

Aprendizaje automático

Repaso de temas del curso



FACULTAD DE
INGENIERÍA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Tipos de aprendizaje



FACULTAD DE
INGENIERÍA



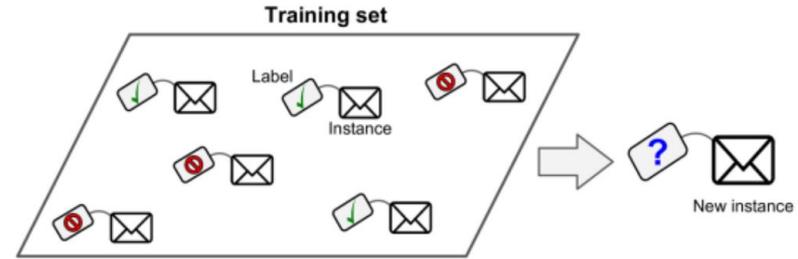
UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Tipos de aprendizaje

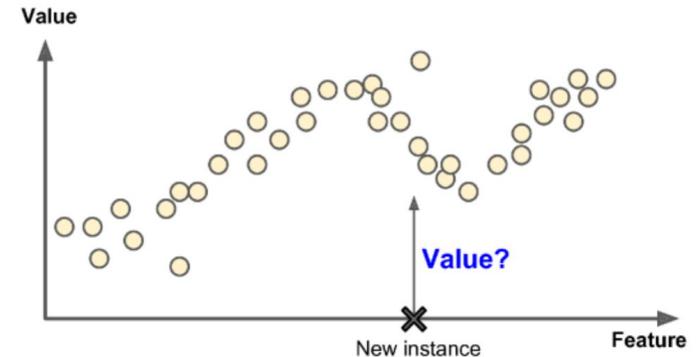
Aprendizaje supervisado

- Tenemos pares de datos de entrada-salida generados por humanos
- Según el tipo de salida dos posibles problemas:
 - salida categórica: clasificación
 - salida numérica: regresión

Problema de Clasificación



Problema de Regresión

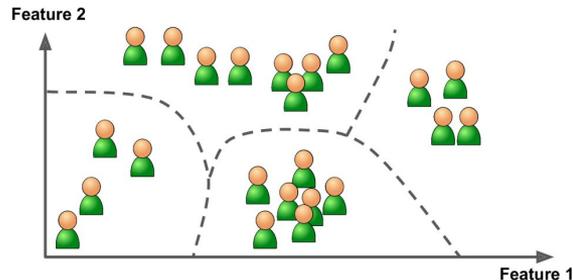
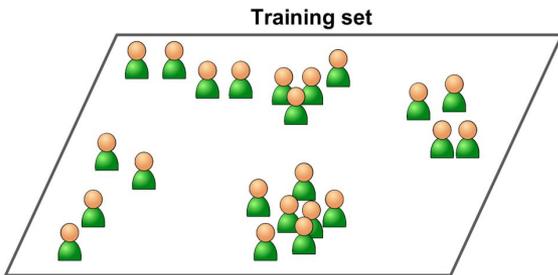


Tipos de aprendizaje

Aprendizaje no supervisado

La mayor parte de los datos disponibles no están etiquetados. Su potencial etiquetado requiere intervención humana (muchas veces expertos especializados) y su costo es muy alto.

- **Clustering:** descubrir estructura dentro de un conjunto de datos, agrupándolos en subconjuntos (clusters) que muestren una cierta coherencia o similitud interna.



Aprendizaje y generalización

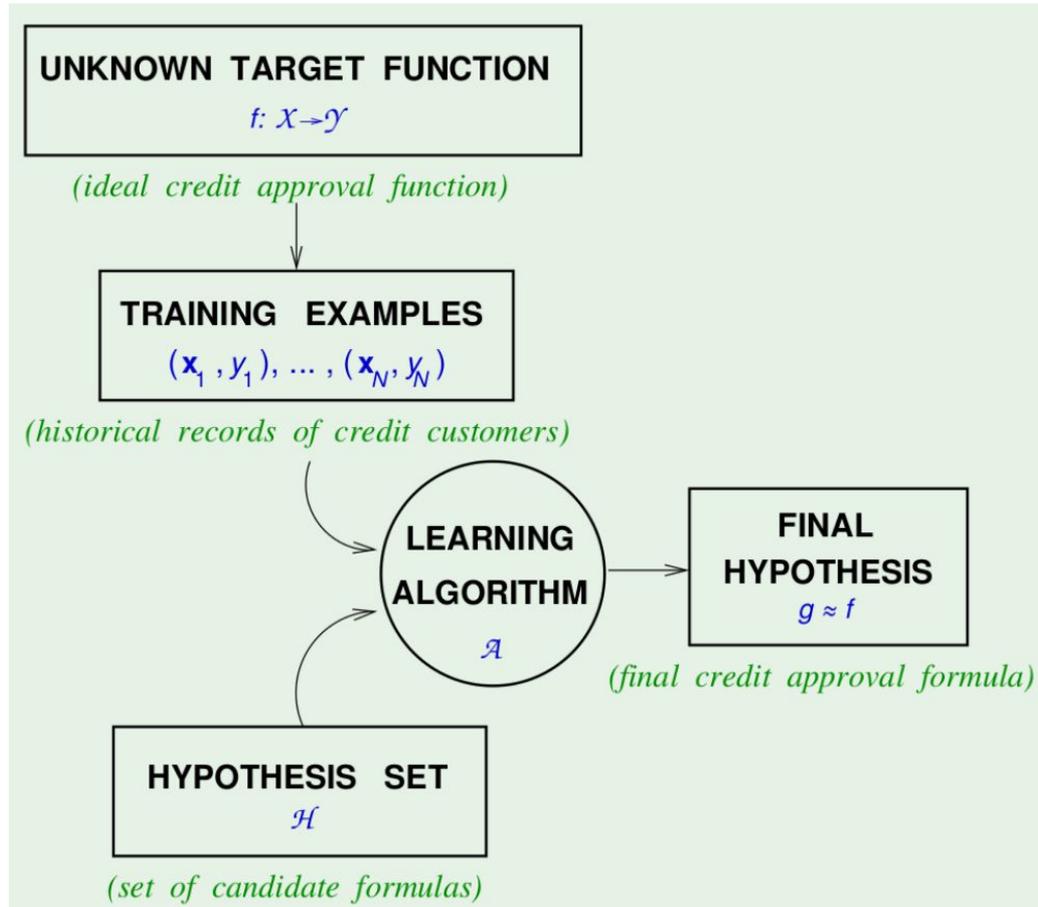


FACULTAD DE
INGENIERÍA



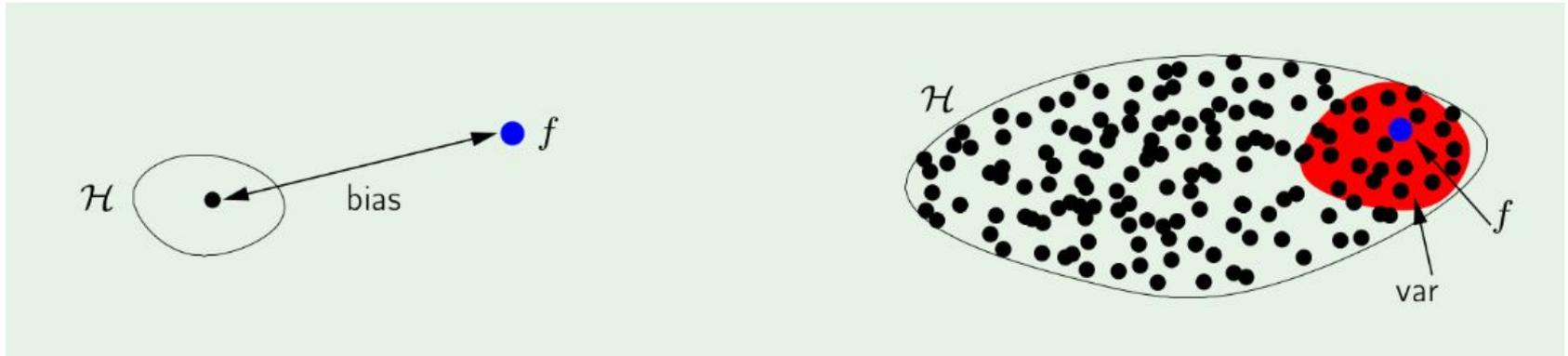
UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

El problema de aprendizaje



El problema de aprendizaje

- El set de hipótesis elegido influye en:
 - El error dentro de la muestra E_{in} (aproximación a las muestras disponibles)
 - El error fuera de la muestra E_{out} (error fuera de muestra)
- Queremos que nuestro modelo
 - aproxime bien a las muestras (x,y) disponibles
 - generalice bien para nuevos datos
- Va a existir un compromiso



Ajuste óptimo y Sobreajuste



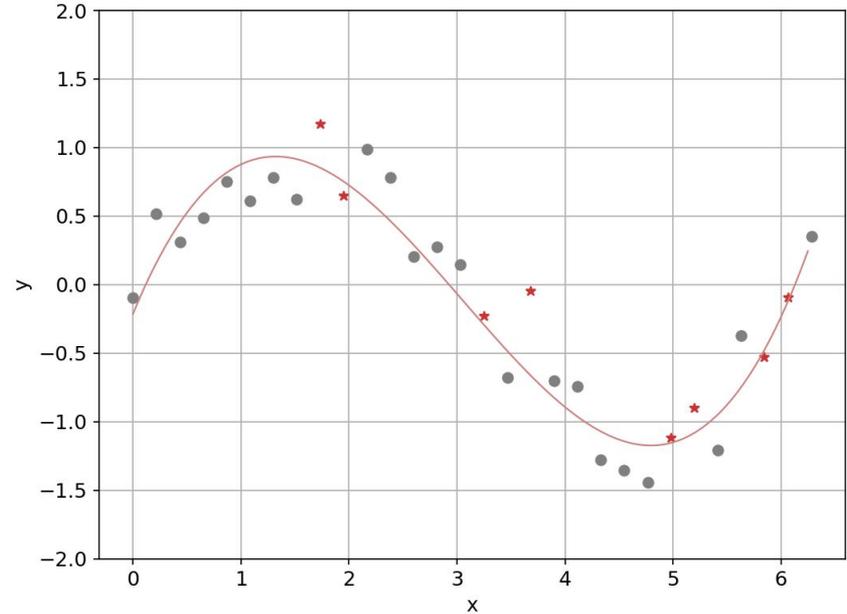
FACULTAD DE
INGENIERÍA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Ajuste óptimo

- Complejidad óptima
- Captura **toda** la estructura
- No captura **particularidades**
- **Cómo** identificar modelo ideal?
 - Selección de modelos
 - Validación cruzada
- Depende de cantidad de datos de entrenamiento disponibles!
- Depende del problema



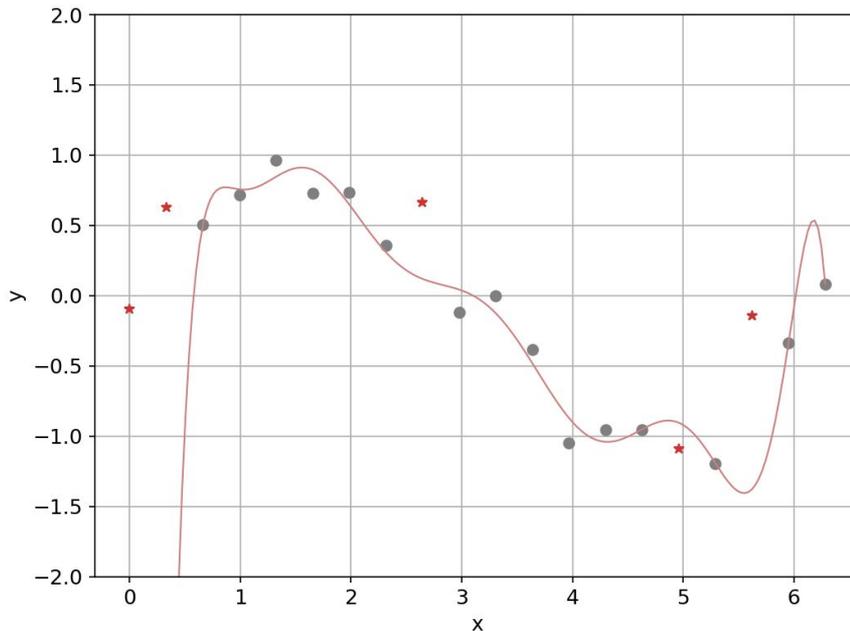
Sobreajuste (overfitting)

Escenario

- $g(\mathbf{x})$ demasiado compleja
- Demasiados parámetros
- Pocas muestras

Consecuencia

- Captura particularidades
- No captura generalidades



El modelo funciona bien en los datos de entrenamiento, pero no generaliza bien.

