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: (nicolas@decsai.ugr.es) ,
  - Tutorías: Miércoles (9.30h-13.30h, 16.00h-18.00h).

- Prácticas:

Francisco J. Baldán Lozano(Grupo 1)
  - Despacho 31 (4ª planta)
  - Correo: fjbaldan@decsai.ugr.es
  - Tutorías: Lunes (12:00-13:00)

Ofelia Retamero Pascual (Grupo-2)
  - D31 (4ª planta). (concertar cita por correo)
  - Correo: : oretamero@decsai.ugr.es
  - Tutorías: Jueves (11.00-12.00)

Pablo Mesejo Santiago (Grupo-3)
  - D01 (4ª planta). (concertar cita por correo)
  - Correo: : pmesejo@decsai.ugr.es
  - Tutorías: Viernes (10:00-11:00)

Bases y Funcionamiento

Información de la asignatura

- Web en la Plataforma Docente de DECSAI
  - Acceder a través de http://decsai.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

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 interés por aplicaciones en casos reales (Realizar TFG en aplicaciones)

Sistema de Evaluación Continua

- **3-Trabajos de Teoría y Prácticas (TTP)**: **75 puntos (individual)**
  - Preguntas y ejercicios sobre los conceptos y técnicas explicadas.
  - Teoría: relación de cuestiones, habrá de 3-5 días para su contestación y envío.
  - PRÁCTICAS: implementación y experimentación con algoritmos
  - Plazo de entrega pre-fijado.

- **Examen FINAL (EF): 25 puntos (individual),** para alumnos con TTP <40 puntos o proporción (27, 2-TTP)
- **PROYECTO FINAL (PF): 25 puntos** (**2 estudiantes**), para alumnos con TTP $\geq$ 40 puntos o proporción (27, 2-TTP)

- **Otros:** Interés y Participación**: hasta 8 puntos (Asistencia a más del 70% de la clases)**

- **Calificación final = (TTP + PF o TF+ Otros)/10**

- **Matrícula de Honor:**
  - Haber obtenido 95 puntos o más en la calificación final
  - **Haber desarrollado un proyecto final de calidad**

- **EVALUACIÓN EXTRAORDINARIA: examen escrito sobre los contenidos de la teoría y algoritmos y prácticas 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.**

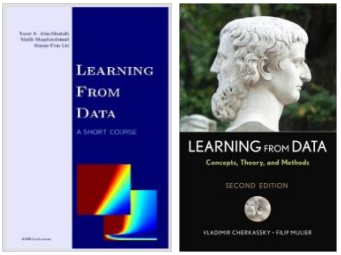
¿Qué necesitamos recordar?

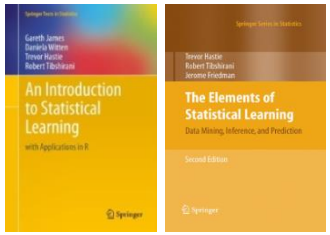
- **Notación y manipulación de matrices**
- **Conceptos básicos de probabilidad**
- **Cálculo básicos de cálculo de derivadas**
- **Cálculo de máximos y mínimos de una función**

- **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 apoyo:
  - 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 laboratorio

- Prácticas: lenguajes Python
  - Lenguaje relevante para análisis de datos: Scikit-learn
    - Instalar Anaconda 3.7 y la librería scikit_learn
  - Descargar e instalar en el ordenador portátil (Windows, Linux, MacOS)
  - Para su uso en las aulas, instalar en un disco/pendrive externo
  - En clase de prácticas se darán los detalles

- **Tres grupos de prácticas: lunes , miércoles y viernes (17.30-19.30):**
  - Apuntarse en la web de DECSAI a partir de las 20.00h de hoy
  - Las prácticas se corrigen por el profesor del grupo en el que se este
  - En caso de sobrecarga de un grupo, se asignarán los alumnos de la forma más razonable posible por parte de los profesores.
  - Ocasionalmente es posible asistir a otro grupo si hay espacio

