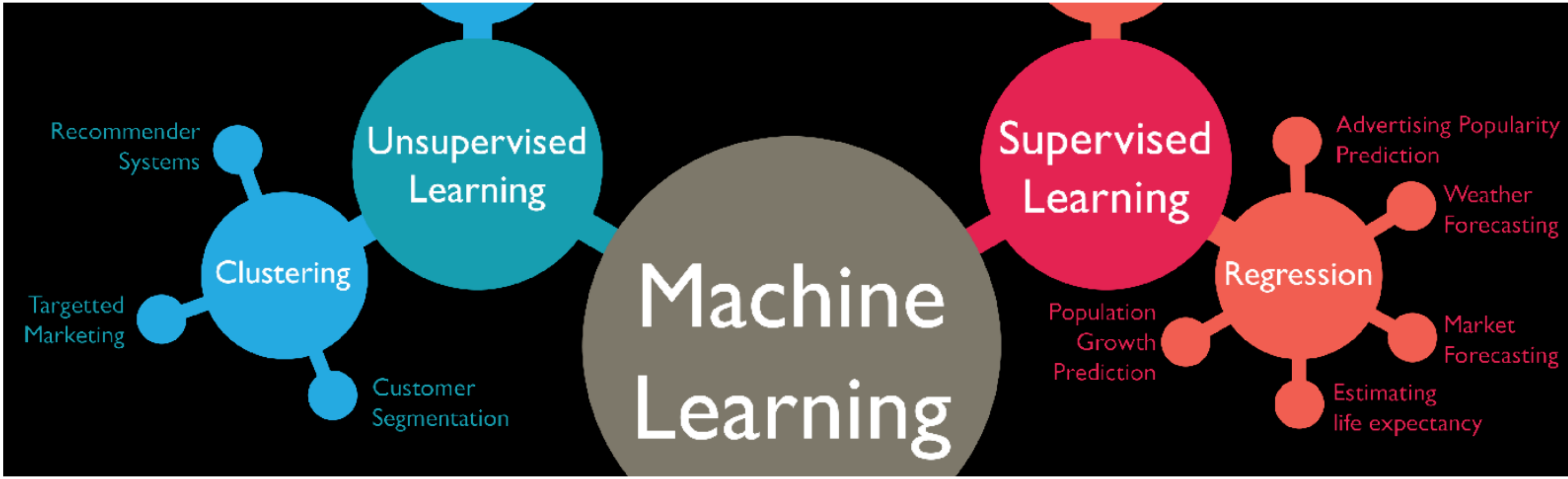


Aprendizaje Automático

Grado en Ingeniería Informática

Computación y Sistemas Inteligentes

Profesores de la asignatura

- Teoría :

Nicolás Pérez de la Blanca Capilla

  - D.5 , Dpto. CCIA, 4ª planta, ETSIIT
  - Correo: a través de PRADO ,
  - Tutorías: Miércoles 9-13 y Jueves de 9-11

- Prácticas:

Pablo Mesejo Santiago (Grupos 1 y 2)

  - Edificio Auxiliar ETSIIT (antiguo FOREM),
  - C/ Periodista Juan Osorio Bueno, 18014
  - (concertar cita por correo)
  - Correo: : pmesejo@decsai.ugr.es
  - Tutorías: Martes y Miércoles de 10 a 13 h
  - Despacho 1.10

Jesús Giráldez(Grupo 3)

  - Edificio Auxiliar ETSIIT (antiguo FOREM),
  - C/ Periodista Juan Osorio Bueno, 18014
  - (concertar cita por correo)
  - Correo: jgiraldez@go.ugr.es
  - Tutorías: jueves y viernes de 15.30-17-30 h
  - Despacho 1.10

Bases y Funcionamiento

Plataforma docente

- Web de la Plataforma PRADO - UGR
  - Acceder a través de http://pradogrado.ugr.es.
  - Toda la información y documentos relativos a la asignatura estarán disponible en dicha web.
  - Todos los alumnos deben verificar que el correo electrónico y la foto están disponibles en la web de la asignatura

Objetivos y Competencias (FICHA)

Competencias: Capacidad para conocer y desarrollar técnicas de aprendizaje computacional y diseñar e implementar aplicaciones y sistemas que las utilicen, incluyendo las dedicadas a extracción automática de información y conocimiento a partir de grandes volúmenes de datos.

Objetivos generales:

- Comprender el aprendizaje como mecanismo para obtener conocimiento, y mostrar las distintas formas en las que se puede realizar el aprendizaje.
- Distinguir entre aprendizaje supervisado, no supervisado y por refuerzo, así como determinar cuál de ellos es apropiado para resolver un determinado problema.
- Descripción y análisis de los distintos modelos de aprendizaje de conjuntos de hipótesis. Estudio de distintos métodos de aprendizaje
- Conocer diferentes modelos de **aprendizaje supervisado** y su aplicación en diferentes problemas. Conocer técnicas de validación y verificación de modelos, experimentar con dichas técnicas en diferentes problemas reales.
- Utilizar herramientas de aprendizaje en aplicaciones reales

Metas a alcanzar

- Al final del curso se debería conocer:
  - El conjunto de problemas, en el que las técnicas de A.A. son una aproximación adecuada.
  - Como identificar los modelos aplicables a un problema dado
  - Como aplicar los modelos estudiados
  - Las garantías que permiten aprender desde datos.