Siempre se puede construir un modelo que funcione perfecto con los datos de entrenamiento.

Métodos de aprendizaje supervisado



FACULTAD DE
INGENIERÍA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Aprendizaje supervisado

- Clasificación
 - K-NN
 - Árboles de decisión
 - Modelos lineales
 - Perceptrón
 - Regresión logística
 - SVM
 - Extensión a modelos no-lineales
 - Kernel SVM
 - Modelo lineal en un espacio de características de mayor dimensión
 - Kernel trick
- Regresión
 - Regresión lineal
 - Regresión polinómica
 - Árboles de decisión
 - SVM y Kernel SVM
 - Regresión lineal con regularización
 - Lasso, Ridge, ElasticNet

Ensamble de predictores

- Entrenamiento en paralelo, sobre un mismo conjunto de entrenamiento
 - Mezcla de expertos:
distintos tipos de predictor (SVM, k-NN, etc.) entrenados sobre los mismos datos
- Entrenamiento en paralelo, con distintos conjuntos de entrenamiento
 - Bagging (Bootstrap Aggregating):
distintos modelos del mismo tipo entrenados sobre distintos muestreos de los datos
 - Random forest:
combina bagging sobre árboles de decisión con selección aleatoria de atributos
- Entrenamiento secuencial o en cascada
 - Boosting: al pasar por la cascada de predictores, a las muestras difíciles de clasificar se les va aumentando el peso en el costo total, para que los predictores cuesta arriba se focalicen en su predicción. La decisión es un voto ponderado por el desempeño de cada predictor.

Training dataset

Obs	X_1	X_2	X_3	X_4	X_5	Y
1	X_{11}	X_{21}	X_{31}	X_{41}	X_{51}	0
2	X_{12}	X_{22}	X_{32}	X_{42}	X_{52}	1
3	X_{13}	X_{23}	X_{33}	X_{43}	X_{53}	0
4	X_{14}	X_{24}	X_{34}	X_{44}	X_{54}	0
5	X_{15}	X_{25}	X_{35}	X_{45}	X_{55}	1

Bootstrap

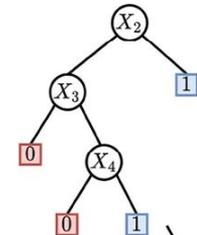
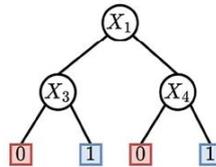
Obs	X_1	X_3	X_4	Y
1	X_{11}	X_{31}	X_{41}	0
2	X_{12}	X_{32}	X_{42}	1
5	X_{15}	X_{35}	X_{45}	1
1	X_{11}	X_{31}	X_{41}	0
5	X_{15}	X_{35}	X_{45}	1

Obs	X_2	X_3	X_4	Y
1	X_{21}	X_{31}	X_{41}	0
3	X_{23}	X_{33}	X_{43}	0
4	X_{24}	X_{34}	X_{44}	0
3	X_{23}	X_{33}	X_{43}	0
2	X_{22}	X_{32}	X_{42}	0

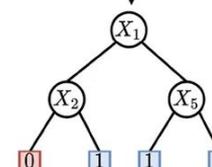
... ..

Obs	X_1	X_2	X_5	Y
2	X_{12}	X_{22}	X_{52}	1
3	X_{13}	X_{23}	X_{53}	0
5	X_{15}	X_{25}	X_{55}	1
5	X_{15}	X_{25}	X_{55}	1
3	X_{13}	X_{23}	X_{53}	0

Ensemble of trees



... ..



Aggregation



Métodos de aprendizaje no supervisado



FACULTAD DE
INGENIERÍA

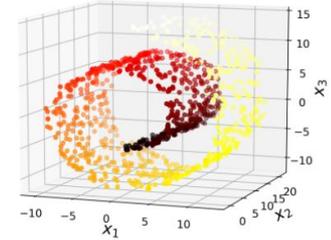
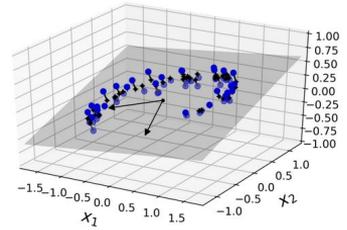


UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Aprendizaje no supervisado

- Reducción de dimensiones

- Maldición de la dimensionalidad / Bendición de la no uniformidad
- Objetivos:
 - Visualización
 - Extraer características relevantes.
 - Reducir costo computacional
- Proyección
 - PCA
 - Kernel PCA
- Aprendizaje de variedades
 - Locally linear embedding



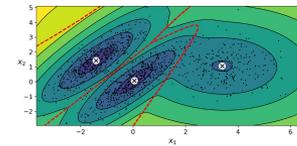
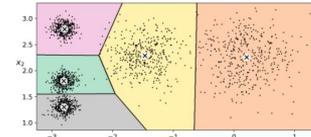
- Clustering

- K-Means
- DBSCAN

- Estimación de densidades

- Mezcla de gaussianas

- Detección de anomalías



Parámetros e hiperparámetros



FACULTAD DE
INGENIERÍA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Parámetros e hiperparámetros

- Parámetros
 - Se determinan durante el entrenamiento
 - Ejemplo:
 - Los pesos que determinan el hiperplano en SVM
- Hiper-parámetros
 - Se eligen
 - No cambian en el entrenamiento
 - Ejemplo:
 - valor de C en SVM
 - tipo de kernel y valor de gamma en Kernel SVM

Entrenamiento de modelos

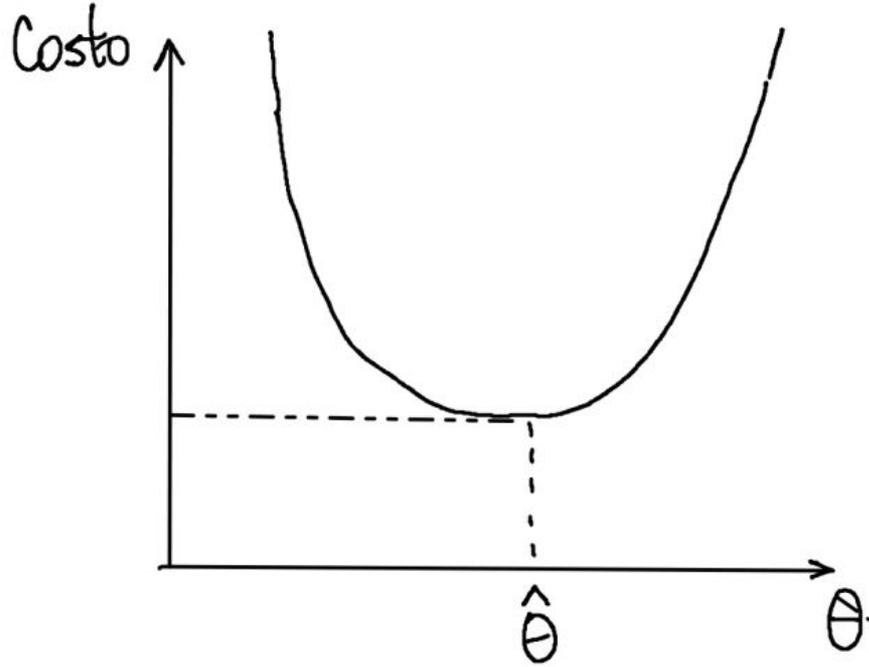


FACULTAD DE
INGENIERÍA



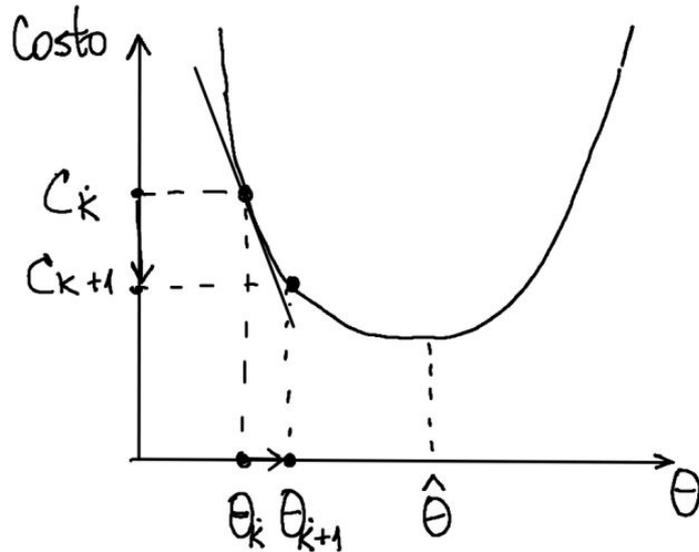
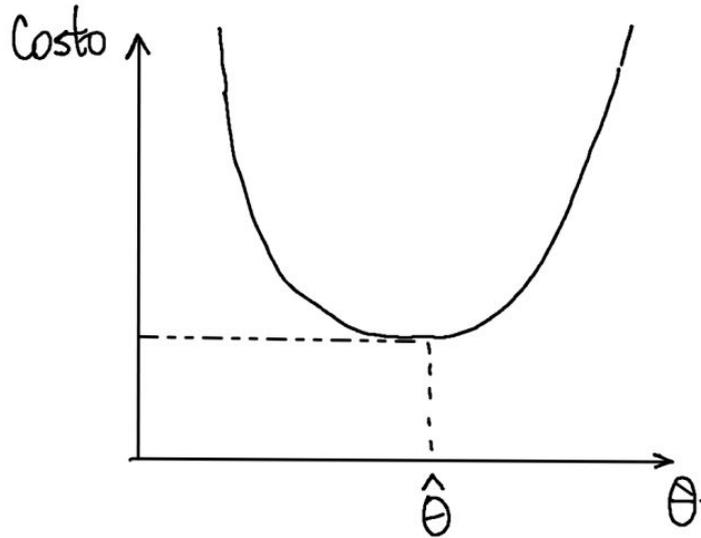
UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Parámetros se determinan en el entrenamiento



Gradient descent (GD)

- Algoritmo genérico de optimización
- Idea
 - Llegar al óptimo en forma iterativa
 - En cada paso:
 - Buscar la dirección de mayor crecimiento del costo y moverse en el sentido opuesto

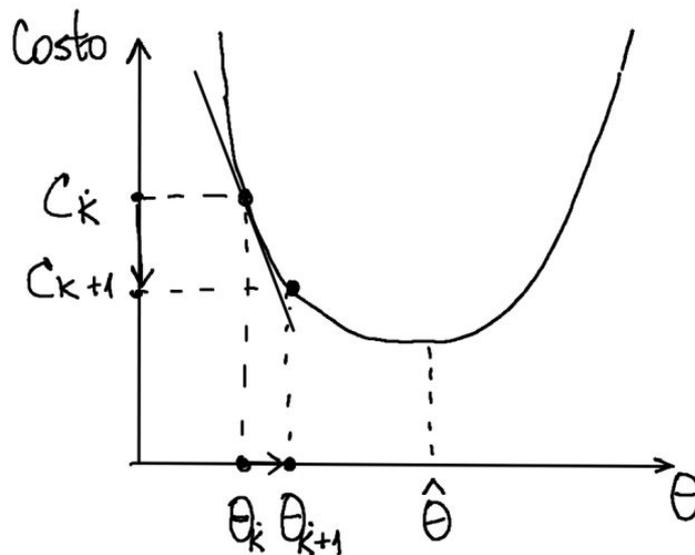


Gradient descent (GD)

