Aprendizaje Automático

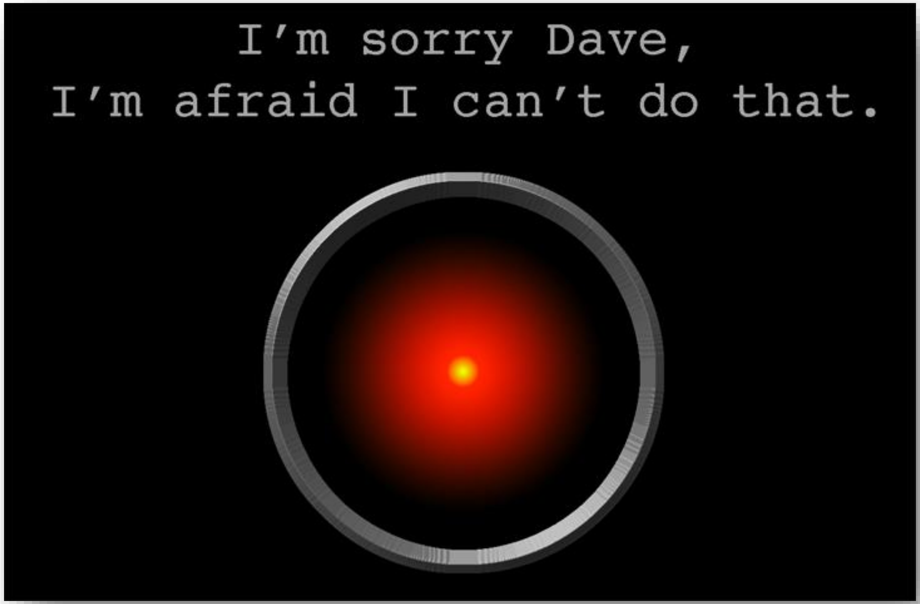
Introduction to Machine Learning

Máster en Bioinformática y Biología Computacional

Machine Learning

What is Machine Learning?

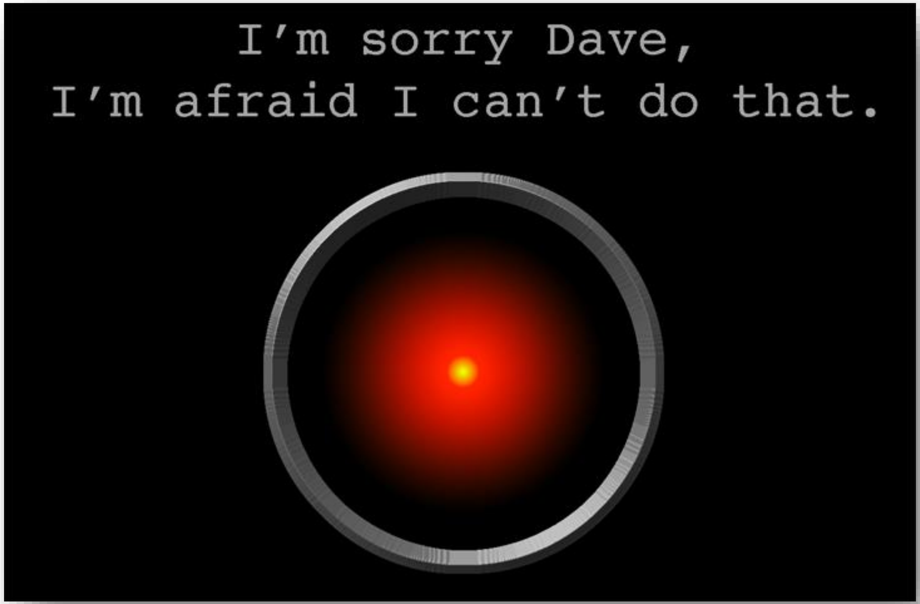
- Construction and study of systems that can learn from data (*Wikipedia*)
- Field of study that gives computers the ability to learn without being explicitly programmed (*Arthur Samuel*)



Machine Learning

What is Machine Learning?

- Construction and study of systems that can learn from data (*Wikipedia*)
- Field of study that gives computers the ability to learn without being explicitly programmed (*Arthur Samuel*)



HAL 9000

2001, A Space Odyssey

Stanley Kubrick, 1968

Machine Learning

What is Machine Learning?

- Arthur Samuel, back in 1950´s wrote a checkers playing program, that was able to learn the best board positions by analyzing 1000's of games. The system learnt by itself how to play checkers better and better.

Machine Learning

What is Machine Learning?