- Haber suscitado el interés por realizar aplicaciones en casos reales (Realizar TFG en problemas de AA)

Sistema de Evaluación Continua

- **2-Cuestionarios de Teoría (TT): 32 puntos**
- **3-Trabajos de Prácticas (TP)**: **36 puntos**
  - PRÁCTICAS: implementación y experimentación con algoritmos
  - Plazo de entrega pre-fijado.

- **PROYECTO FINAL (PF): 25 puntos** (**2 estudiantes**), para alumnos con TT1+TT2+TP1+TP2 $\geq$ 28
- **Examen TEORIA (ET): 25 puntos (individual),** para alumnos con TT1+TT2+TP1+TP2 < 28

- **Otros:** Interés y Participación**: hasta 6 puntos (participación en bonus y asistencia)**

- **Calificación final = (TTP + PF o ET+ Bonus)/10**

- **Matrícula de Honor:**
  - Haber superado claramente 105 puntos en **TTP + TF + Bonus**
  - **Haber desarrollado un proyecto final de alta calidad**

- **EVALUACIÓN EXTRAORDINARIA: examen escrito sobre los contenidos de la teoría y la implementación de los algoritmos de la asignatura**

- **EVALUACIÓN ÚNICA: se podrá elegir hacer un único examen final escrito de teoría y prácticas. Solicitar en la Sede Electrónica de la página web de la UGR.**

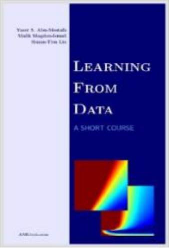
¿Qué necesitamos recordar?

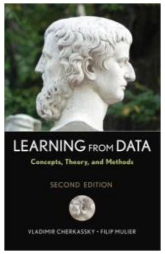
- **Notación y cálculo con matrices**
- **Conceptos básicos de probabilidad**
- **Cálculo de derivadas**
- **Cálculo de máximos y mínimos de una función**

- **Para repasar todos estos conceptos hay disponibles en la web documentos de ayuda y repaso.**
- **Si necesita ayuda con alguno de ellos acuda a tutorías**

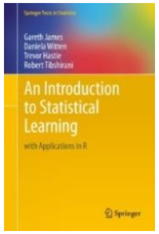
Documentos de consulta y apoyo

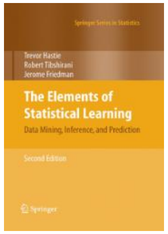
- El curso se intenta que sea lo más auto contenido posible.
- Transparencias de clase y otros documentos de apoyo están en la web de la asignatura (Inglés)
- Monografías de consulta:
  - Y.S. Abu-Mustafa, M. Magdom-Ismail, H. Lin, **Learning from Data**, AMLbook.com, 2012 (biblioteca)
  - V.Cherkassky, F.Mulier, **Learning from Data: concepts, theory and methods,** Wiley-Interscience, 2007 (en pdf)





- Otros libros complementarios:
  - G. James, D. Witten, T. Hastie and R. Tibshirani : An Introduction to Statistical Learning with Applications in R. Springer (http://www-bcf.usc.edu/~gareth/ISL/index.html)
  - Hastie, Tibshirani, Friedman, The Elements of Statistical Learning, (en pdf)





Prácticas de Implementación

- Prácticas: lenguaje Python + scikit-learn
  - Entornos de desarrollo: Google-Colab.
  - En clase de prácticas se darán los detalles

- **Tres grupos de prácticas: Martes(G1) ,Jueves(G3), Viernes(G2), (17.30-19.30):**
  - Los grupos han sido asignados por la ETSIIT
  - Las prácticas se corrigen por el profesor del grupo en el que se este
  - Ocasionalmente es posible asistir a otro grupo si hay espacio y el profesor lo permite.

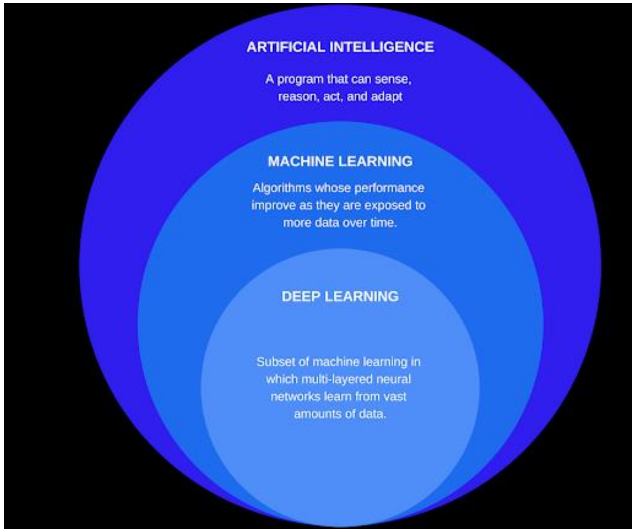
Código de Honor

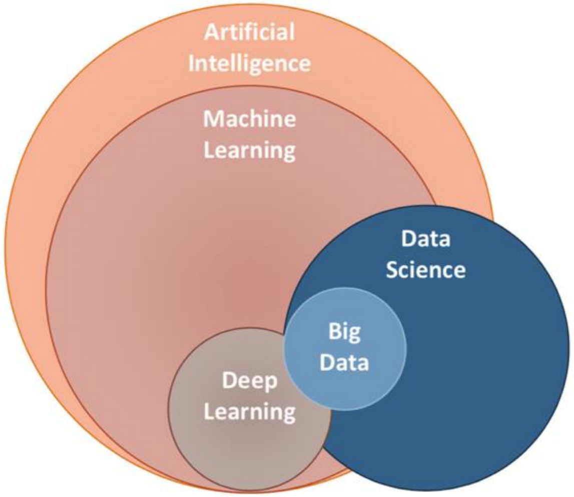
- **Trabajos de Teoría y Prácticas** :
  - **Se fomenta la colaboración entre alumnos a nivel de comprensión de conceptos e ideas**
  - **El desarrollo y escritura de los trabajos ES estrictamente individual**
  - **Si se usa información de alguna fuente debe explicitarse claramente en el TRABAJO de donde/ de quien se ha obtenido. En caso contrario se entenderá como COPIA.**