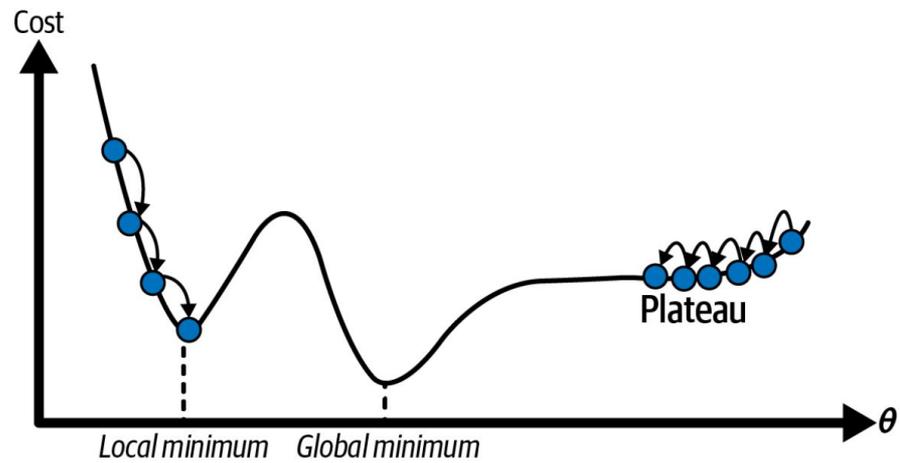
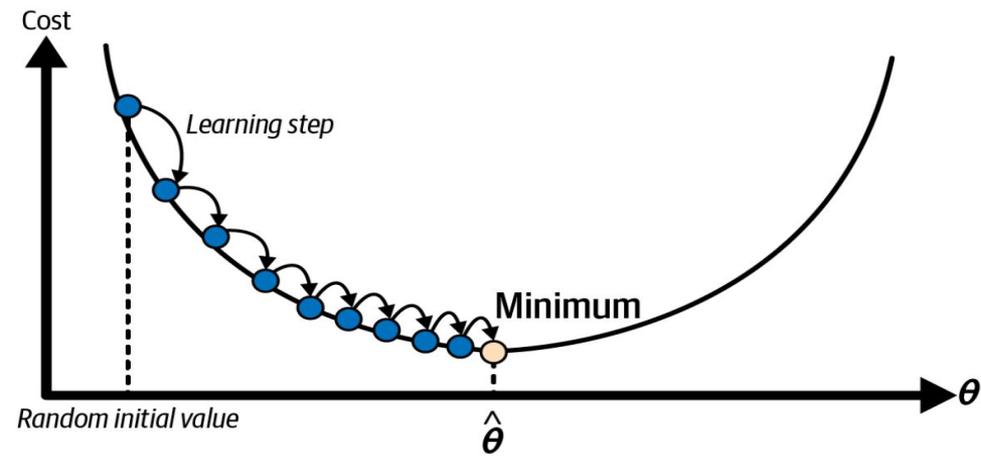
- Llegar al óptimo en forma iterativa
- En cada paso:
 - Buscar la dirección de mayor crecimiento del costo y moverse en el sentido opuesto

$$\nabla C(\theta) = \begin{bmatrix} \frac{\partial C}{\partial \theta_0} \\ \frac{\partial C}{\partial \theta_1} \\ \vdots \\ \frac{\partial C}{\partial \theta_n} \end{bmatrix}$$

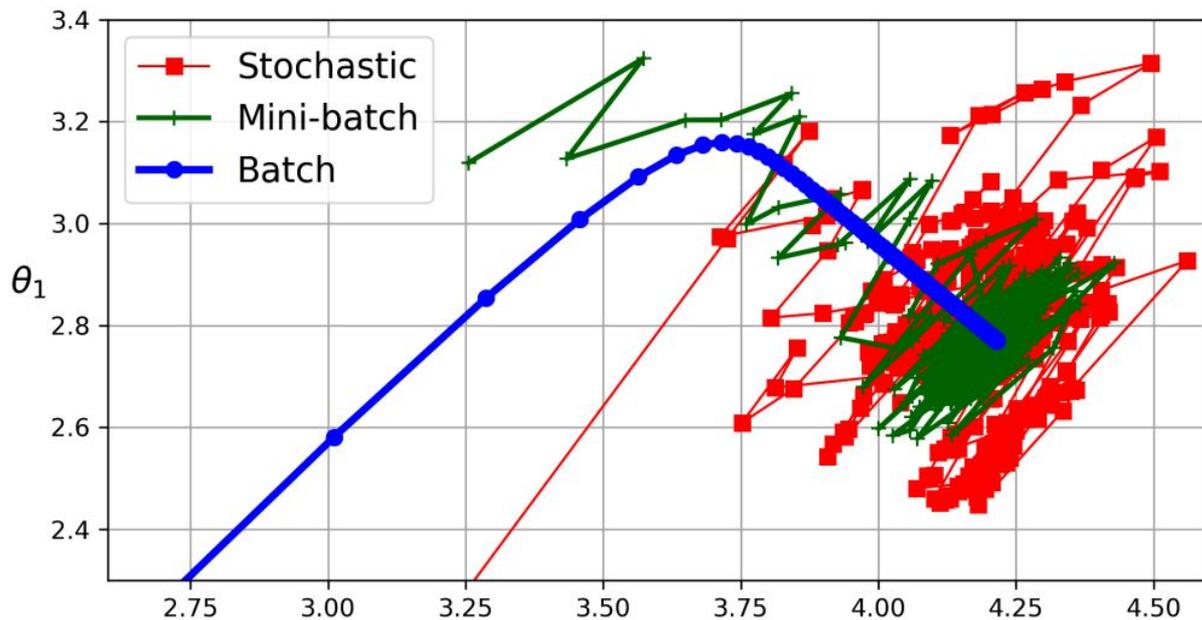
$$\theta^{k+1} = \theta^k - \eta \nabla_{\theta} C(\theta)$$



Gradient descent (GD)



Batch GD, Stochastic GD y Batch GD



Búsqueda de hiperparámetros y evaluación de modelos



FACULTAD DE
INGENIERÍA

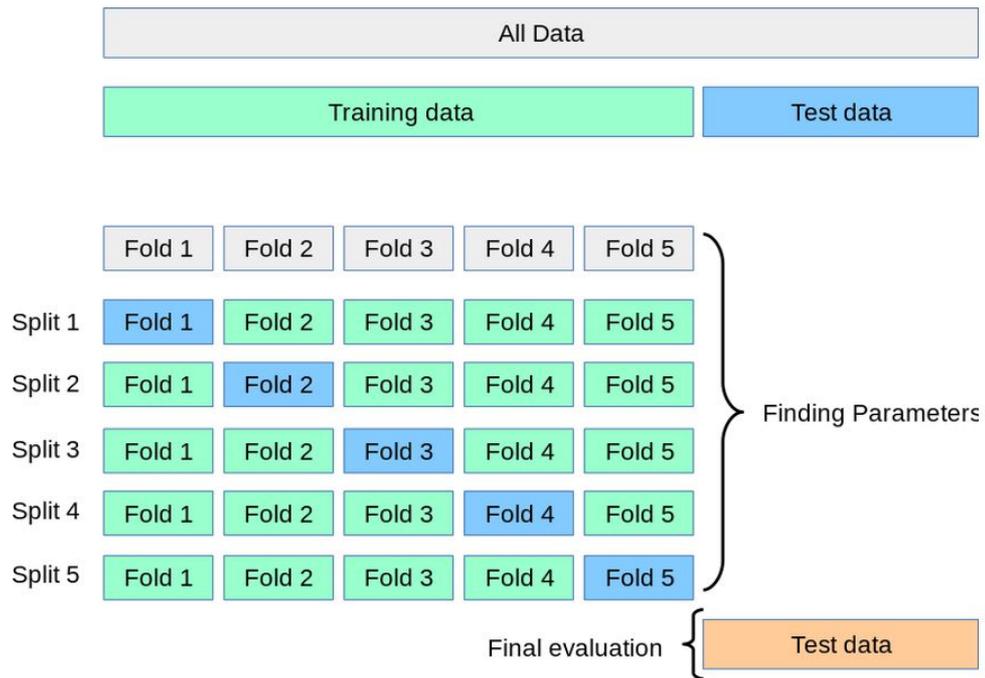
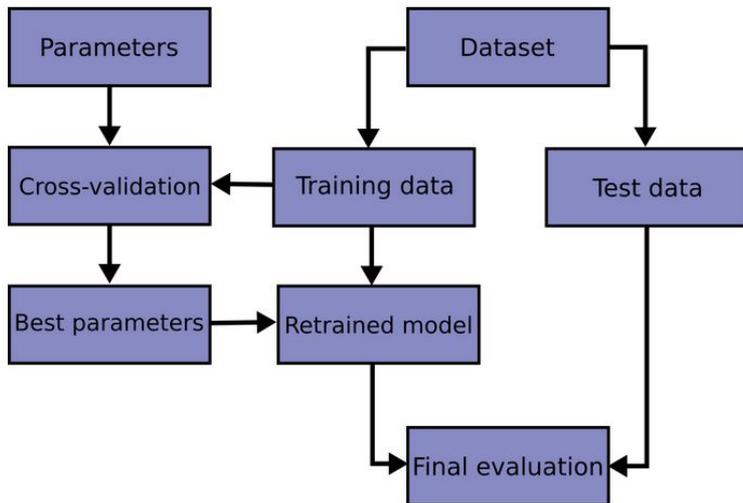


UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Búsqueda de mejores hiperparámetros y evaluación de modelos

- Dos tareas:
 - Elegir los mejores hiperparámetros
 - Evaluar el modelo
- No se debe realizar sobre los mismos datos
 - Riesgo de sobre-ajuste
- Partir los datos
 - Entrenamiento / Validación / Test
 - Validación cruzada

Búsqueda de mejores hiperparámetros y evaluación de modelos



Búsqueda de mejores hiperparámetros con validación cruzada

- Una búsqueda consta de:
 - Un estimador
 - Un espacio de parámetros donde buscar
 - Un método de búsqueda
 - Un esquema de validación cruzada
 - Una función de score
- **GrisSearchCV**
 - Búsqueda exhaustiva sobre una grilla de parámetros
- **RandomizedSearchCV**
 - Búsqueda en forma de muestreo aleatorio sobre el espacio de parámetros

Métricas de desempeño



FACULTAD DE
INGENIERÍA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Performance measurements

42 17 1 20 4

Total samples: $N = TP + FP + TN + FN$

Accuracy: $ACC = \frac{(TP + TN)}{N}$

¿cuánto se acerca a los valores reales? exactitud

Sensitivity/Recall: $TPR = \frac{TP}{(TP + FN)}$

tasa de enfermos correctamente detectados

Specificity: $TNR = \frac{TN}{(TN + FP)}$

tasa de sanos correctamente clasificados

Precision: $PPV = \frac{TP}{(TP + FP)}$

de los clasificados positivos ¿cuántos son los realmente positivos?

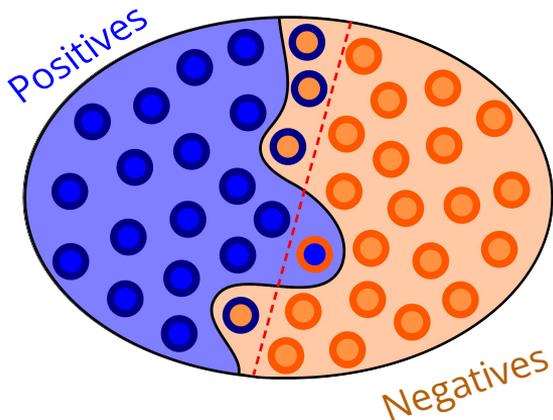
False positive rate: $FPR = \frac{FP}{(FP + TN)}$

probabilidad de una falsa alarma.

False discovery rate: $FDR = \frac{FP}{(FP + TP)}$

F-score: $F = \frac{2TP}{(2TP + FP + FN)}$

Media armónica entre TPR y PPV.