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	Proyectos Finales
1	15 febrero	Presentación de la Asignatura (1h) Definición de Aprendizaje Automático (1h)	Software de prácticas.		
2	22 febrero	Modelo lineal: Regresión y Clasificación	Software de prácticas.		
3	1 marzo	Modelo lineal: Estimación de la probabilidad Transformaciones no lineales	PRÁCTICA-1 Conceptos y algoritmos básicos	Ejercicios Python	
4	8 marzo	Compromiso Sesgo-varianza Justificación del Aprendizaje Estadístico	PRÁCTICA-1 Conceptos y algoritmos básicos		
5	15 marzo	Teoría de la generalización La dimensión VC	PRÁCTICA-1 Conceptos y algoritmos básicos		
6	22 marzo	Sobreajuste Regularización	PRÁCTICA-2: Modelo lineales		
7	29 marzo	Validación Principios Generales	PRÁCTICA-2 Modelo lineales	25 marzo: Entrega T1	
8	5 abril	SVM	PRÁCTICA-2 Modelo lineales		
9	12 abril	SVM+Núcleos	PRÁCTICA-2 Modelo lineales		
	19 abril	VACACIONES			
10	26 abril	Árboles "Random Forest"	PRÁCTICA-3 Boosting, RN, FBR	22 Abril: Entrega T2	
11	3 mayo	"Boosting" Redes Neuronales	PRÁCTICA-3 Boosting, RN, FBR		Oferta de proyectos
12	10 mayo	Redes Neuronales	PRÁCTICA-3 Boosting, RN, FBR		Selección de proyectos
13	17 mayo	Extracción automática de características	PRÁCTICA-3 Boosting, RN, FBR		
14	24 mayo	KNN - Funciones de base radial K-Medias & Mixturas Gaussianas	Desarrollo Proyecto F.C.	20 mayo: Entrega T3	Presentación objetivos del Proyecto
15	31 mayo	Reducción de dimensionalidad	Desarrollo Proyecto F.C.		
7	Junio				Entrega de proyectos y examen final

Learning **from Data**
(Machine Learning)

What is Machine Learning ?

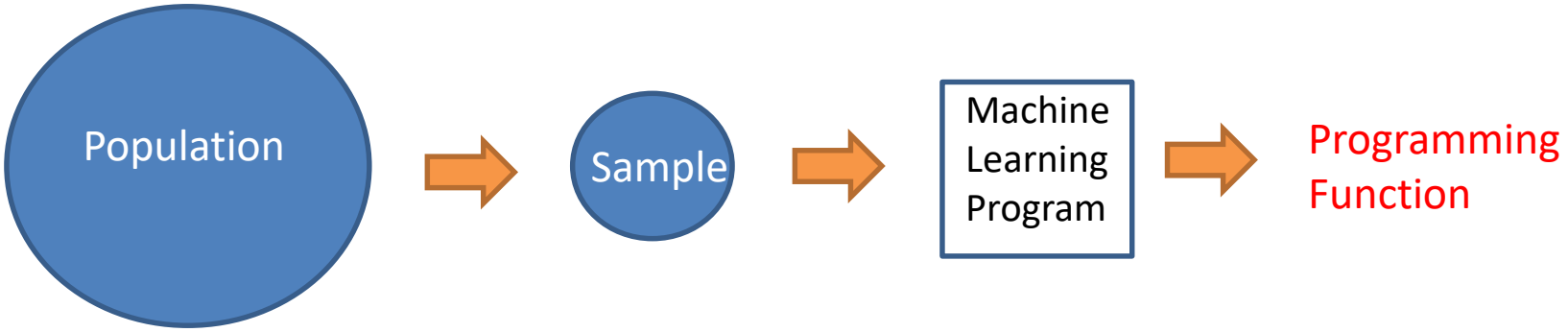
TRADITIONAL PROGRAMMING

MACHINE LEARNING

Computer learns a calculation to solve a task

Suitable for computer but difficult for human being

A learning task sketch



- Hypothesis:
  - The only available information is the sample
  - The function is computed from the sample using a Learning program
  - The function computes a prediction

- Main question: How to choose a sample and a Machine Learning Program to find an accurate prediction-function on the whole population?