- **Detección positiva de copia**
  - **Se aplicará el Reglamento de exámenes de la UGR**

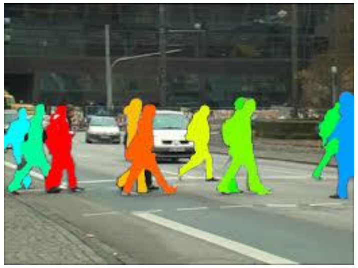
A.A.: Programa de la Asignatura

Sesión	Semana	CLASES DE TEORÍA	PRÁCTICAS-SEMINARIOS	ENTREGA DE TRABAJOS	
1	20 Febrero	Presentación de la Asignatura (1h) Definición de Aprendizaje Automático (1h)	Software de prácticas.		
2	27 Febrero	Modelo lineal: Regresión y Clasificación	Software de prácticas.		
3	6 Marzo	Modelo lineal: Estimación de la probabilidad Transformaciones no lineales	PRÁCTICA-1 Conceptos y algoritmos básicos	12 Entrega P0	
4	13 Marzo	Compromiso Sesgo-varianza Justificación del Aprendizaje Estadístico	PRÁCTICA-1 Conceptos y algoritmos básicos		
5	20 Marzo	Teoría de la generalización La dimensión VC	PRÁCTICA-1 Conceptos y algoritmos básicos		
6	27 Marzo	Sobreajuste Regularización	PRÁCTICA-2: Modelo lineales	2 Abril Entrega P1	
	3 Abril	VACACIONES			
7	10 Abril	Validación Principios Generales	PRÁCTICA-2 Modelo lineales		
8	17 abril	SVM	PRÁCTICA-2 Modelo lineales		
9	24 abril	SVM+Núcleos	PRÁCTICA-2 Modelo lineales	30 Abril: Entrega P2	
10	1 Mayo	Árboles "Random Forest"	PRÁCTICA-2 Modelo lineales		
11	8 Mayo	"Boosting" Redes Neuronales	PRÁCTICA-3 Boosting, RN, FBR		
12	15 Mayo	Redes Neuronales	PRÁCTICA-3 Boosting, RN, FBR		
13	22 Mayo	KNN - Funciones de base radial K-Medias & Mixturas Gaussianas	PRÁCTICA-3 Boosting, RN, FBR	21 Mayo: Entrega P3	
14	29 Mayo	Extracción automática de características	PRÁCTICA-3 Boosting, RN, FBR		
15	5 Junio	Extracción automática de características			
	20 Junio	EXAMEN TEORIA		19 Junio	





Learning **from Data**
(Machine Learning)



Interesting readings

- ***Machine learning: Trends, perspectives, and prospects***. M. I. Jordan and T. M. Mitchell (Science, 2015) (availables in PRADO)

- Want to Work in Artificial Intelligence? 14 AI Careers & Job Outlook [2022]
- **https://onlinedegrees.sandiego.edu/artificial-intelligence-jobs/**

- ***https://itchronicles.com/artificial-intelligence/where-is-ai-used-today/***

- ***https://www.infoworld.com/article/3438322/artificial-intelligence-today-whats-hype-and-whats-real.html***

- ***https://portal.mineco.gob.es/es-es/digitalizacionIA/Paginas/ENIA.aspx*****.**
  Secretaria de Estado de Digitalización e Inteligencia Artificial, 2023

The paradigm

Consider the following situation:

a) A student follows all the lessons of a ML course, and he/she thinks he/she know all the concepts.

b) In addition, he/she works on all the exercises and additional assignments.

The question is: has this student really learned the subject of Machine Learning?

Ok, Let's give him/her a test/quiz:

What do we ask to ensure that the test result answers the question?

A) Questions about the concepts or exercises worked on during the course.

B) Questions that can be answered from the concepts and techniques studied but not included in the material explained.

C) Mix of the two previous options.