 TP: true positives
success

 TN: true negatives
correct rejection

 FP: false positives
type I error (false alarm)

 FN: false negatives
type II error

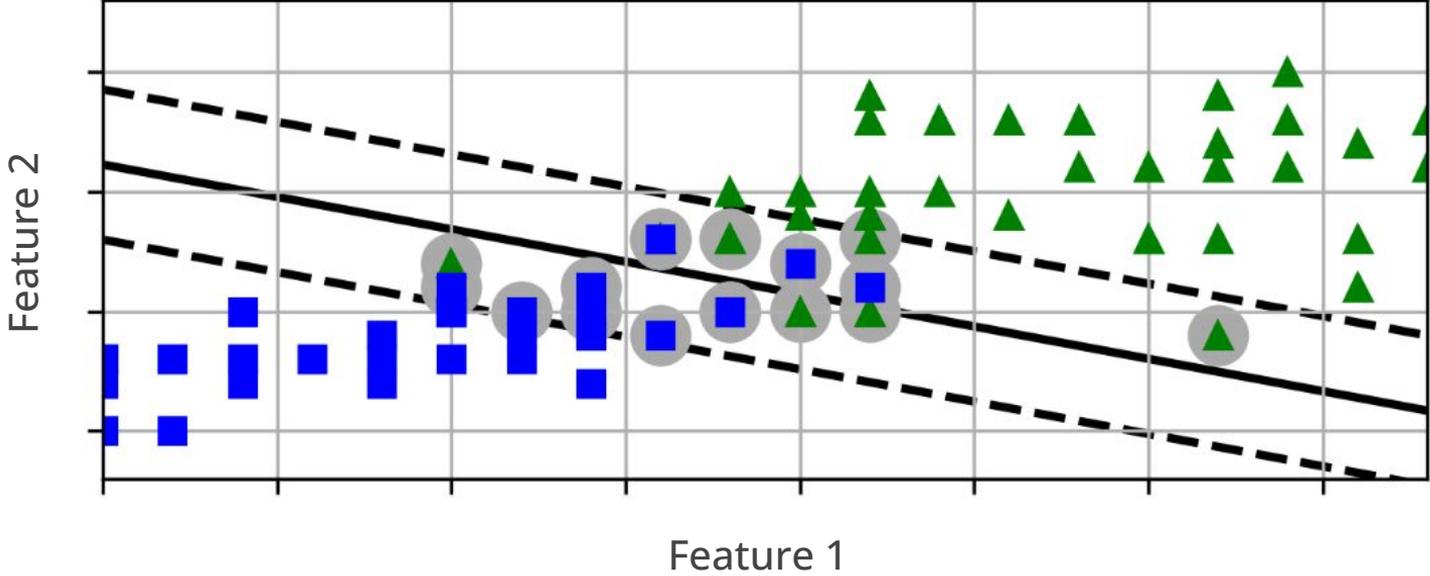
system login

disease detection

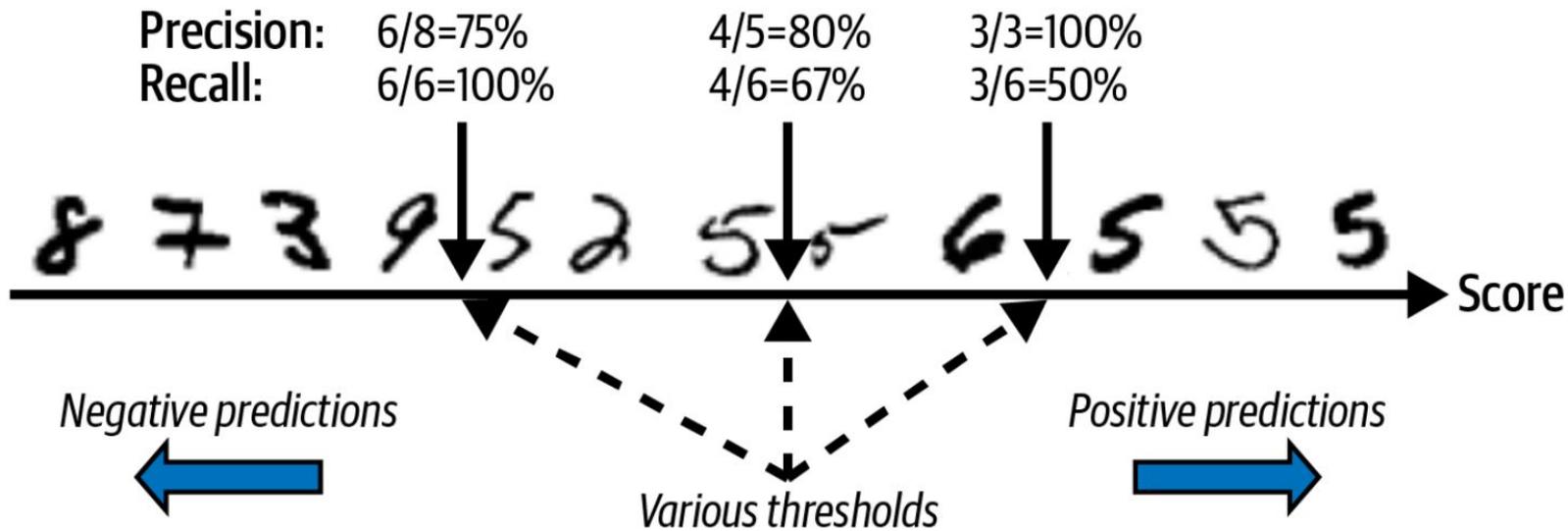
Confusion matrix

		Positive (sick)	Negative (healthy)		
		TP	FN		
Real labels	Positive (sick)	TP	FN		
	Negative (healthy)	FP	TN		
		Predicted labels			

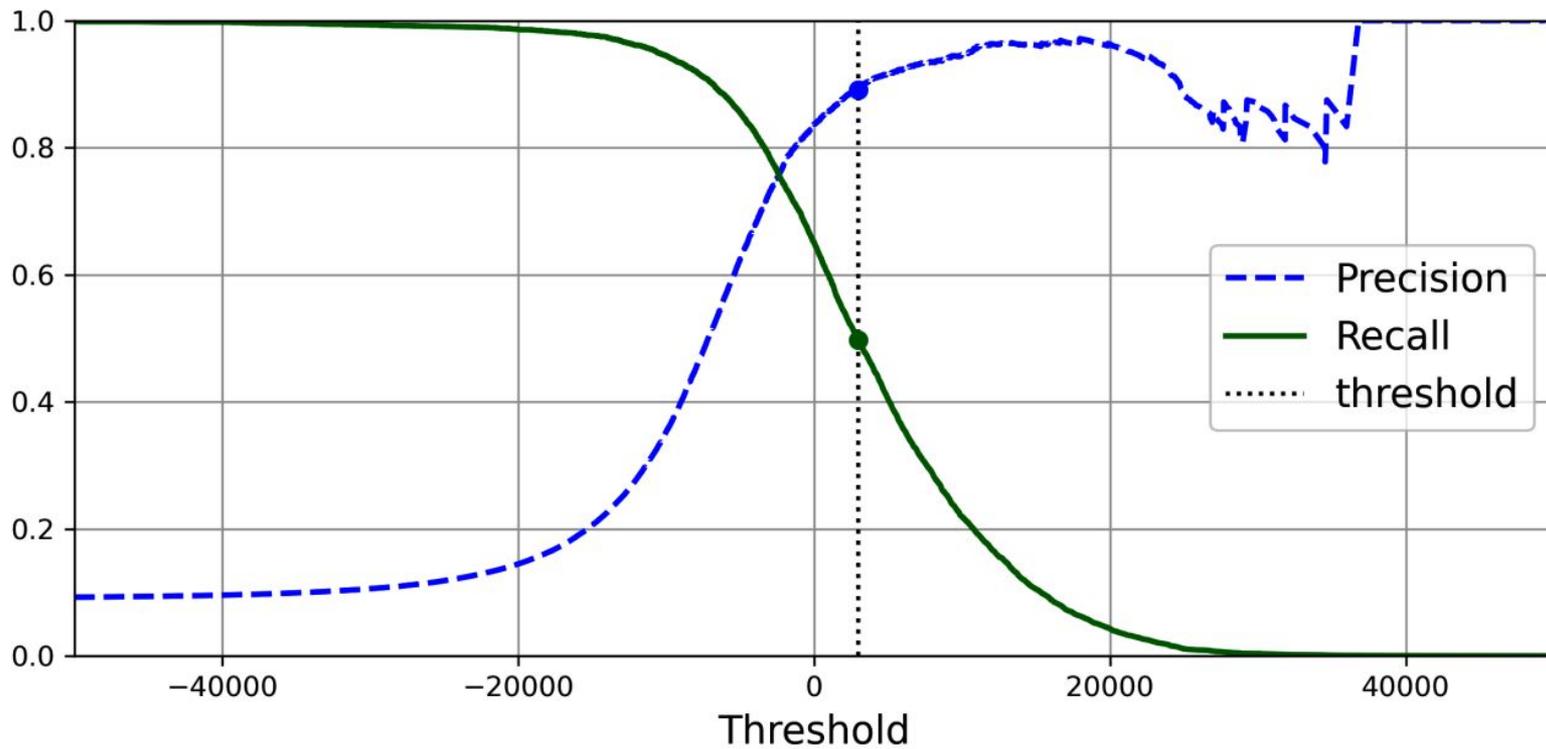
Score de confianza. Distancia al hiperplano