- On May 11, **1997**, chess grandmaster **Garry Kasparov** resigns after 19 moves in a game against Deep Blue, a chess-playing computer developed by scientists at IBM.

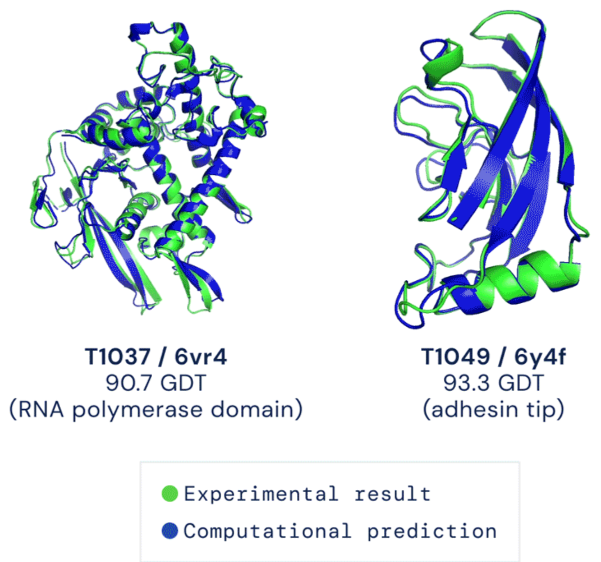
Machine Learning

What is Machine Learning?

- In 2016, Google (AlphaGo) defeated the Go World Champion. This game was considered during decades ones of the great challenges in AI.

Machine Learning

What is Machine Learning?



- In 2020, Google (AlphaFold) predicts the structure of proteins

> We have been stuck on this one problem – how do proteins fold up – for nearly 50 years. To see DeepMind produce a solution for this, having worked personally on this problem for so long and after so many stops and starts, wondering if we'd ever get there, is a very special moment.

PROFESSOR JOHN MOULT
CO-FOUNDER AND CHAIR OF CASP, UNIVERSITY OF MARYLAND

What is Ma

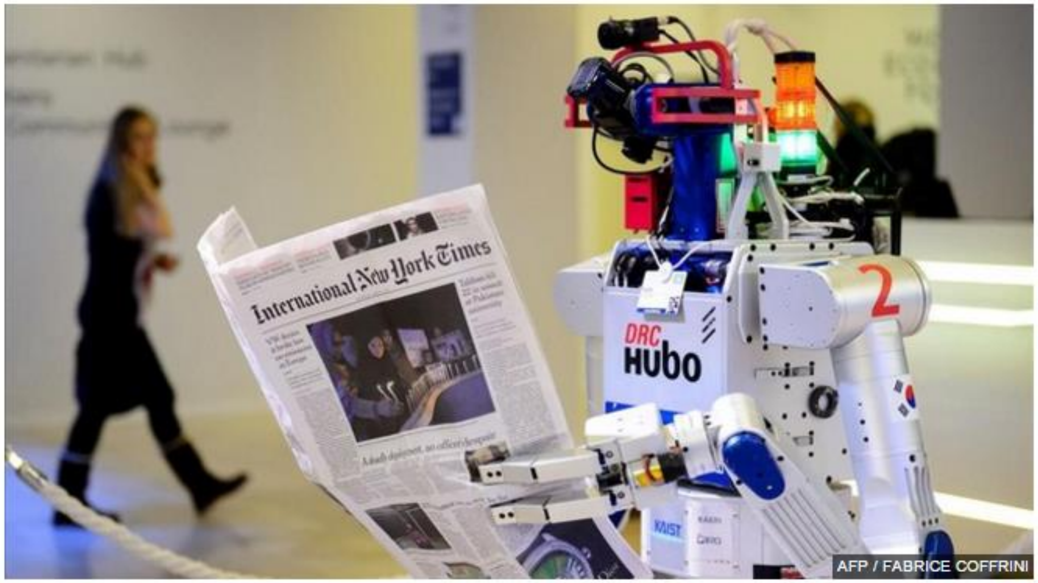
- Constructi … rn from data (*Wik*
- Field of st … o learn without be … *amuel*)

LA REVOLUCIÓN CIBERNÉTICA

La inteligencia artificial publica sus secretos

- Google, Facebook, Microsoft o IBM liberan software de reconocimiento de imágenes y voz para involucrar a programadores y empresas
- El objetivo es conseguir avances en ordenadores que aprendan a funcionar como la mente humana

46 2 COMENTARIOS 5



Robot leyendo un periódico en el foro de Davos este viernes.