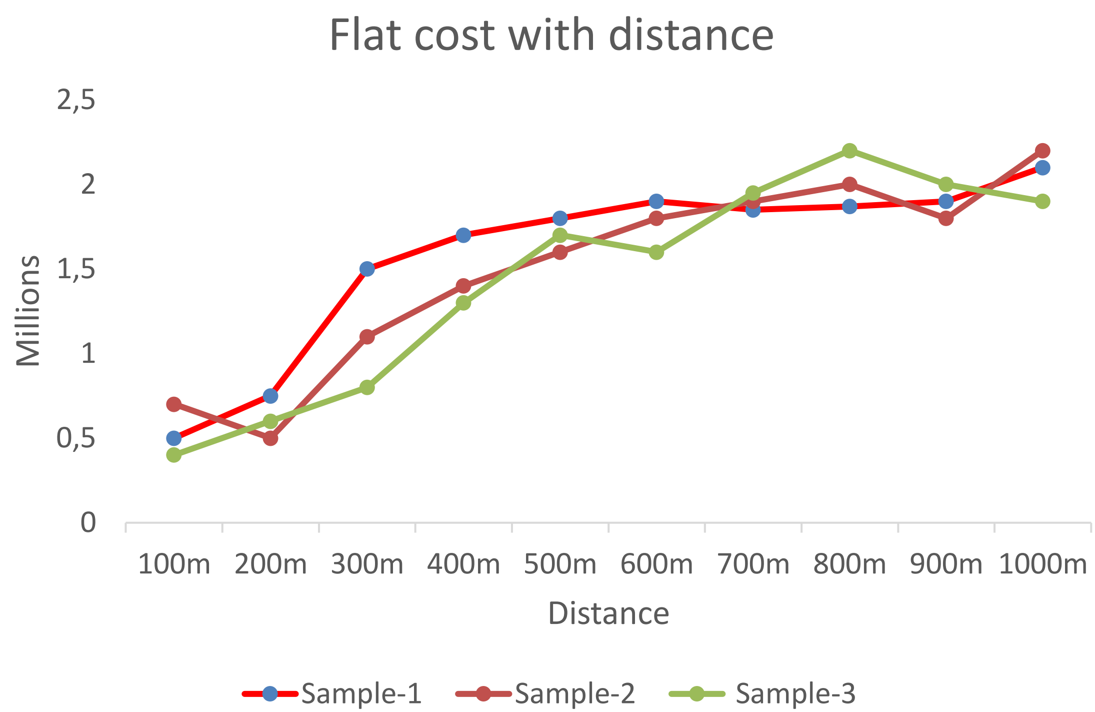
The Paradigm

- A ML- problem has two different stages:
  - Learning (Training)
    - You should learn as much as possible from the available material
  - Verification/Test
    - Must be successful on all new questions on the topic.

The difficult goal is to learn the RULES/FUNCTIONS/CONCEPTS valid to answer any possible question on the subject, not only for those previously trained.

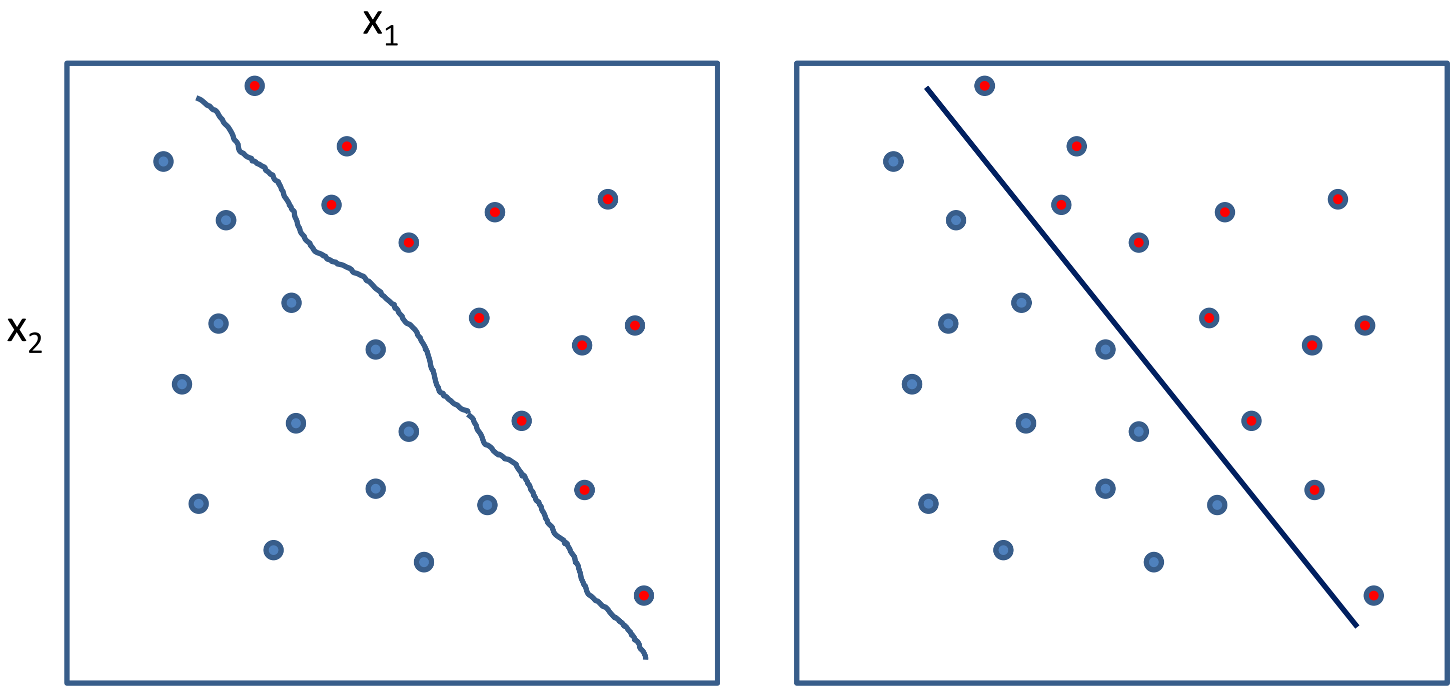
A simple example of ML

- Let suppose we know sales data of flats in a city, and we also know the flats distance to the city center: $(s_i, d_i), i = 1, .. N$
- Question: How to choose the function that best predicts the cost given a distance?



Try to guess the decisions to be made and the tasks to be solved....

Another simple example



- Color points represent a 2d-representation of two different objects
- For a given new point, how do we know which object it represents?