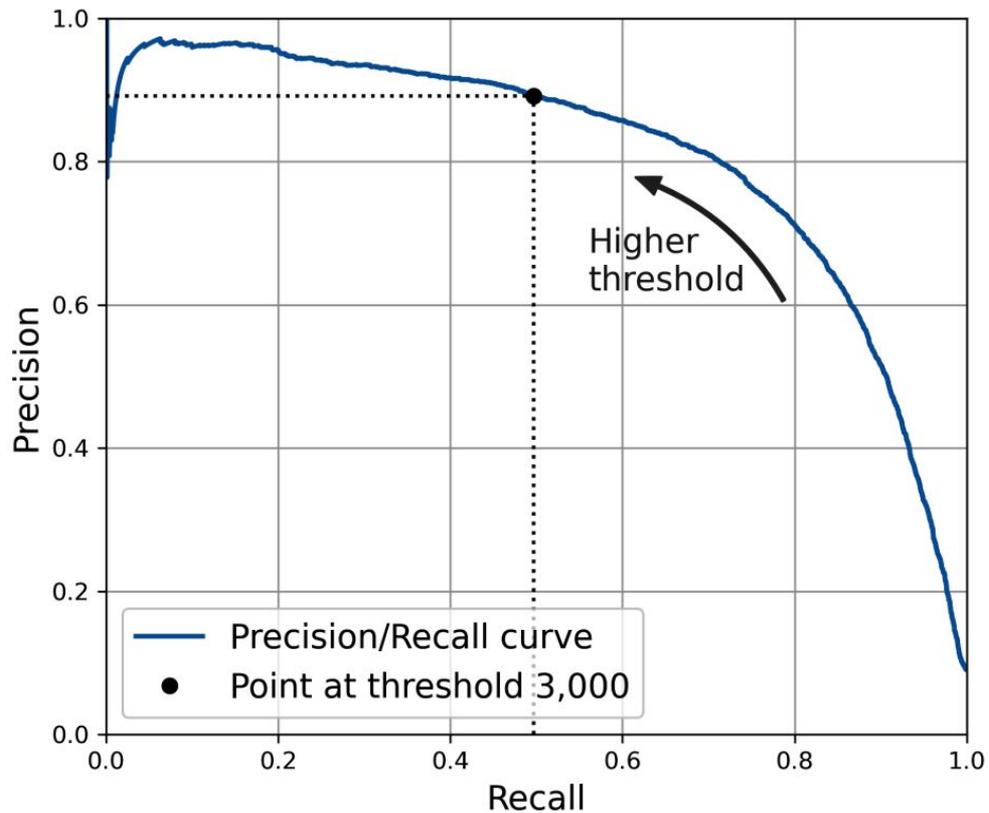
Compromiso precision/recall



Compromiso precision/recall

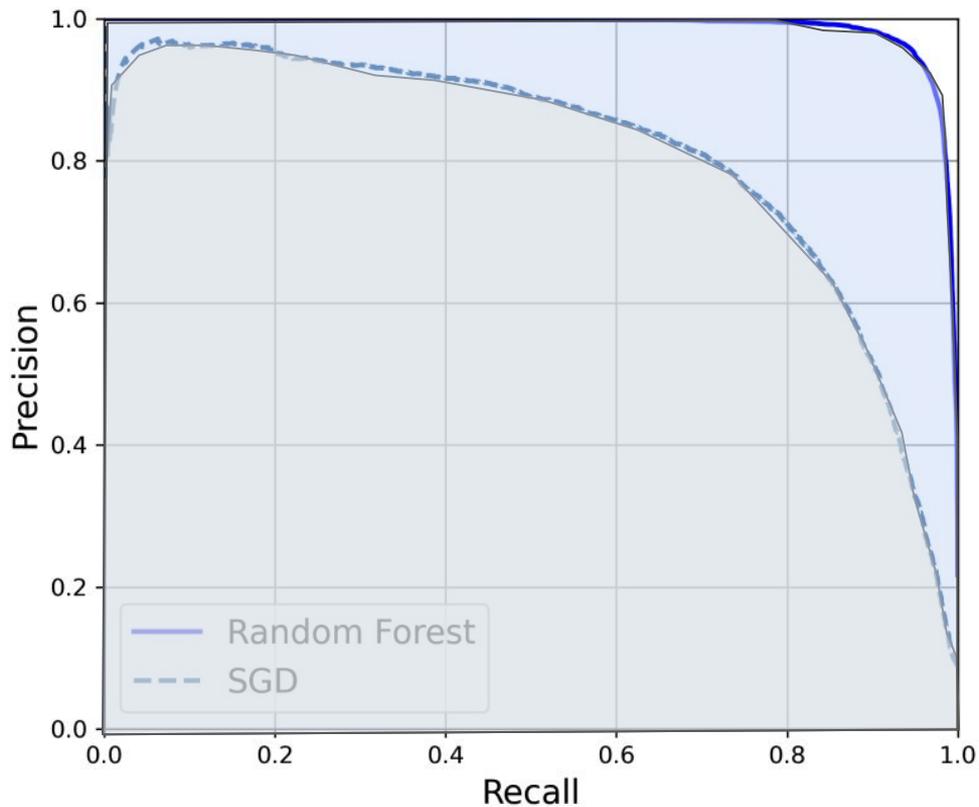


Curva precision vs recall



Curva precision vs recall. Comparación de algoritmos

PR-AUC
Area under curve



Redes neuronales



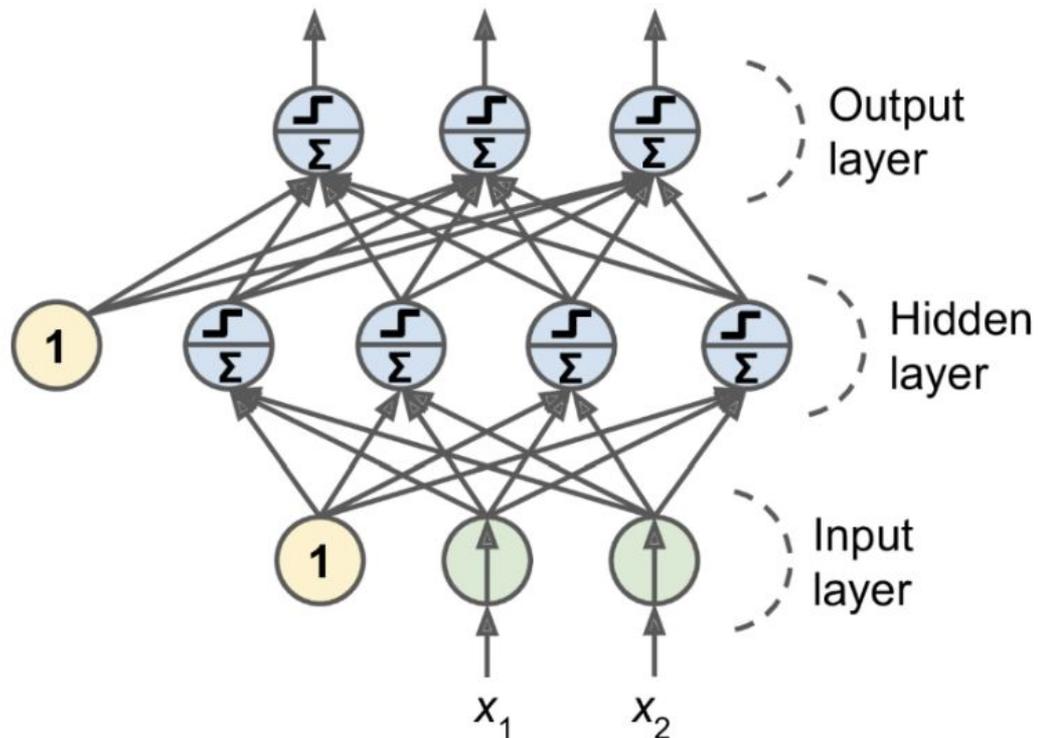
FACULTAD DE
INGENIERÍA



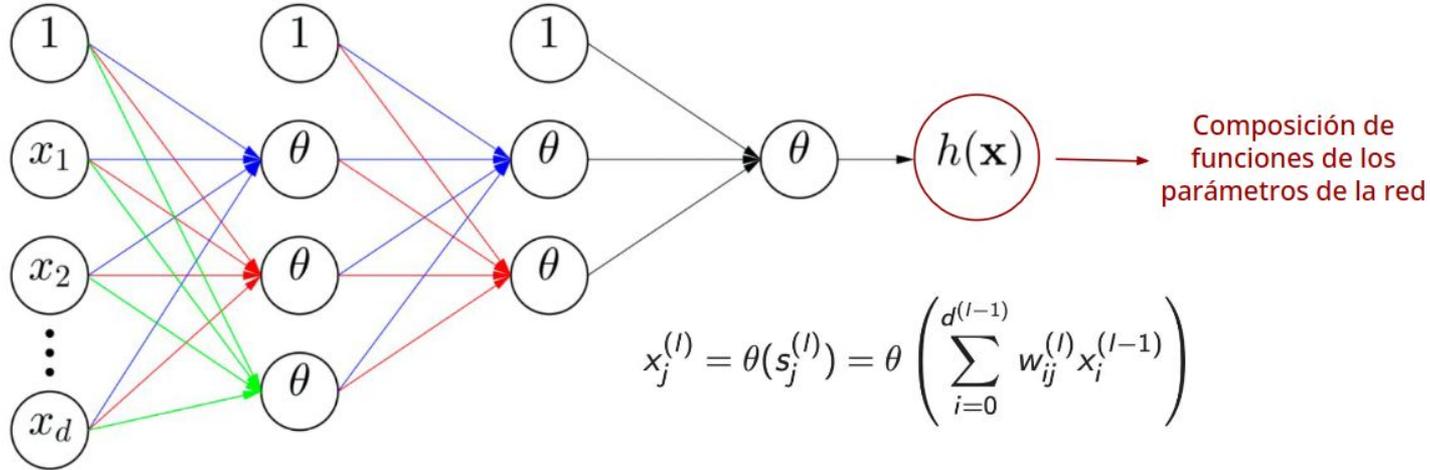
UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Perceptron multicapa

- Varias capas de perceptrones
- Una capa de entrada
- Una o varias capas ocultas (hidden layers)
- Una capa de salida
- Todas menos la capa de salida tiene una neurona para el bias
- Todas son capas densas



Entrenamiento - Backpropagation



$$e(w): \text{función error de la red}$$
$$\nabla e(w) = \frac{\partial e(w)}{\partial w_{ij}^{(l)}} \quad \forall i,j,l$$
$$w_{ij}^{(l)} = w_{ij}^{(l)} - \eta \frac{\partial e(w)}{\partial w_{ij}^{(l)}} \quad \forall i,j,l$$

Regla de la cadena

$$\left. \begin{array}{l} y = f(u) = f(g(x)) \\ \frac{dy}{dx} = \frac{du}{dx} \frac{dy}{du} \end{array} \right\}$$

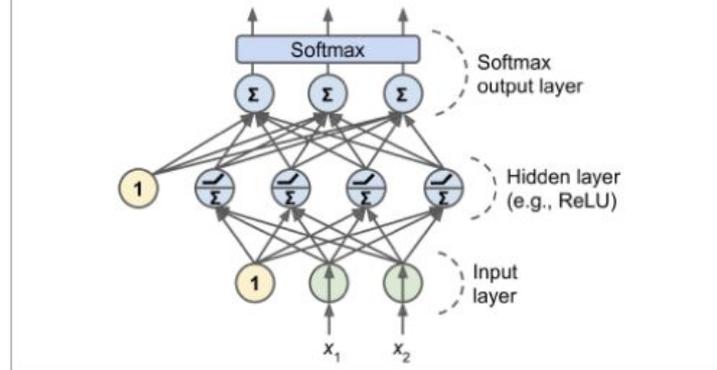
Regresión

- Los MLP pueden utilizarse para regresión
- Determinar si se quiere una o varias salidas
- En general la última capa de neuronas no tiene función de activación
- Función de error típica: MSE, MAE, Huber

Hyperparameter	Typical value
# input neurons	One per input feature (e.g., $28 \times 28 = 784$ for MNIST)
# hidden layers	Depends on the problem, but typically 1 to 5
# neurons per hidden layer	Depends on the problem, but typically 10 to 100
# output neurons	1 per prediction dimension
Hidden activation	ReLU (or SELU, see Chapter 11)
Output activation	None, or ReLU/softplus (if positive outputs) or logistic/tanh (if bounded outputs)
Loss function	MSE or MAE/Huber (if outliers)

Clasificación

- Los MLP pueden utilizarse para clasificación
- En la capa de salida, tantas neuronas como clases
- Clasificación binaria: sigmoid
- Clasificación más de dos clases: softmax
- Salida de la red: probabilidad de pertenencia a cada clase
- Función de error típica: entropía cruzada



Hyperparameter	Binary classification	Multilabel binary classification	Multiclass classification
Input and hidden layers	Same as regression	Same as regression	Same as regression
# output neurons	1	1 per label	1 per class
Output layer activation	Logistic	Logistic	Softmax
Loss function	Cross entropy	Cross entropy	Cross entropy

Redes profundas



FACULTAD DE
INGENIERÍA



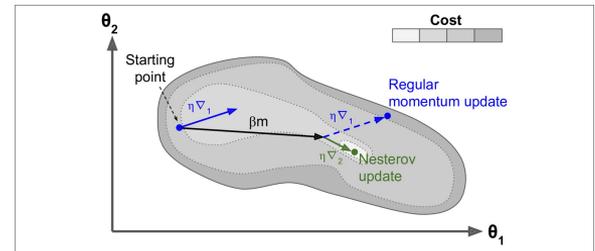
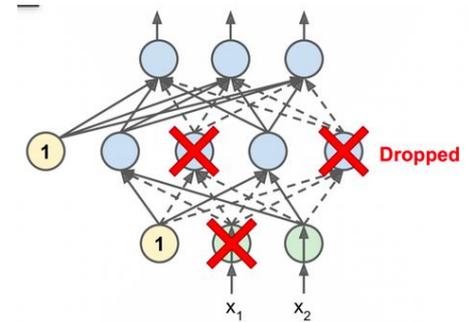
UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Redes profundas - Desafíos

- Desvanecimiento y explosión de gradiente
- Generalización. Evitar el sobreajuste
- Entrenamiento costoso. Acelerar el GD