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 classifications:
  - Supervised learning
  - Unsupervised learning.

Supervised Learning

Visual Patterns

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.

Example: Netflix Competition

See Wikipedia

- October 2006: Released user ratings data spanning six years
  - [anon ID, date, title, year, rating]
  - Some noise added
  - 100M+ ratings
  - 480K+ users
  - 17K+ movies
- Objective: perform significantly better than Netflix's internal algorithm in terms of root mean squared error (RMSE) on held-out test data.
- Starting point: 0.9525 RMSE
- Objective: 0.8572 RMSE
- $1M Grand Prize

More examples

Computational linguistics: tagging parts of speech

Input:

Profits soared at Boeing Co., easily topping forecasts on Wall Street, as their CEO Alan Mulally announced first quarter results.

Output:

Profits/N soared/V at/P Boeing/N Co./N, easily/ADV topping/V forecasts/N on/P Wall/N Street/N, as/P their/POSS CEO/N Alan/N Mulally/N announced/V first/ADJ quarter/N results/N.

(M. Collins)

How difficult is to define a tree?

A brown trunk moving upwards and branching with leaves . . .

Are these trees ?

Defining is Hard; Recognizing is Easy

Hard to give a complete mathematical definition 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.

Two main supervised paradigms

- Both represent PREDICTIONS !!

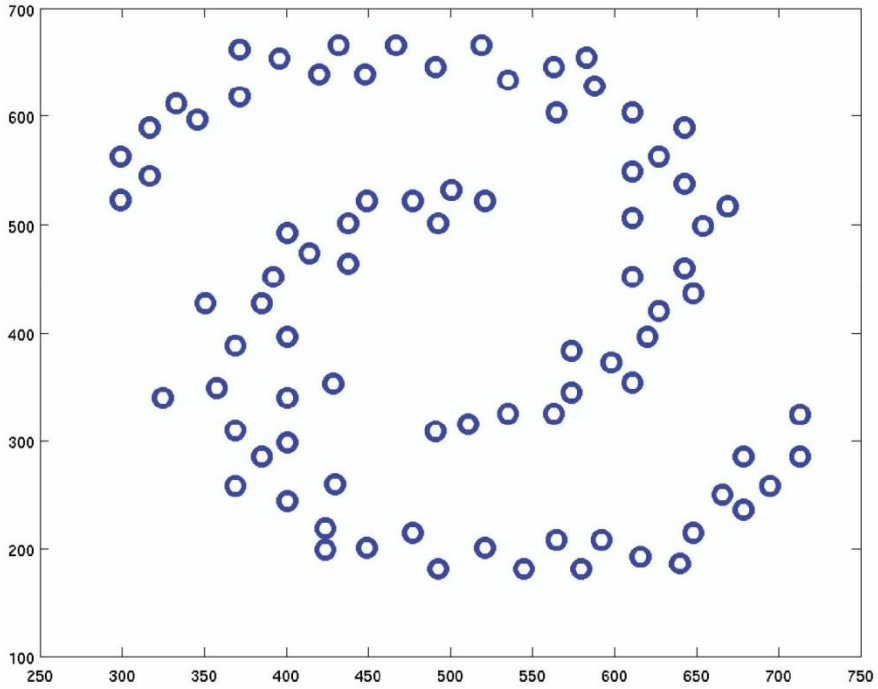
- Regression: the output is a real number (**continuous variable**)
  - Predict the height of a person from a data sample:
    - Features: weight; (weight, feet length); (weight, feet length, shoulders width), etc
  - Predict the temperature for the next day from a previous record of temperatures