- *How to choose a color separation function using only data?*

Approaches to learning

- **Machine Learning (computer science):**
  - Focuses on accurated and efficient algorithms for large scale problems (out-of-sample generalization is the goal!)
  - Depends on the advances in optimization and regularization techniques.
  - Cons: bad models are always a possibility
- Statistical Learning: (classical approach)
  - Main focus is inference (explaining the data) using specific probability distributions
  - Good results only under the assumed hypothesis
  - Very poor attention to very large scale problems
- Bayesian Learning (full probabilistic)
  - A full probabilistic approach incorporating prior knowledge
  - Complex mathematically and computationally
  - Very poor attention to algorithm and computational issues
  - Good models when the hypothesis fit the data.

Machine Learning

TRADITIONAL PROGRAMMING

MACHINE LEARNING

Computer learns a calculation to solve a task

Suitable for computer but difficult for human being

Learning vs Design: What Learning is not !

- Some approaches only use data to fix some parameters of a well specified problem: this is design !
- Example: Let assume we want to build a model to recognizing coins from its size and mass: { (size_i,mass_i), i=1,..,N}
- Design: We collect information on the size and mass of each coin type and the number of coins in use. We build a physical model for mass and size, taking into account the variations by the use and the measured errors. Finally, we build a probability distribution on (size,mass) that we use to classify.
- Learning: We collect labeled data from each type of coin. The learning algorithm searches for a hypothesis that classifies the data well. We use the learned hypothesis to classify new coins
- *Available information* about the problem *is the key* to adopting one or the other approach

Machine Learning definitions

- Arthur Samuel : "the field of study that gives computers the ability to learn without being explicitly programmed." This is an older, informal definition.

- Tom Mitchell provides a more modern definition: "A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

- Example: playing checkers.

- E = the experience of playing many games of checkers
- T = the task of playing checkers.
- P = the probability that the program will win the next game.

- In general, any machine learning problem can be assigned to one of two broad classes:
  - Supervised learning : learn a function to PREDICT labels from sample labeled data
  - Unsupervised learning: Learns unlabeled data properties transforming the observed features.

Machine learning: Trends, perspectives, and prospects. M. I. Jordan and T. M. Mitchell (Science, 2015)

Summary of ML Paradigms

- **Supervised learning** : sample data + teacher (label) :
  - Semi-Supervised, Active Learning, On-line Learning**.**

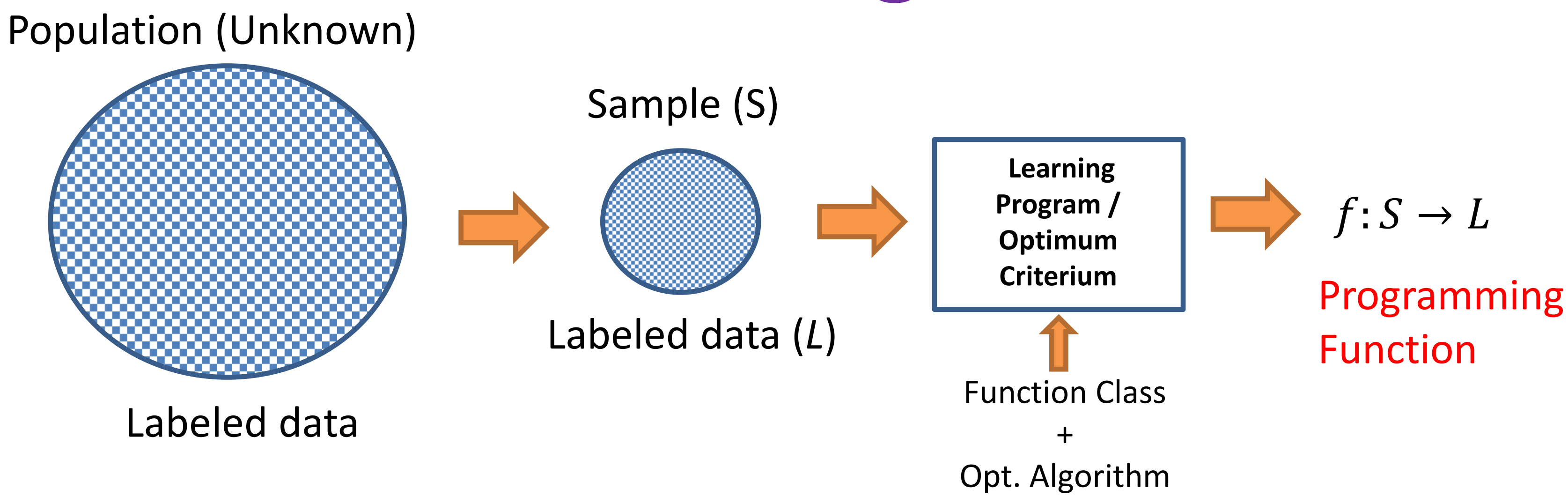
- **Unsupervised learning**: only data (it learns new subspaces for data representation)
  - **Self-Supervised:** Transform raw data into outstanding features by solving pretext-tasks that exploit data dependencies.

- **Transfer Knowledge:** How to share knowledge among different tasks.

- **Reinforcement Learning**: interaction with the environment. Learns a local dynamics(**dynamic learning**) not a global function.

Supervised Learning

The learning sketch



- Assumptions:
  - The only available information is the sample
  - The function is computed from the sample using a *Learning criterium*
  - The function computes a prediction

- Main task: *How to choose a sample and a machine learning program (How to learn) to find an accurate prediction function over the whole population?*

Try to make a proposal for the function class and opt. algorithm of the previous examples

Feature Extraction & Data Representation

How difficult is to represent a tree?

What features/measurement will encode a tree ?

Are they trees ?

Defining is Hard; Recognizing is Easy

Hard to give a complete encoding of a tree.
Even a 3 year old can tell a tree from a non-tree.
The 3 year old has learned from data.