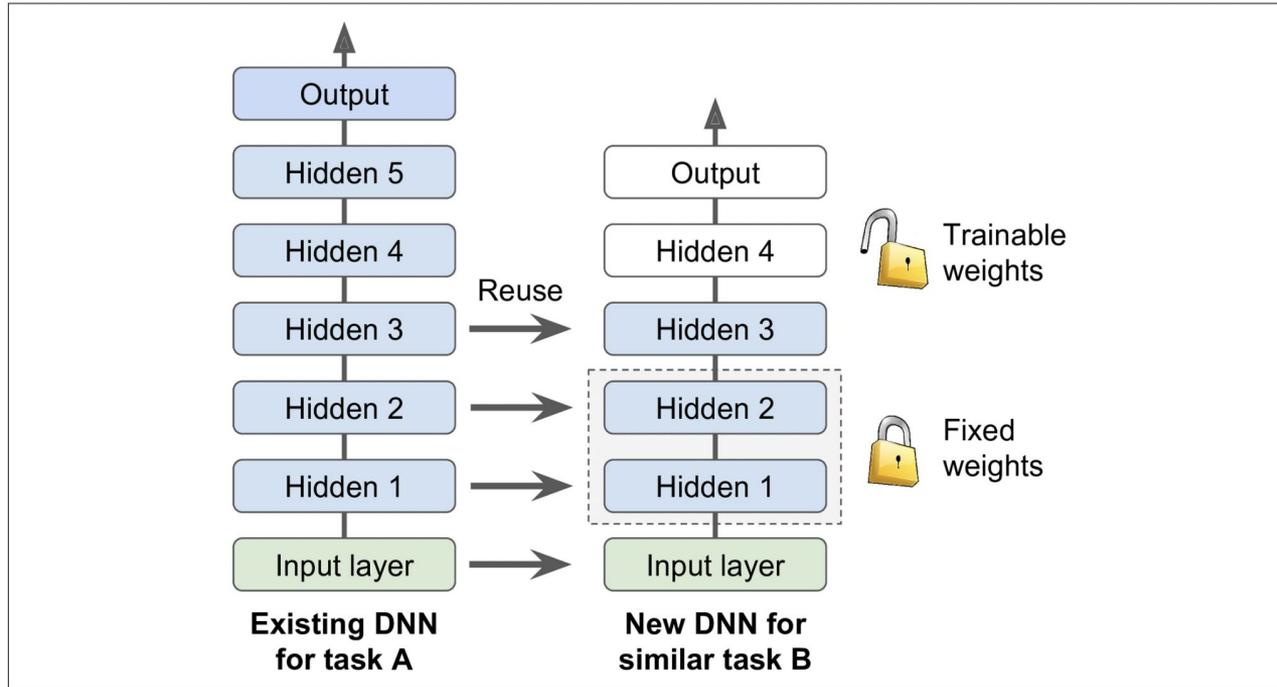
Redes profundas - Desafíos

- Desvanecimiento y explosión de gradiente
 - Inicializaciones de Glorot, He o Lecun
 - Funciones de activación que no saturan → ReLU → Leaky ReLU → ELU
 - Batch normalization
 - Gradient clipping
 - Aprendizaje por transferencia
- Generalización. Evitar el sobreajuste
 - Batch normalization
 - Dropout
 - Early-stopping
- Entrenamiento costoso. Acelerar el GD
 - Optimizadores más rápidos que el GD regular (Momentum, Nesterov, AdaGrad, RMSProp, Adam)
 - Learning rate scheduling



Aprendizaje por transferencia

- En vez de entrenar de zero una DNN, re usar una DNN ya entrenada para una tarea similar y re usar algunas de sus capas
- Acelera la fase entrenamiento y requiere menos datos



Redes neuronales convolucionales (CNN)

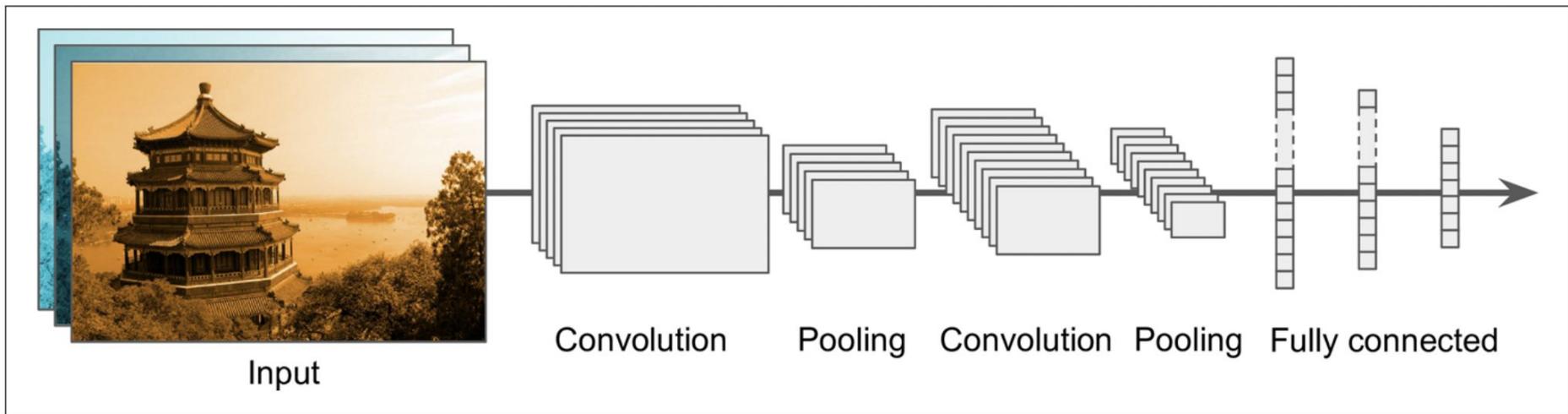


FACULTAD DE
INGENIERÍA

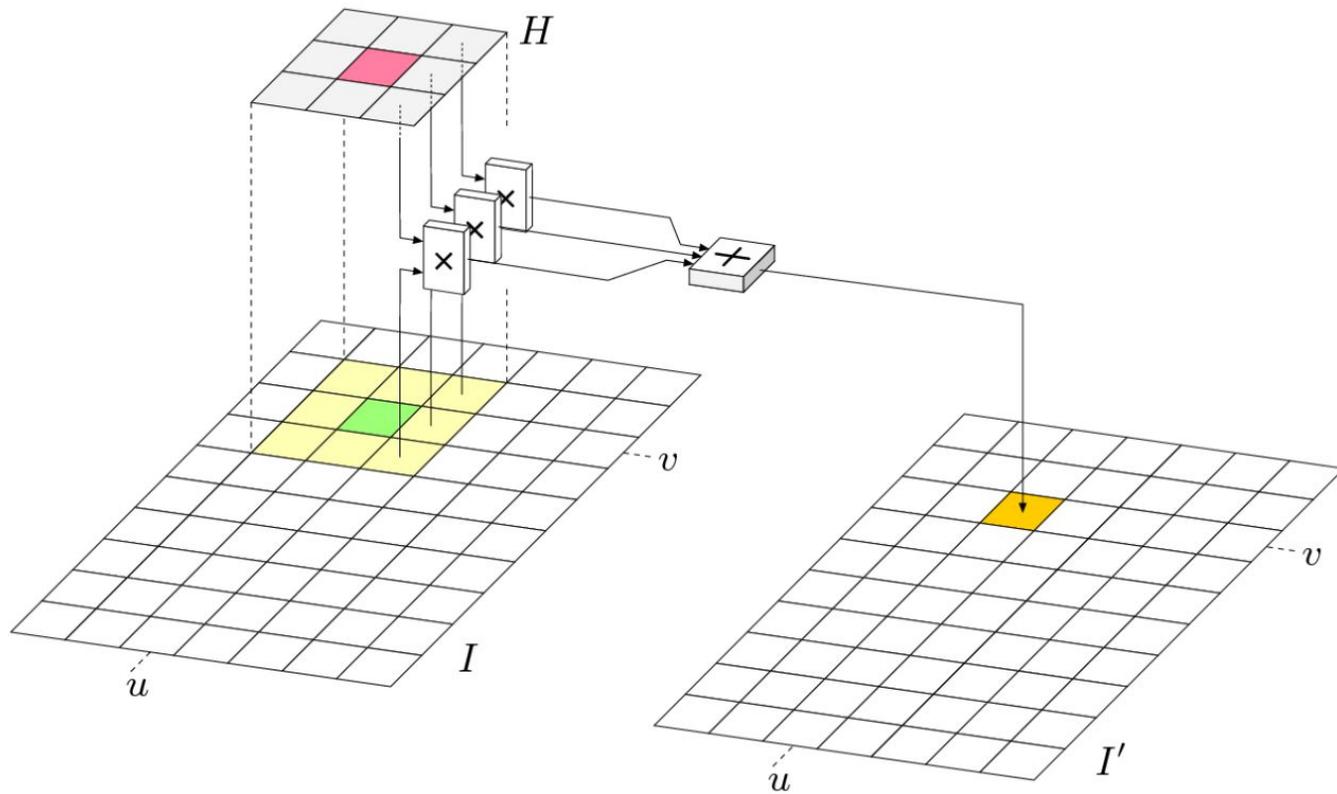


UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

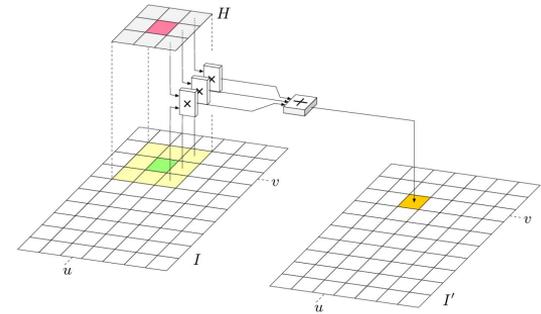
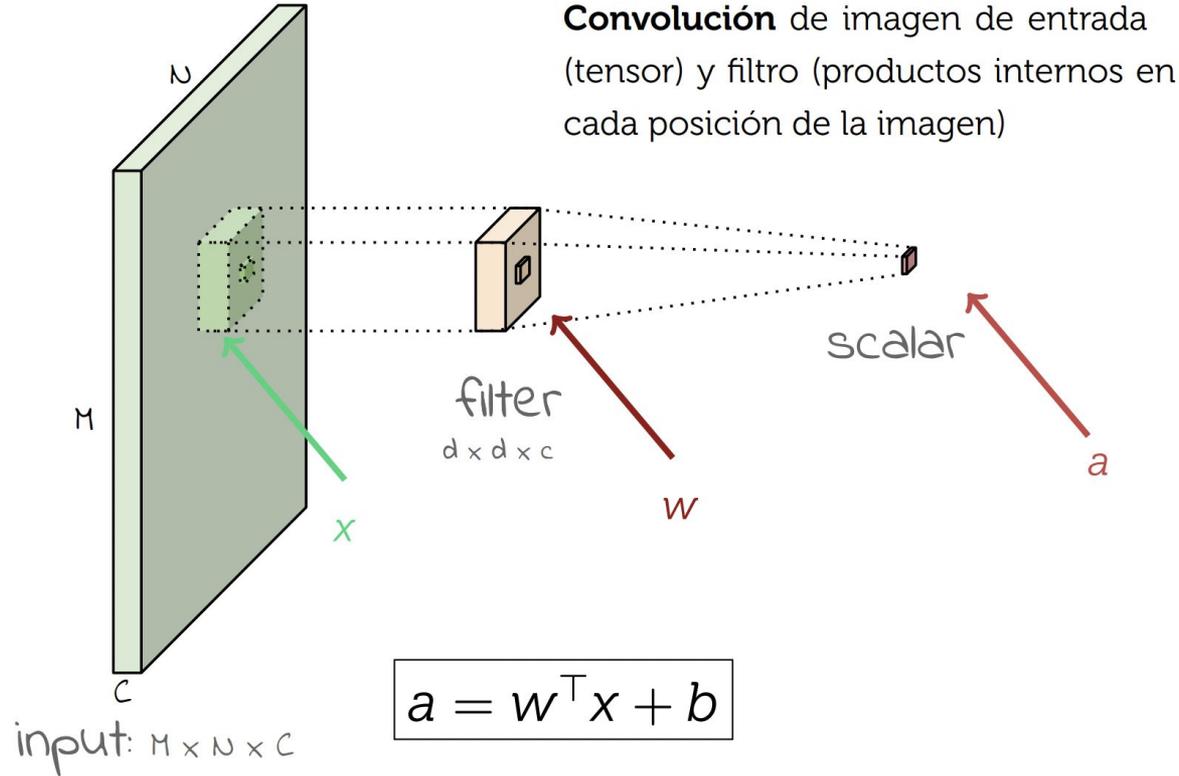
Ejemplo de CNN típica