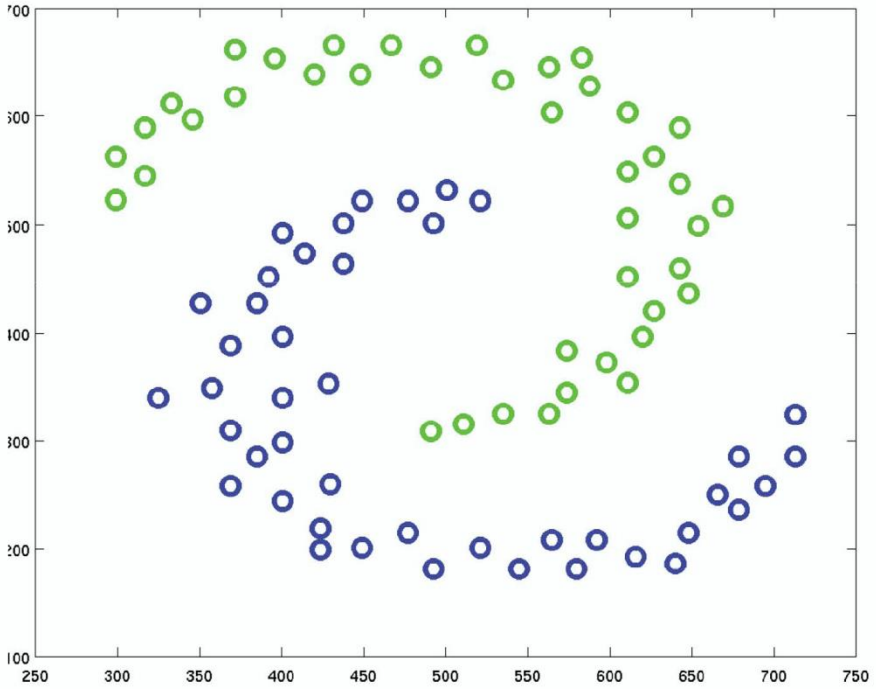
- Classification: the output is a class label (**discrete variable**)
  - Predict the weather for tomorow : (sunny, cloudy, windy)
  - Predict wether an image contains a face: (Yes,No), (0,1), (1,-1), etc
  - Predict wether an email is SPAM or not: (Yes, No)

Unsupervised Learning

- Data modeling to discover what is inside:
  - Geometric structure: cluster
  - Dependences discovering: patterns
  - Dimensionality reduction: relevant features
  - etc