A pattern exists. We don't know it. We have data to learn it.

PATTERN means ENCODINGS/ REPRESENTATION

Visual Patterns (encodings)

Identify handwritten digits in ZIP codes

Detecting faces in an image

A pattern exists. We don't know it. We have data to learn it.

Credit Approval

- Using salary, debt, years in residence, etc., approve for credit or not.
- No magic credit approval formula.
- Banks have lots of data.
  - customer information: salary, debt, etc.
  - whether or not they defaulted on their credit.

age	32 years
gender	male
salary	40,000
debt	26,000
years in job	1 year
years at home	3 years
. . .	. . .

Approve for credit?

A pattern exists. We don't know it. We have data to learn it.

Many many… applications

- Speech and speaker recognition
- Object and Face recognition
- Document translation and composition
- Recommendation systems
- Stock Market Prediction
- Automatic Prognosis and Diagnostic in Medicine
- And a very large etc

Main supervised tasks

- All PREDICTIONS !!
- Regression: the output is a real number (**continuous variable**)
  - Predict the height of a person from a data sample:
    - Features: weight; (weight, feet length); (weight, feet length, shoulders width), etc
  - Predict the temperature for the next day from a previous record of temperatures
- Classification: the output is a class label (**discrete/categorical variable**)
  - Predict the weather for tomorow : (sunny, cloudy, windy)
  - Predict if an image contains a face: (Yes,No), (0,1), (1,-1), etc
  - Predict if an email is SPAM or not: (Yes, No),), (0,1), (1,-1), etc
- Probabilistic Classification: the output is a probability vector over the labels
  - Predict real values
  - Solve the same problems as Classification

Unsupervised Learning

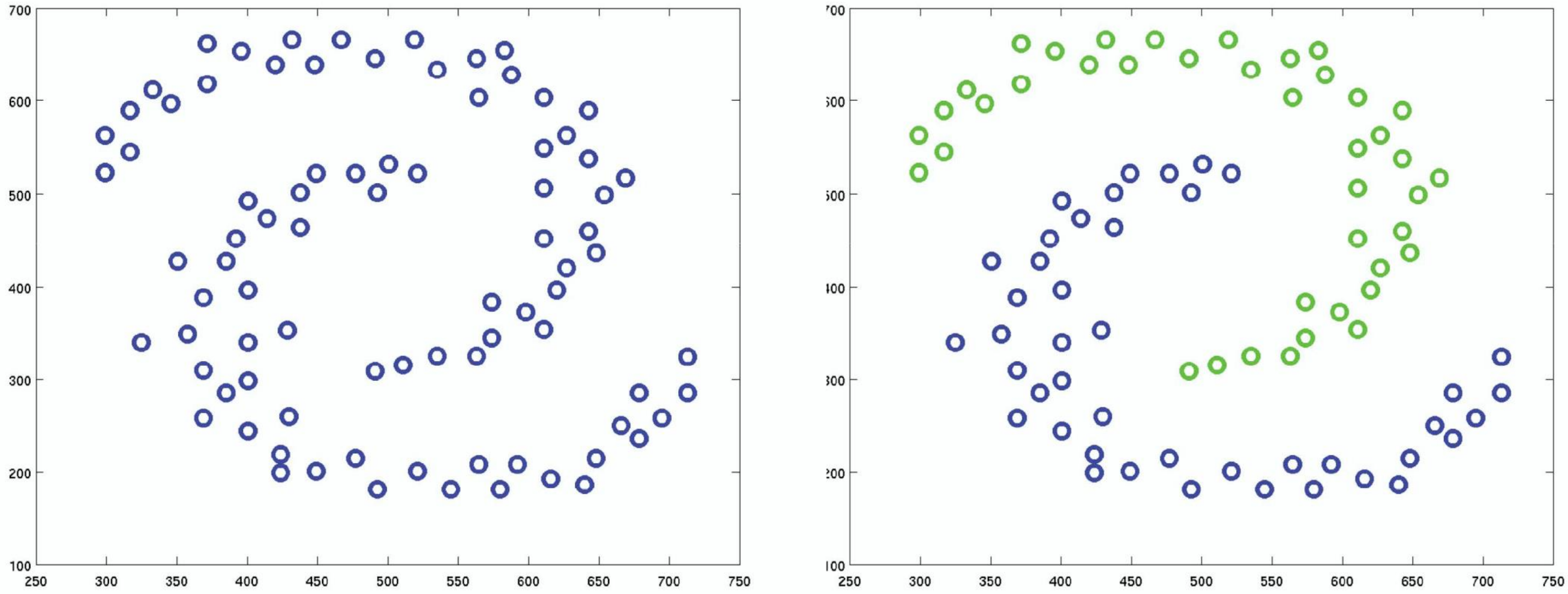
Unsupervised Learning

- The main objective is to transform the sample data to a better representation space by introducing a new distance measure between samples.