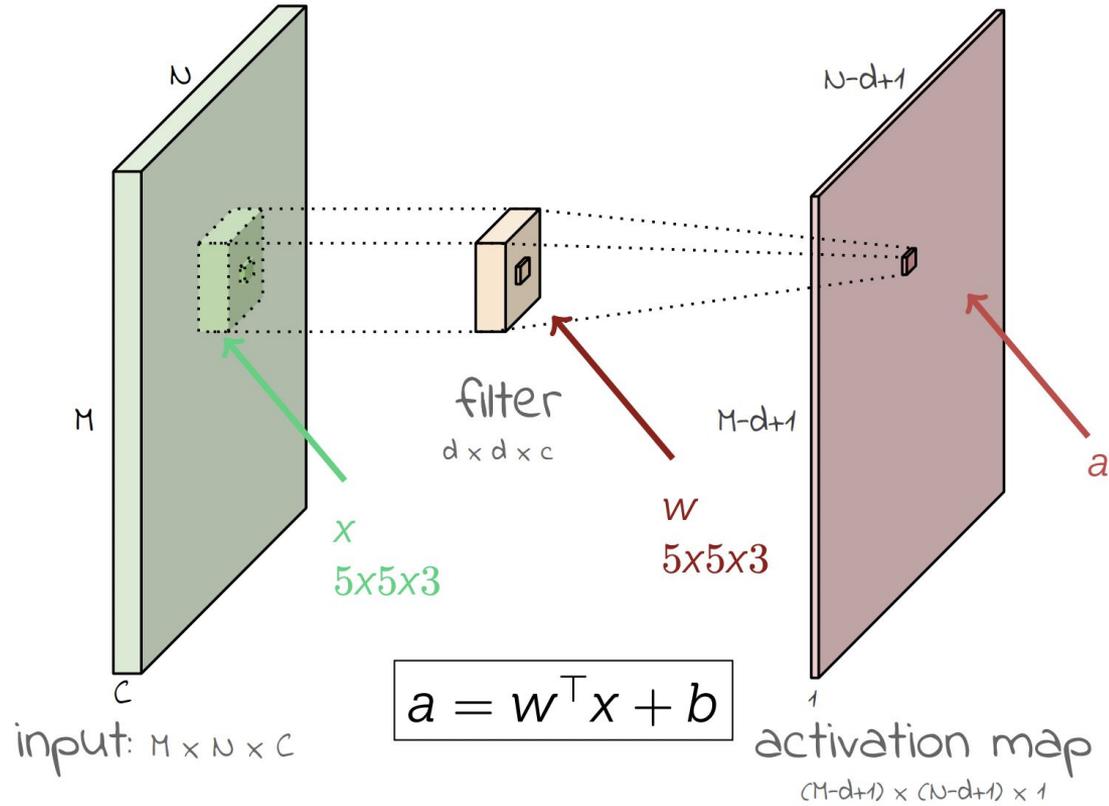
Convolución



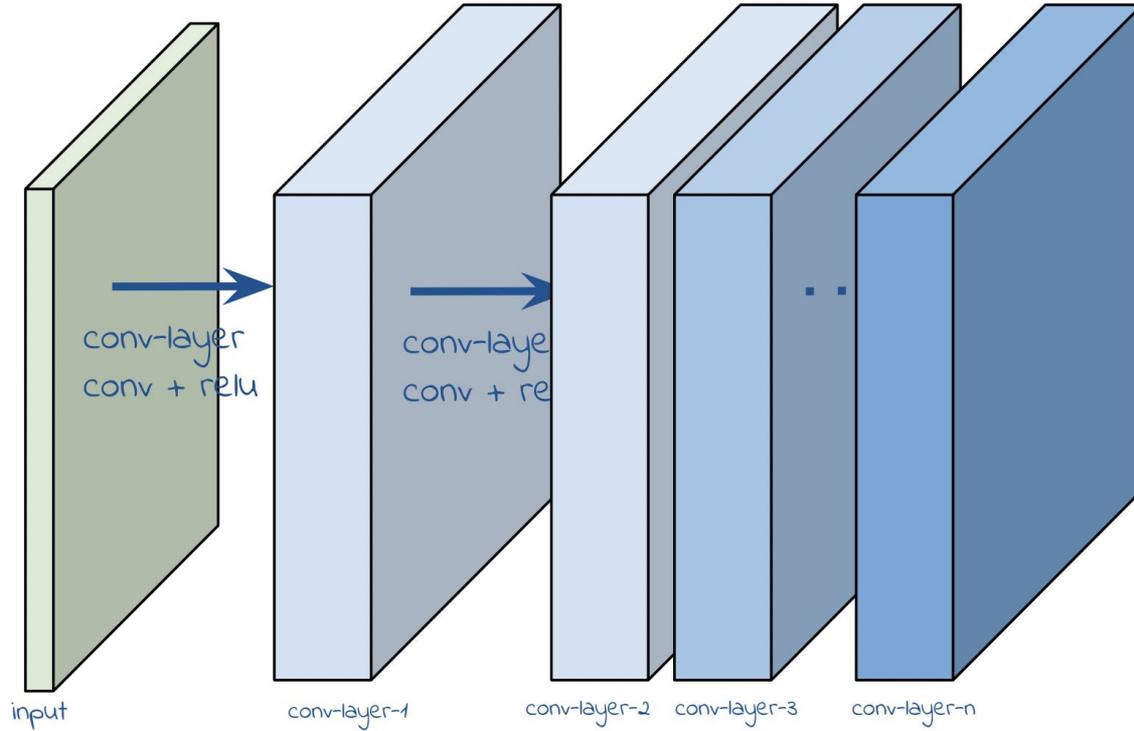
Capa de convolución



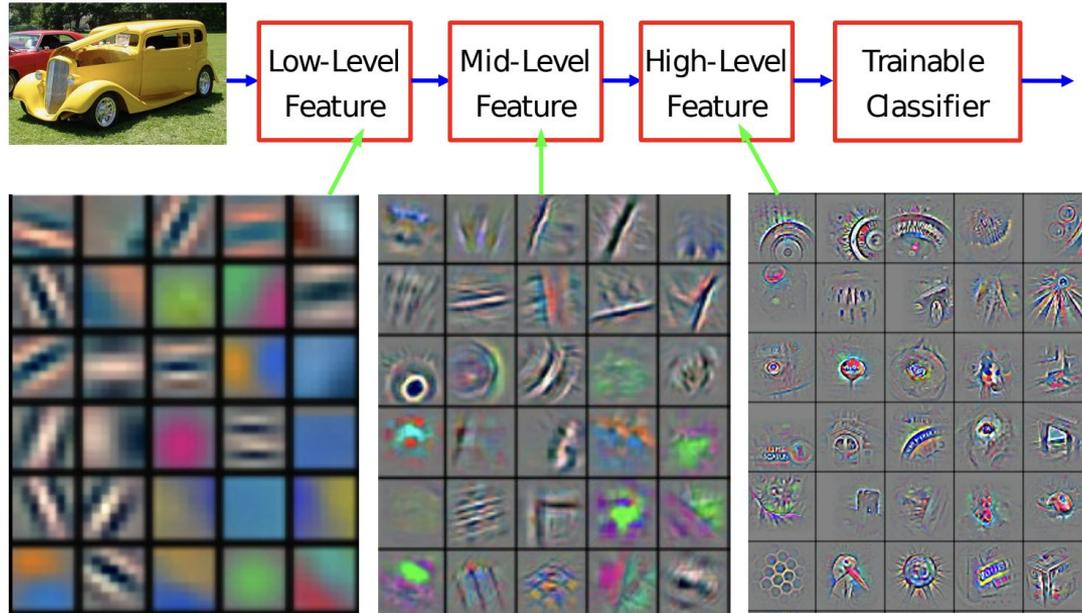
Capa de convolución



Capa de convolución: convolución + activación

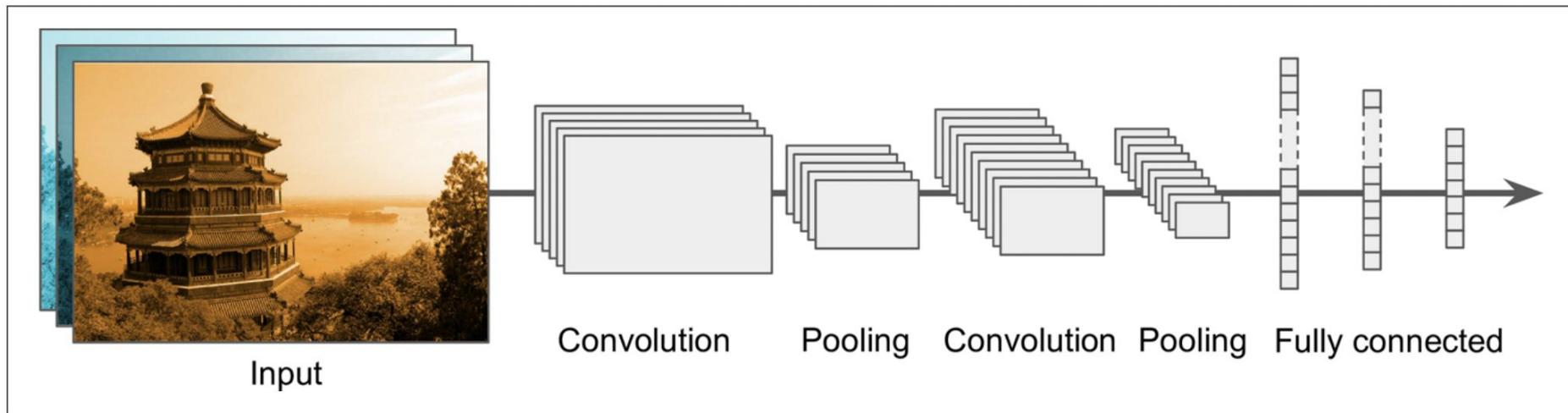


Visualización de los *feature maps*



Feature visualization of convolutional net trained on ImageNet from [Zeiler & Fergus 2013]