Clustering


700
600
500
400
300
200
100
250
300
350
400
450
500
550
600
650
700
750

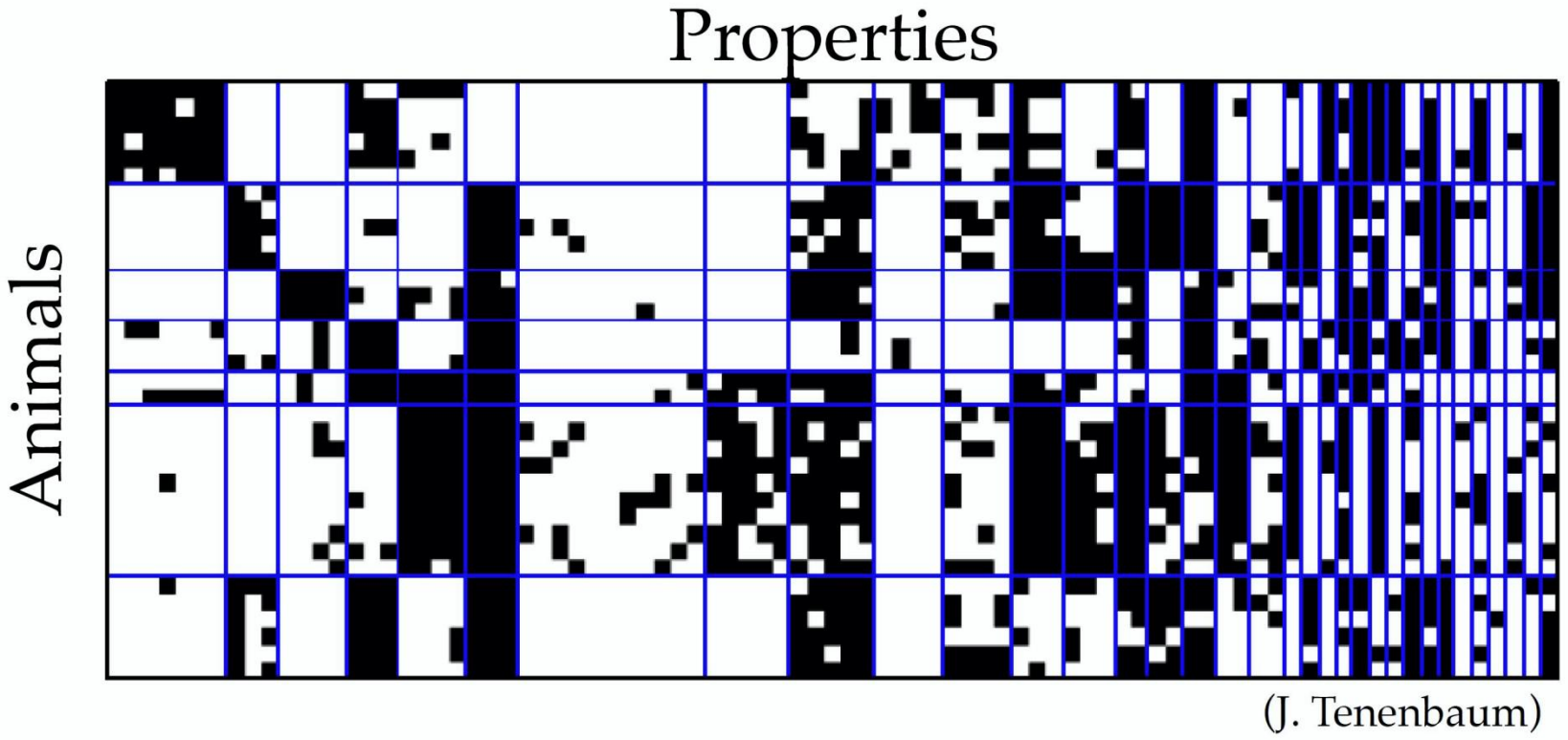

700
600
500
400
300
200
100
250
300
350
400
450
500
550
600
650
700
750

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)

Review

A) Suppose we feed a learning algorithm a lot of historical weather data, and have it learn to predict weather. In this setting, **what is T**?

1. The probability of it correctly predicting a future date's weather.
2. The weather prediction task.
3. None of these.
4. The process of the algorithm examining a large amount of historical weather data.

B) Suppose you are working on weather prediction, and use a learning algorithm to predict tomorrow's temperature (in degrees Centigrade/Fahrenheit). **Would you treat this as a classification or a regression problem?**

C) Suppose you are working on stock market prediction. You would like to predict whether or not a certain company will declare bankruptcy within the next 7 days (by training on data of similar companies that had previously been at risk of bankruptcy). **Would you treat this as a classification or a regression problem?**

D) Some of the problems below are best addressed using a supervised learning algorithm, and the others with an unsupervised learning algorithm. **Which of the following would you apply supervised learning to?** (Select all that apply.) In each case, assume some appropriate dataset is available for your algorithm to learn from.

1) Take a collection of 1000 essays written on the US Economy, and find a way to automatically group these essays into a small number of groups of essays that are somehow "similar" or "related".

2) Given historical data of children's ages and heights, predict children's height as a function of their age.

3) Examine a large collection of emails that are known to be spam email, to discover if there are sub-types of spam mail.

4) Given 50 articles written by male authors, and 50 articles written by female authors, learn to predict the gender of a new manuscript's author (when the identity of this author is unknown).