- Classical approaches:
  - Geometric structure: **clustering**
  - Dependences discovering: **patterns**
  - **Dimensionality reduction**: relevant features
  - Etc

- Disentangle the sampled data into more independent features that facilitate downstream processing

Clustering



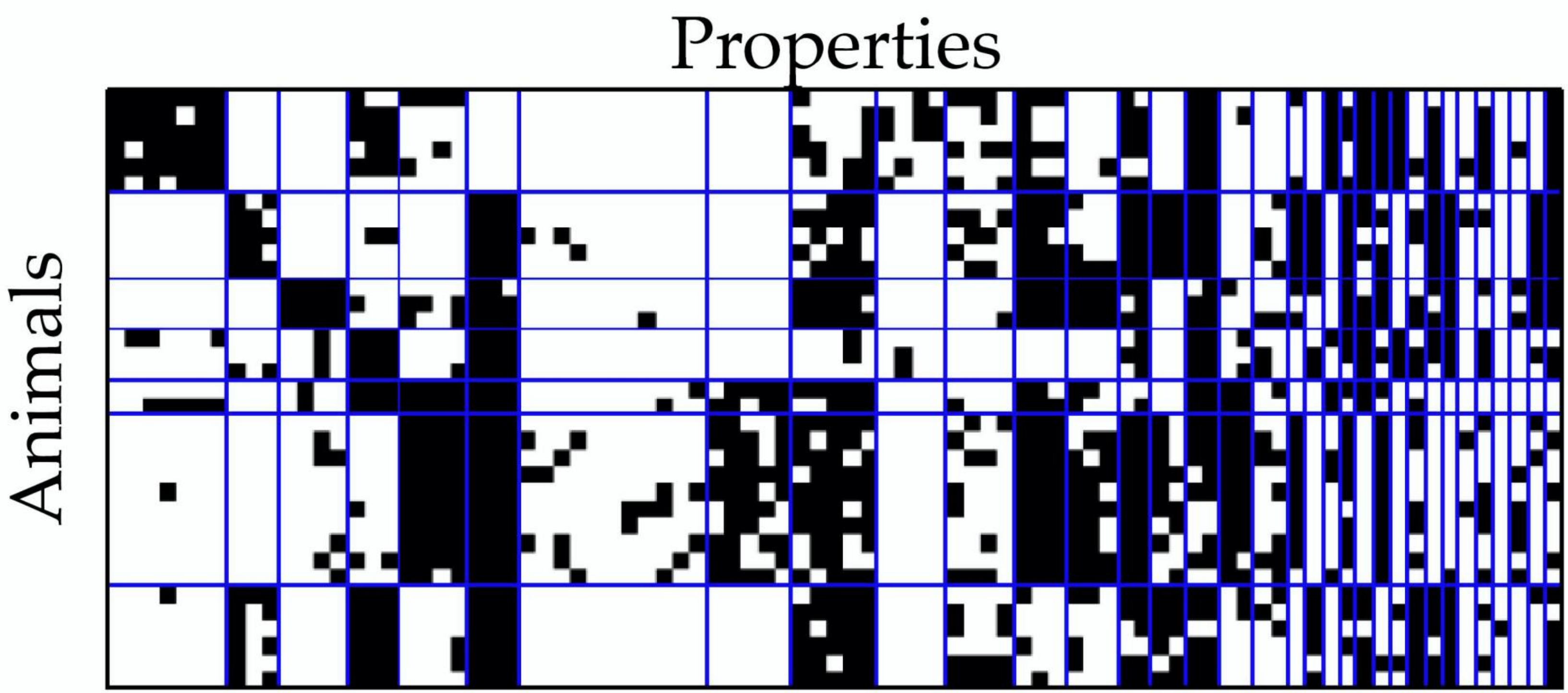
What distance is used ?

Example: Characteristics of Animals

- 50 animals
- 80 binary features
- Find interpretable groupings

O1	killer whale, blue whale, humpback, seal, walrus, dolphin
O2	antelope, horse, giraffe, zebra, deer
O3	monkey, gorilla, chimp
O4	hippo, elephant, rhino
O5	grizzly bear, polar bear

F1	flippers, strain teeth, swims, arctic, coastal, ocean, water
F2	hooves, long neck, horns
F3	hands, bipedal, jungle, tree
F4	bulbous body shape, slow, inactive
F5	meat teeth, eats meat, hunter, fierce
F6	walks, quadrapedal, ground



(J. Tenenbaum)

Dimensionality Reduction

Dimensionality Reduction

Self-Supervised

- Solve a supervised task to disentangle the sampled data into more independent features that facilitate downstream processing.
- The supervised task called pretext-task is easily set from the own data.
- By now is mainly focused on vision, speech and natural language tasks.

Jigsaw task

(a) (b) (c)

It allows to transform the input data into a new representation space useful for any downstream task.

In a nutshell

- Supervised:
  - Input data transformation : $T: D \rightarrow R$
  - Minimize an error mapping: $f: D \rightarrow R \rightarrow L$ (GT-labels)

- Unsupervised:
  - Input data transformation : $T: D \rightarrow R$
    - Discovering new distances between samples

- Self-supervised:
  - Input data transformation : $T: D \rightarrow R$
  - Minimize an error mapping: $f: D \rightarrow R \rightarrow L$
  - Easy labels not related to any specific task

D- data domain, *T* – transformation of the observed data
f: the function from *T* to *L* (lables), GT: Ground-Thruth

Formal elements of the supervised approach

1. **The dataset:**
   - Where the data come from? : a static probability distribution $\mathcal{P}(\mathcal{D})$
   - What features to use? $\mathcal{X}$
   - Any sampling condition?: independent identically distributed (i.i.d.)

2. **The prediction task:** $f: \mathcal{X} \rightarrow \mathcal{Y}$ $(\mathcal{Labels})$

3. **The model:**
   - What representation to use?:
     - Which class of function are we going to use? $\mathcal{H}$
     - How to characterize each element $h \in \mathcal{H}$? parameters, architecture

4. **Learning Algorithm: How to search inside $\mathcal{H}$?**
   - What criteria optimize to guarantee learning? $\mathcal{A}$ ERM, SRM, MDL, Similarity-Measure
   - What optimization to use to find the best function of $\mathcal{H}$? SGD

Let’s try to identify the elements….