Ejemplo de CNN típica



Automatización de procesos de AA

AutoML y MLOps



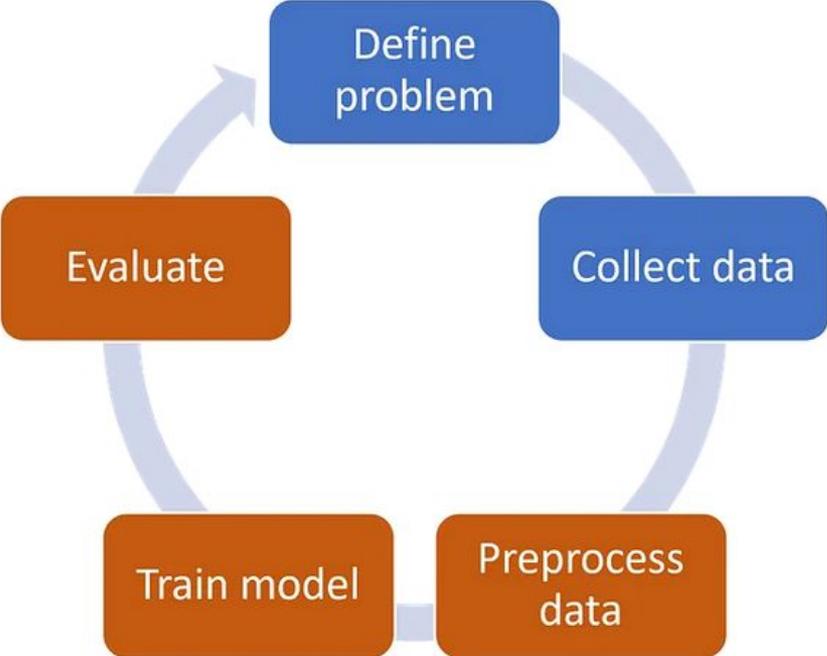
FACULTAD DE
INGENIERÍA



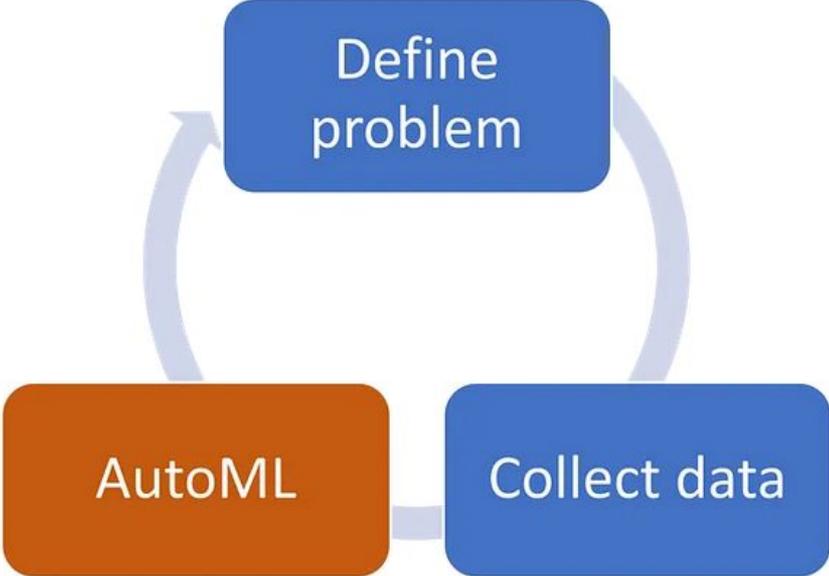
UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

AutoML vs Manual ML

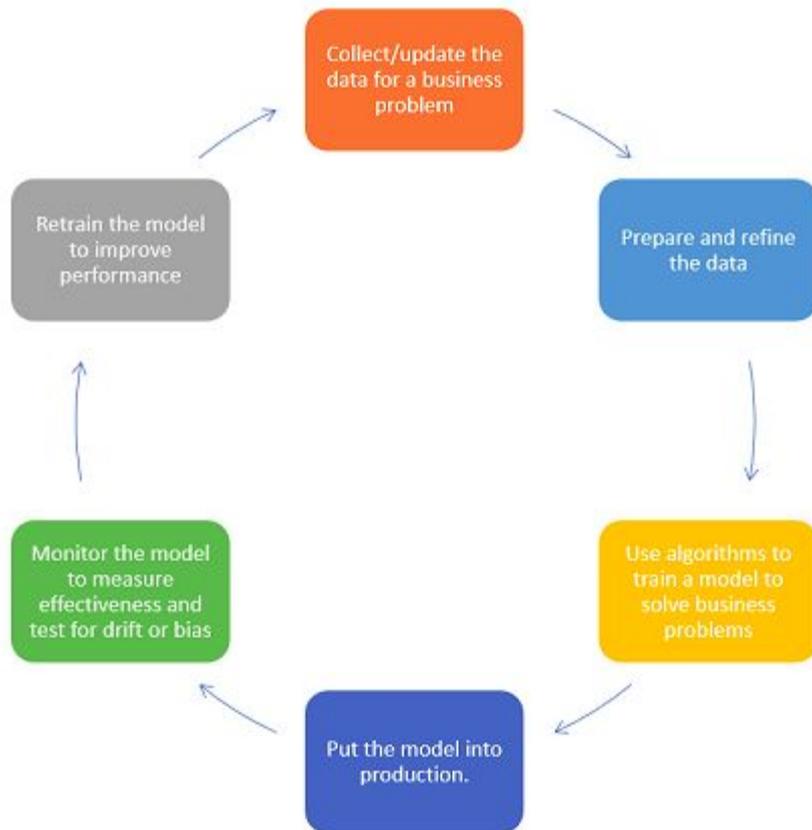
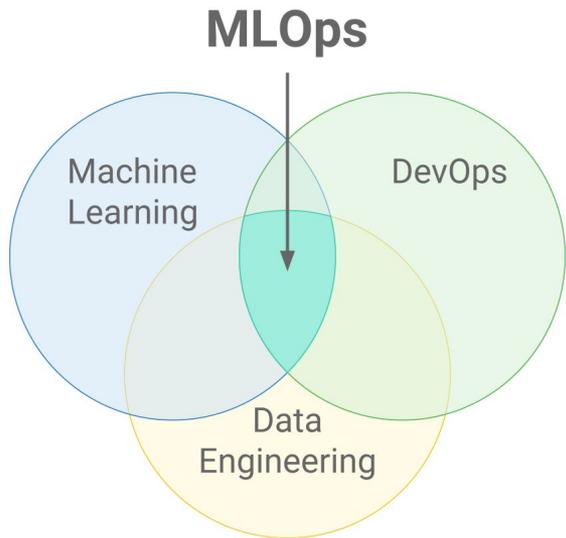
Traditional ML training workflow



AutoML workflow



MLOps



MLOps (Machine Learning Operations)

Integración de la ciencia de datos y el desarrollo de modelos de machine learning en los procesos operativos y de desarrollo de software

Algunos componentes:

- **Automatización del flujo de trabajo**
 - ciclo de vida completo desde la adquisición y preparación de datos hasta el despliegue y monitorización en producción.
- **Gestión de versiones y control de código**
 - garantizar la reproducibilidad
- **Despliegue de modelos en entornos de producción**
 - gestión de dependencias, escalabilidad, monitoreo continuo del rendimiento
- **Infraestructura escalable y reproducible**
 - servidores, contenedores y orquestadores de contenedores
- **Monitorización y mantenimiento**
 - rendimiento de los modelos en producción, detección de problemas, ajustes o actualizaciones a los modelos
- **Gobierno y cumplimiento normativo**
 - gobierno de datos y aseguramiento de la calidad, garantizar la transparencia, la ética y el cumplimiento normativo

¿Cómo se puede seguir?

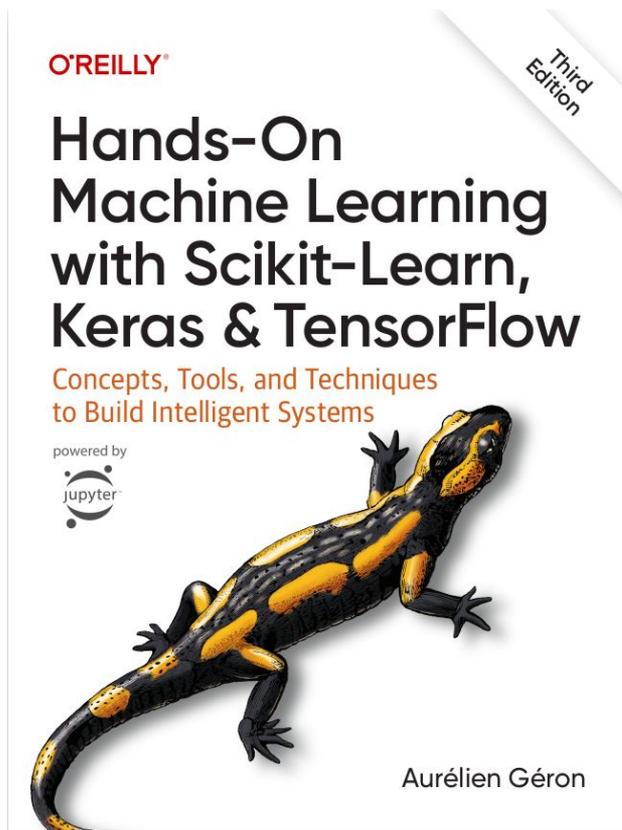


FACULTAD DE
INGENIERÍA

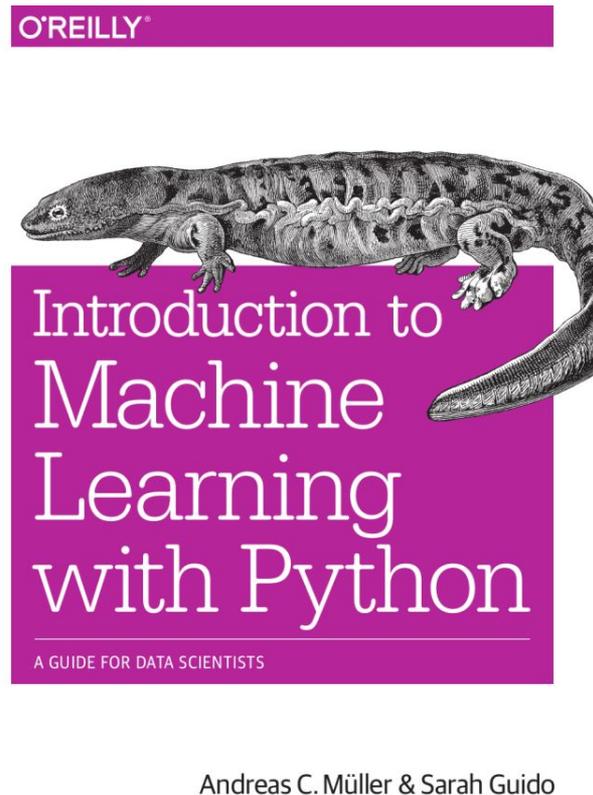


UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY

Bibliografía con código



<https://github.com/ageron/handson-ml3>



<https://github.com/amueller/introduction to ml with python>

Otros cursos sobre los que nos apoyamos

- **Fundamentos de Aprendizaje Automático**

Curso de grado de Ingeniería Eléctrica

- Material: <https://eva.fing.edu.uy/course/view.php?id=1252>
- Videos: <https://open.fing.edu.uy/courses/fuaa/>

- **Taller de Aprendizaje Automático**

Curso de grado de Ingeniería Eléctrica

- Material: <https://eva.fing.edu.uy/course/view.php?id=1492>
- Videos: <https://open.fing.edu.uy/courses/taa/>

Otros cursos y posgrados

- **DLVis (Deep Learning para Visión Artificial)**
Dictado como curso de posgrado y actualización
 - Material: <https://eva.fing.edu.uy/course/view.php?id=1046>
 - Clases: <https://www.youtube.com/@DLVISFing>
- **Aprendizaje Automático para Datos en Grafos**
Dictado como curso de grado/posgrado/educación permanente
 - Material: <https://eva.fing.edu.uy/course/view.php?id=1626>
- **Posgrados en Ingeniería Eléctrica (maestría y doctorado)**
 - <https://iie.fing.edu.uy/ensenanza/posgrado/carreras/>
- **Maestría en Ciencia de Datos y Aprendizaje Automático**
 - <https://www.fing.edu.uy/maestria-ciencia-de-datos/>

Difusión

- **Two Minute Papers**

- DeepFake Detector

<https://www.youtube.com/watch?v=RoGHVI-w9bE>

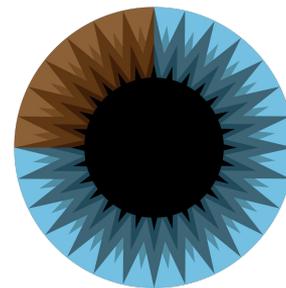
- **Dot-CSV**

- Whisper

<https://www.youtube.com/watch?v=JuMEmF-2FsA&t>

- **3Blue1Brown**

- Backpropagation: <https://www.youtube.com/watch?v=llg3gGewQ5U>
- GPT: <https://www.youtube.com/watch?v=wjZofjX0v4M&t>



Difusión

- **Papers with Code**

- <https://paperswithcode.com/>



- **Competencias Kaggle (con código)**

- <https://www.kaggle.com/datasets/tejasreddy/iam-handwriting-top50/data>

- **Huggingface**

- <https://huggingface.co>



kaggle



FACULTAD DE
INGENIERÍA



UNIVERSIDAD
DE LA REPÚBLICA
URUGUAY