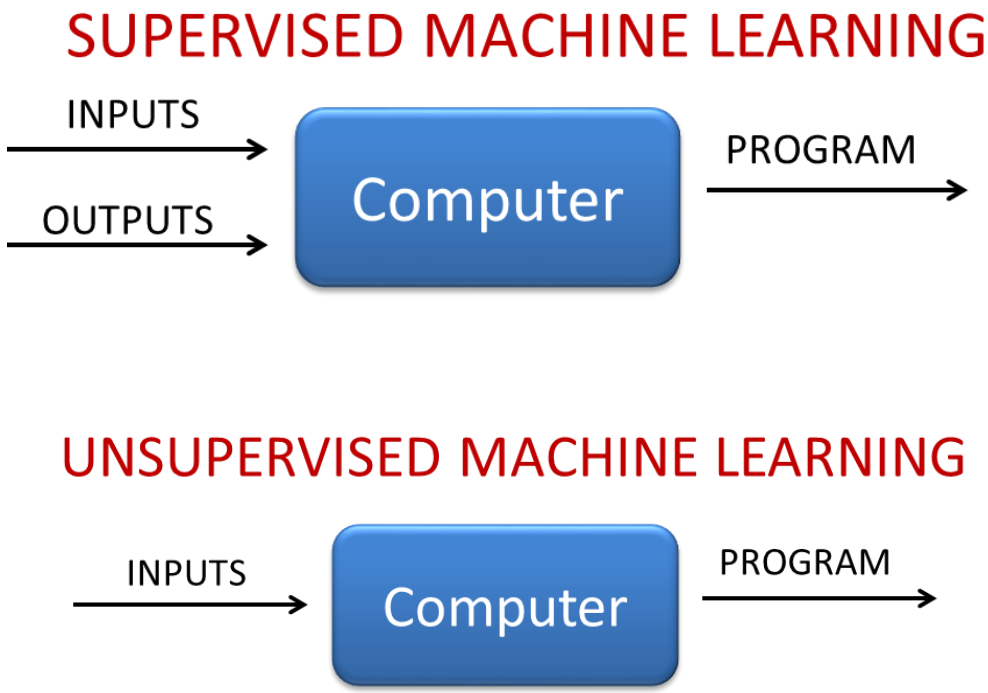
E) **Which of these is a reasonable definition of machine learning?**

1. Machine learning is the field of study that gives computers the ability to learn without being explicitly programmed.
2. Machine learning learns from labeled data.
3. Machine learning is the science of programming computers.
4. Machine learning is the field of allowing robots to act intelligently.

Scope of the course

- What is learning ?
- Can we do it ?
- How to do it?
- How to do it well ?
- General principles ?
- Advanced supervised techniques:
  - kNN, SVM, Kernels, Neural Networks(NN,CNN), Boosting, Random Forest
- Non-supervised techniques:
  - K-means, Mixture of Gaussians, Radial Base Functions, Principal Component Analysis.

What we already know



"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."

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 of 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

Try to identify the main elements….

Identify handwritten digits in ZIP codes

Detecting faces in an image

Regression or classification ?

Supervised or unsupervised ?

Try to identify the Key Elements

Let us assume you work for a company that is interested in discover a rule to fix 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?

Let's formalize the approach

1. What is the available information?
   - Where the data come from? $\mathcal{X}$, $f: \mathcal{X} \rightarrow \mathcal{Y}$
     - What features to use?
   - Is the data collection relevant? Why?: $\mathcal{P}$
     - Should we assume some hypothesis ? independent identically distributed (i.i.d.)

2. How to set-up a model ?
   - What representation use?:
     - Which class of function are we going to use? $\mathcal{H}$
     - How to characterize each element $h \in \mathcal{H}$? h

3. How to search inside $\mathcal{H}$?
   - What criteria to use to guarantee learning? ERM, SRM, MDL
     - The function to optimize: (Loss function)
   - What algorithm to use to find the best function of $\mathcal{H}$? $\mathcal{A}$

Examples

1. Use socio-economic data of families living in different areas of a city joint to the data of the last sales to predict the price of a house/flat.

2. Medicine: in cancer diagnosis the mitosis detection is relevant problem. Collect cells image to learn a binary classifier to predict mitotic cells.