Identify handwritten digits in ZIP codes

Detecting faces in an image

Regression or classification ?
Supervised or unsupervised ?

Elements of the Credit Approval problem

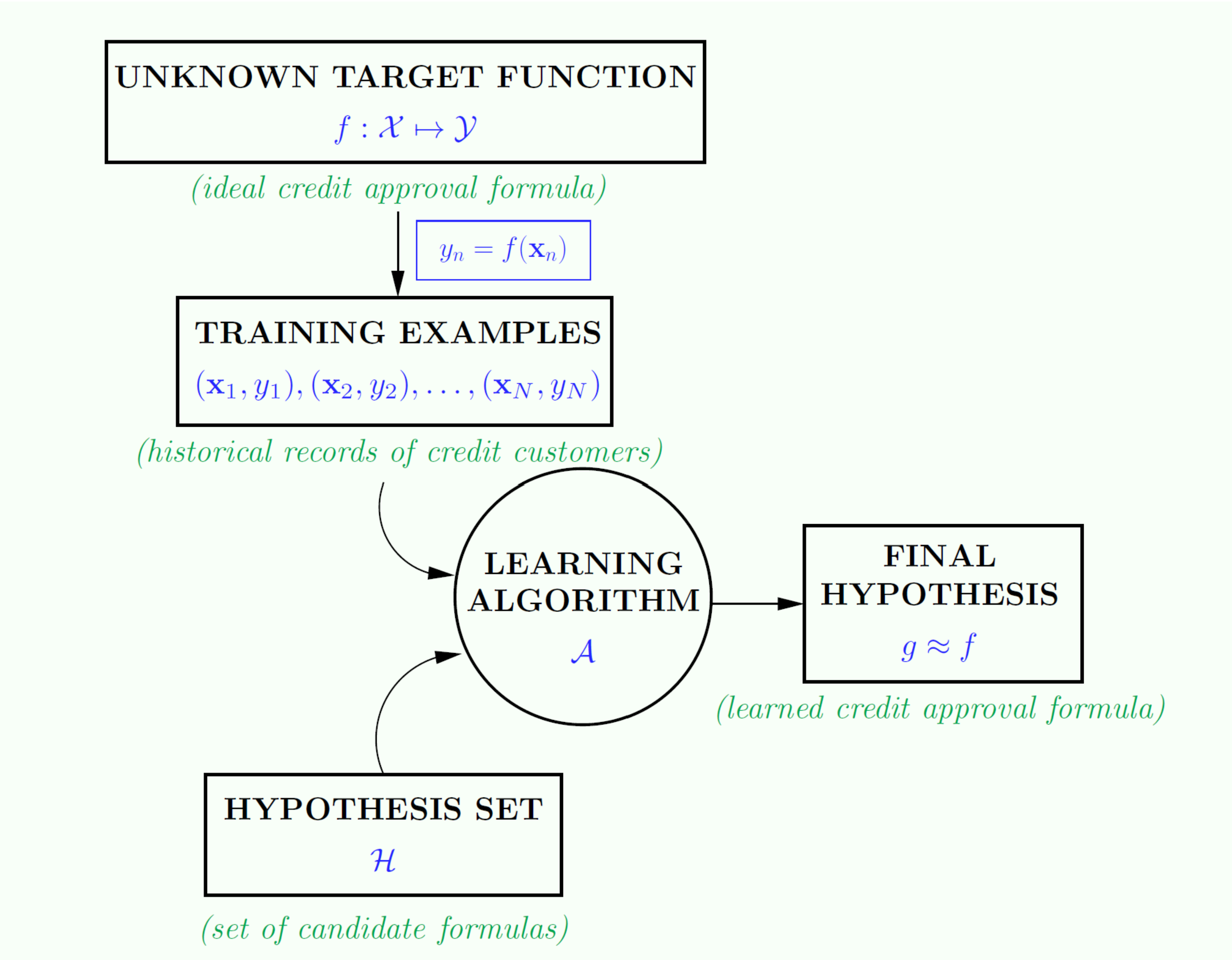
- Salary, debt, years in residence, … — *input* $\mathbf{x} \in \mathbb{R}^d = \mathcal{X}$.
- Approve credit or not — *output* $y \in \{-1, +1\} = \mathcal{Y}$.
- True relationship between $\mathbf{x}$ and y — *target function* $f : \mathcal{X} \mapsto \mathcal{Y}$.
  (The target f is *unknown*.)
- Data on customers — *data set* $\mathcal{D} = (\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_N, y_N)$.
  $(y_n = f(\mathbf{x}_n)$.)

We have identified some of the main elements of a learning task:

Input: feature vector
Output: class or label
Target function: unknown
Data Sample: i.i.d $\mathbf{x}_i$
Training sample: labeled data

When it is understood data set means training sample

Diagram of the initial Learning Setup



Try to identify the Key Elements

Let us assume you work for a company that is interested in discover a rule to set the price to pay for the mango harvest in a large extension crop.

- The company is only interested in paying for the tasty mangos that can be sold in the next days.
- Unfortunately you knowledge about mangos crop is very limited or null
- But, you can visit the farm and take a sample of fruit to identify the mangos properties (photo, color, size, texture, etc) associated with tasty/non-tasty mangos.

What to do?

Can you identify the principal elements of the problem?

How set-up a dataset that can be helpful?

What we already should know

- The formal elements of a learning-from-data approach
- The elements to be fixed before starting
- The learning-from-data goal