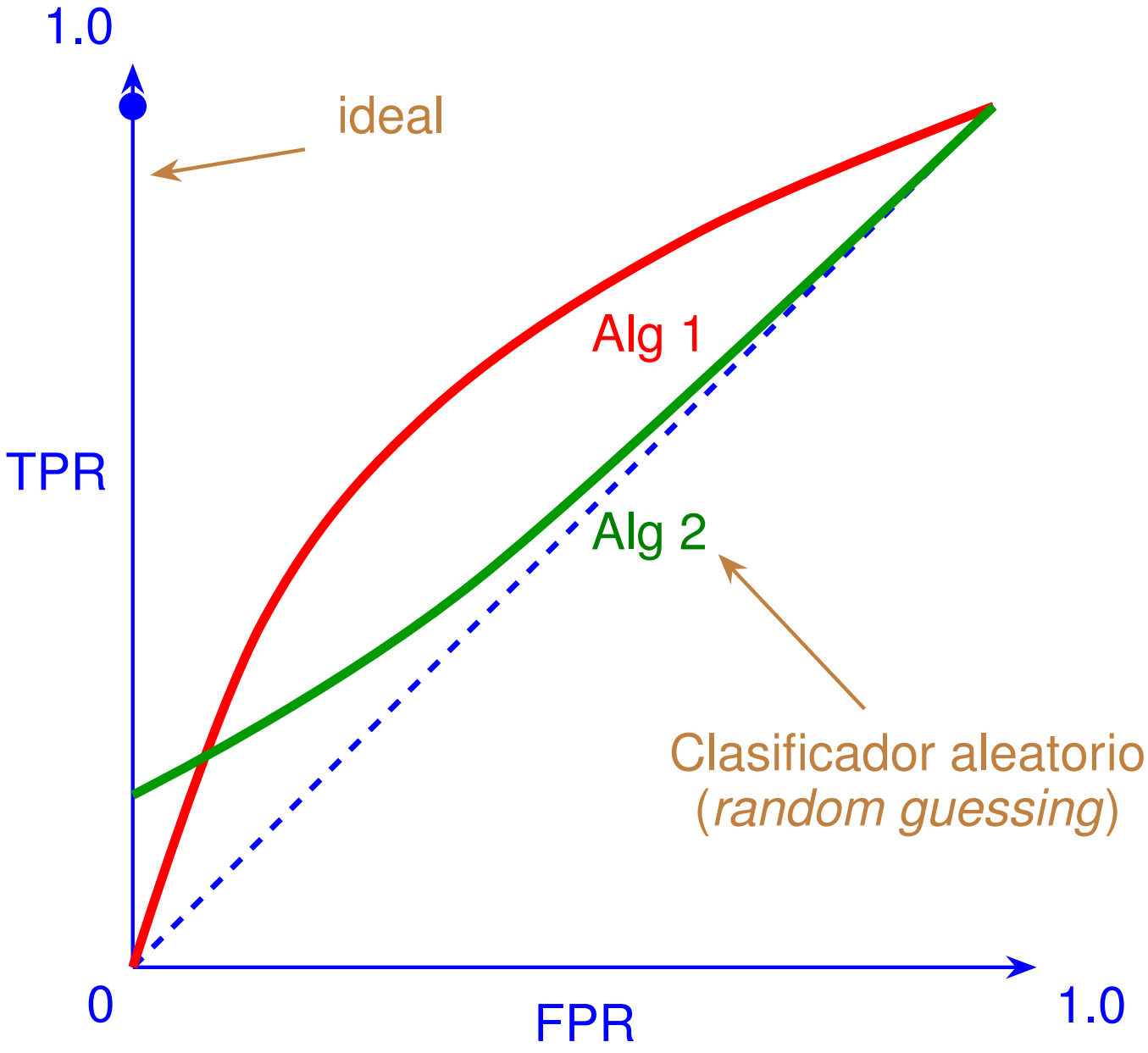
Por ejemplo, para un modelo detección de enfermedades extrañas se obtiene la siguiente matriz de confusión:

		Actual	
Predicción		**1**	**0**
1		**0**	**0**
0		**10**	**990**

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

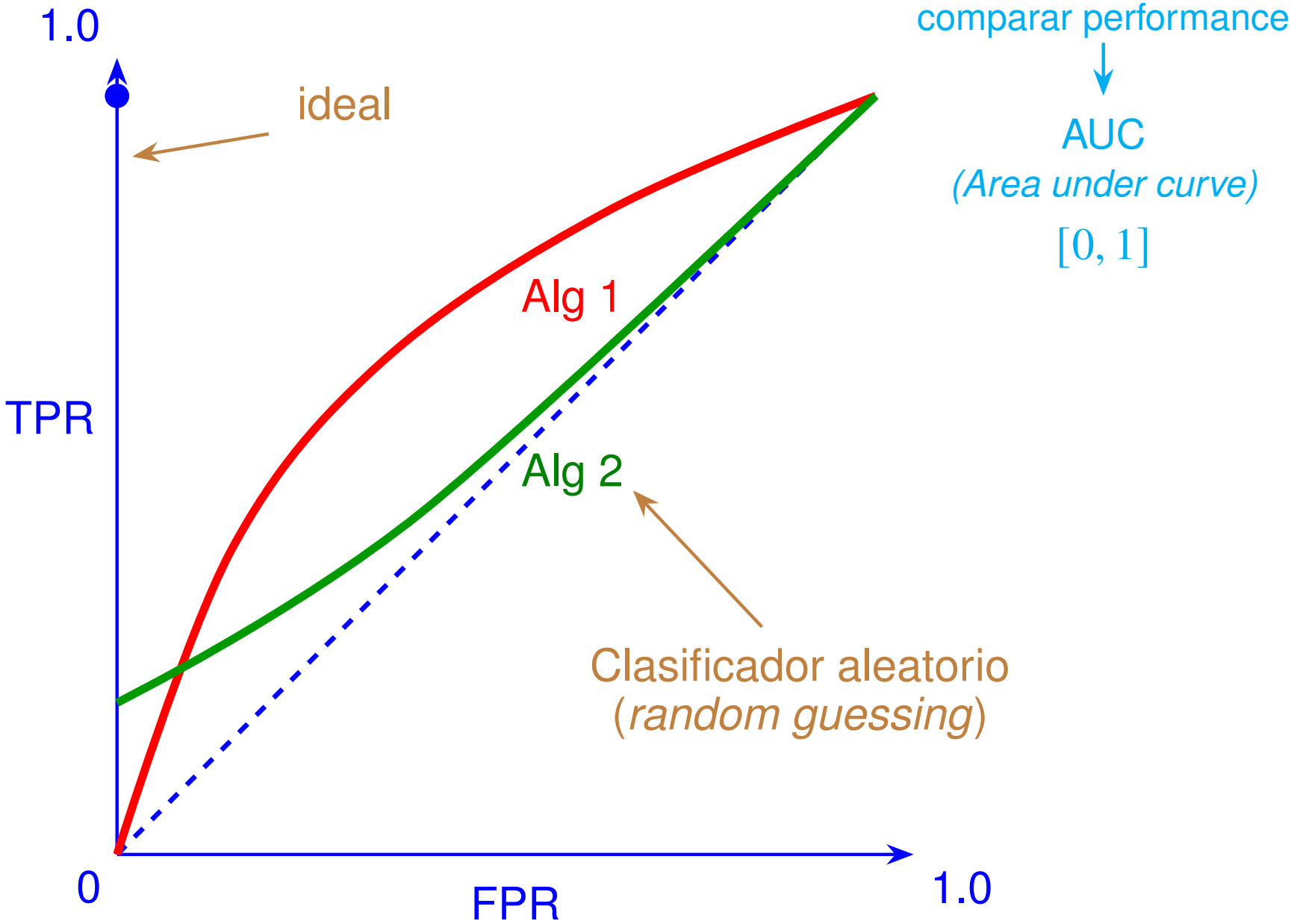
$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} = \frac{0 + 990}{1000} = 99\,\%$$

Métricas - Curvas ROC



Adaptado de `http://www.cs.wisc.edu/~dpage/cs760/`

Métricas - Curvas ROC



Adaptado de `http://www.cs.wisc.edu/~dpage/cs760/`

Aspectos Éticos

Ética de la ciencia

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

Ética en IA (y ML) [2]

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

Ética en IA (y ML) [2]

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

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

Ética en IA (y ML) [2]

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

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

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

Ética en IA (y ML) [2]

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

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

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

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

Ética en IA (y ML) [2]

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

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

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

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

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

Ética en IA (y ML) [2]

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

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

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

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

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

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

Ética en IA (y ML) [2]

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

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

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

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

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

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

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

Ética en IA (y ML) - cont'd

Why Amazon's Automated Hiring Tool Discriminated Against Women

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

Ética en IA (y ML) - cont'd

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

Ética en IA (y ML) - cont'd

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

Ética en IA (y ML) - cont'd

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]

Próximas clases

- Clasificación del Aprendizaje Automático y algortimos
- 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 automágico?: 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., SCHARENBORG, 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.