3. Build a classifier for wine quality from physicochemical data of different type of wines and vineyard.

4. Build a classifier to recognize la presence of an object/face in an image.

5. Build a classifier to recognize a generic object(categories): dog, cat, tree, pedestrian, etc

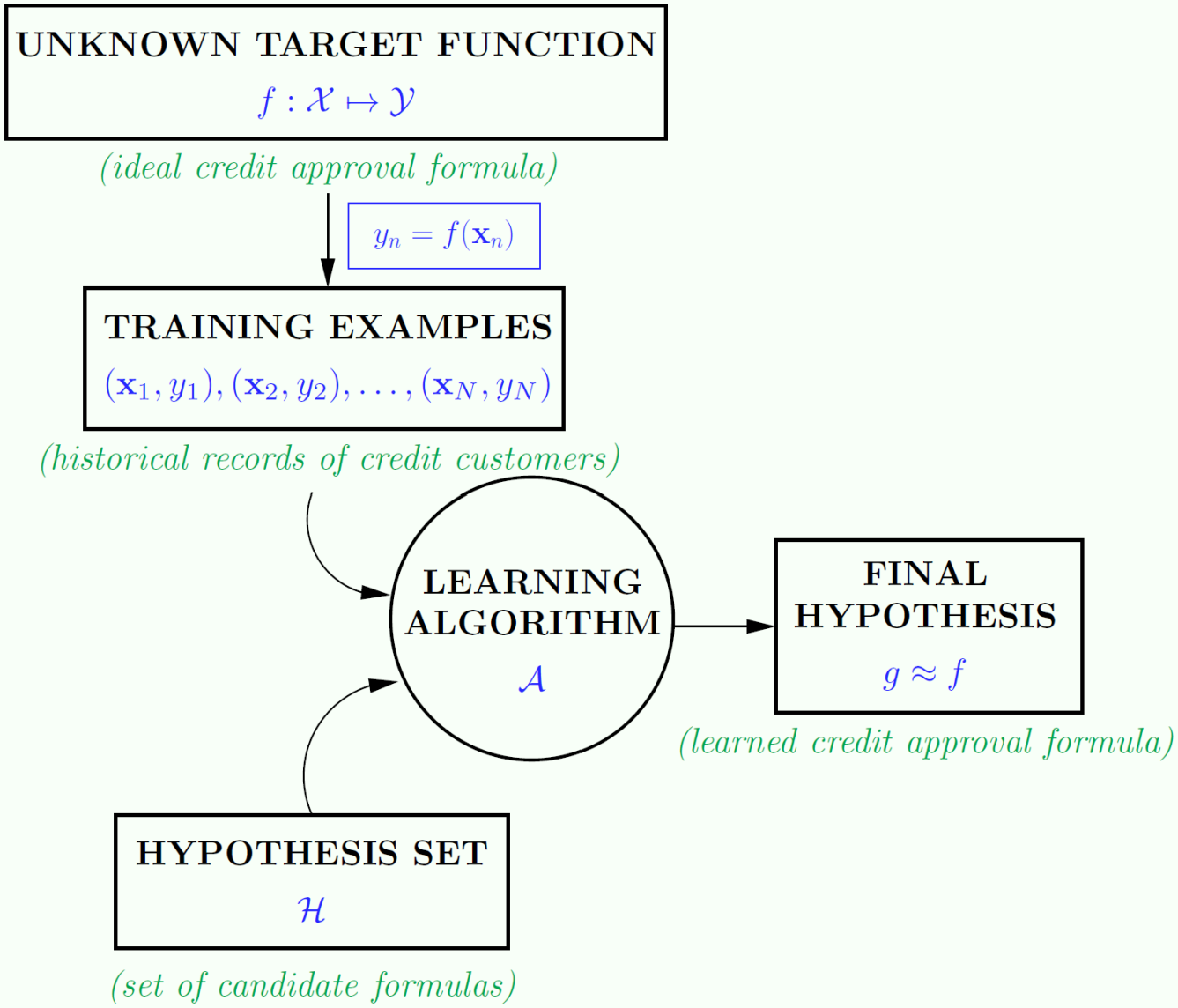
6. Classify documents according to its topic

7. Voice recognition

8. Automatic translation

9. etc, etc

Summary of the initial Learning Setup



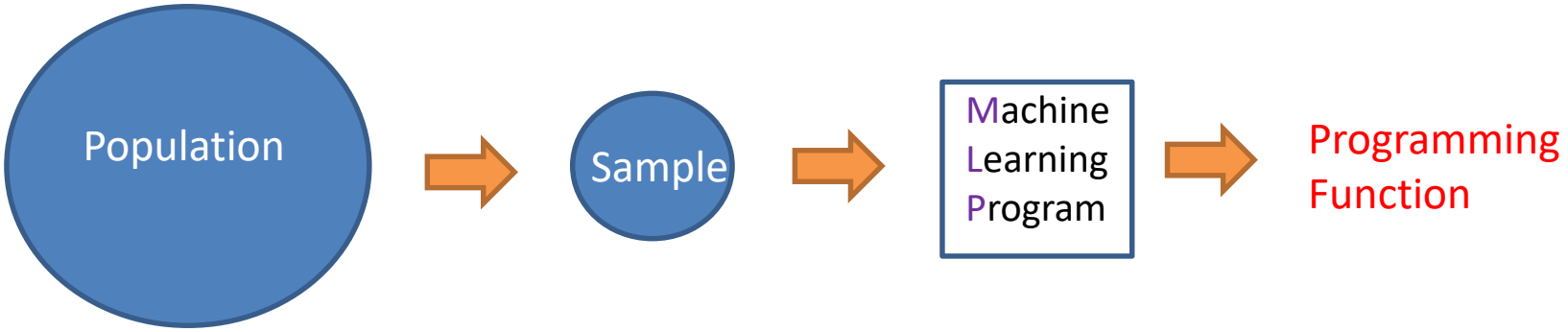
Summary of the approach

- A feature space/ Domain set/ Set of feature: $\mathcal{X}$
- A label set: $\mathcal{Y}$ (discrete: $\{(0,1),(-1,1),(1,2,3,..,M)\}$ o continuos: $\mathbb{R}$)
- Unknown target/label function: sample→ label ($f: \mathcal{X} \to \mathcal{Y}$ The unknown true function)
- Training examples: set of labeled samples , $\mathcal{D} = \{(x_i, y_i)), i = 1, \cdots, N\}$
  - independent and identically distributed (i.i.d.) samples from $(\mathcal{X}, \mathcal{Y})$

Modelling

- Hypothesis Set: $\mathcal{H}$ (Set of candidate function where the approximated solution is chosen from)
  - Solution hypothesis: a function $g \in \mathcal{H}$
- Learning optimization criteria: Criteria used to choice the function g (ERM,SRM, MDL)
- **Loss Function**: $L(h) \in \mathbb{R}^+, h \in \mathcal{H}$ (Criteria used to assign a value of merit to each function)
- Learning Algorithm: $\mathcal{A}$ (Algorithm that determines, **from samples**, the best global hypothesis.)

A learning task sketch



Main question: How to choose a sample and learn from it an accurate prediction-function in the whole population?

- Sample: i.i.d from the population
- MLP: is equivalent to fix $(\mathcal{H}, \mathcal{A})$

Review

- What we already know:
  - The formal elements of a learning-from-data approach
  - The elements to be fixed before starting
  - The learning-from-data goal

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. To classify a new coin we use the learned hypothesis

- The information about *f* is key to adopt one or another approach.

Approaches of learning

- Machine Learning (computer science):
  - Main focus is on accurated prediction from large scale problems (generalization is important!)
  - Algorithm efficiency is an issue
  - Very dependent on the advances in optimization and regularization techniques.
  - Cons: Overfitting is always a possibility

- Statistical Learning: (statistical goals)
  - Main focus is inference (explaining the data) using probability distributions
  - Good results only under the assumed hypothesis
  - Very poor attention to very large scale problems

- Data Mining (statistical & computer science):
  - The main focus is extract dependences between variables in large databases. That is large inference
  - Shares many tools with M.L.
  - Algorithms and hardware with high level of scalability are important

- Bayesian Learning (probabilistic)
  - A full probabilistic approach based on a-priori distributions as prior knowledge
  - Overfitting is not in general an issue
  - Much more complex mathematically and computationally
  - Very poor attention to algorithm and computational issues