CARMEN JANÉ

@carmenjane

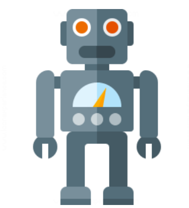
Enviar por correo
Cuerpo de letra
Imprimir noticia
ENVÍA UNA CARTA DEL LECTOR

SÁBADO, 23 DE ENERO DEL 2016 - 18:10 CET

La mayoría de las **grandes compañías tecnológicas** (**Google**, **Facebook**, **Microsoft**, **Nvidia**, **IBM**) han hecho público durante este último mes el **software de complejos programas de inteligencia artificial** para dar un impulso a esta disciplina, que

From Classical Programming to Machine Learning

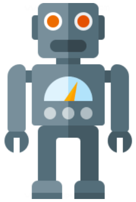
Experience → **Learning** → Responses



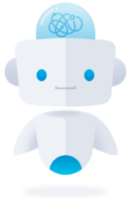
Rules, Data → **Classic programming** → Responses

From Classical Programming to Machine Learning

Experience → **Learning** → Responses



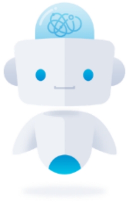
Rules, Data → **Classic programming** → Responses



Data, Responses → **Machine Learning** → Rules

From Human Bevavior to Machine Behavior: The Standard Model

Human Behavior 

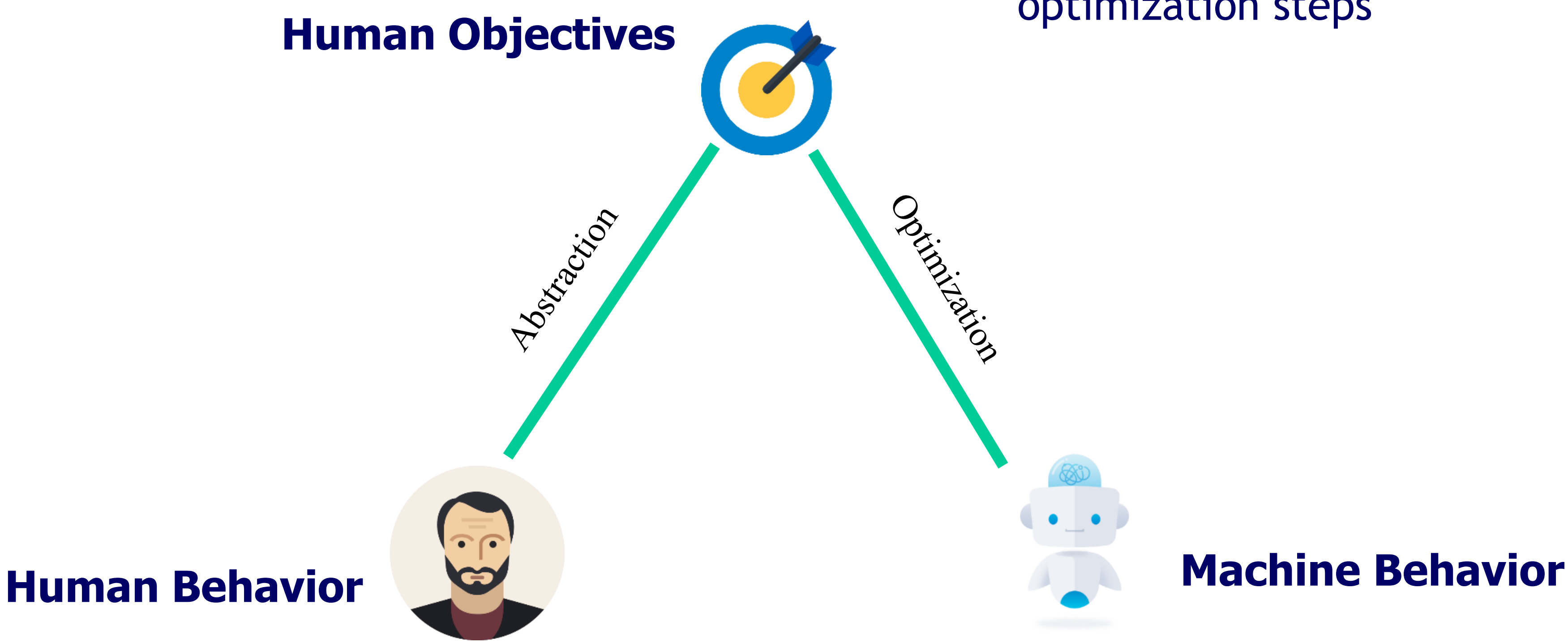
 Machine Behavior

From Human Bevavior to Machine Behavior: The Standard Model

- Learning process is guided by pre-defined objectives (i.e. simplification of complex behaviors).

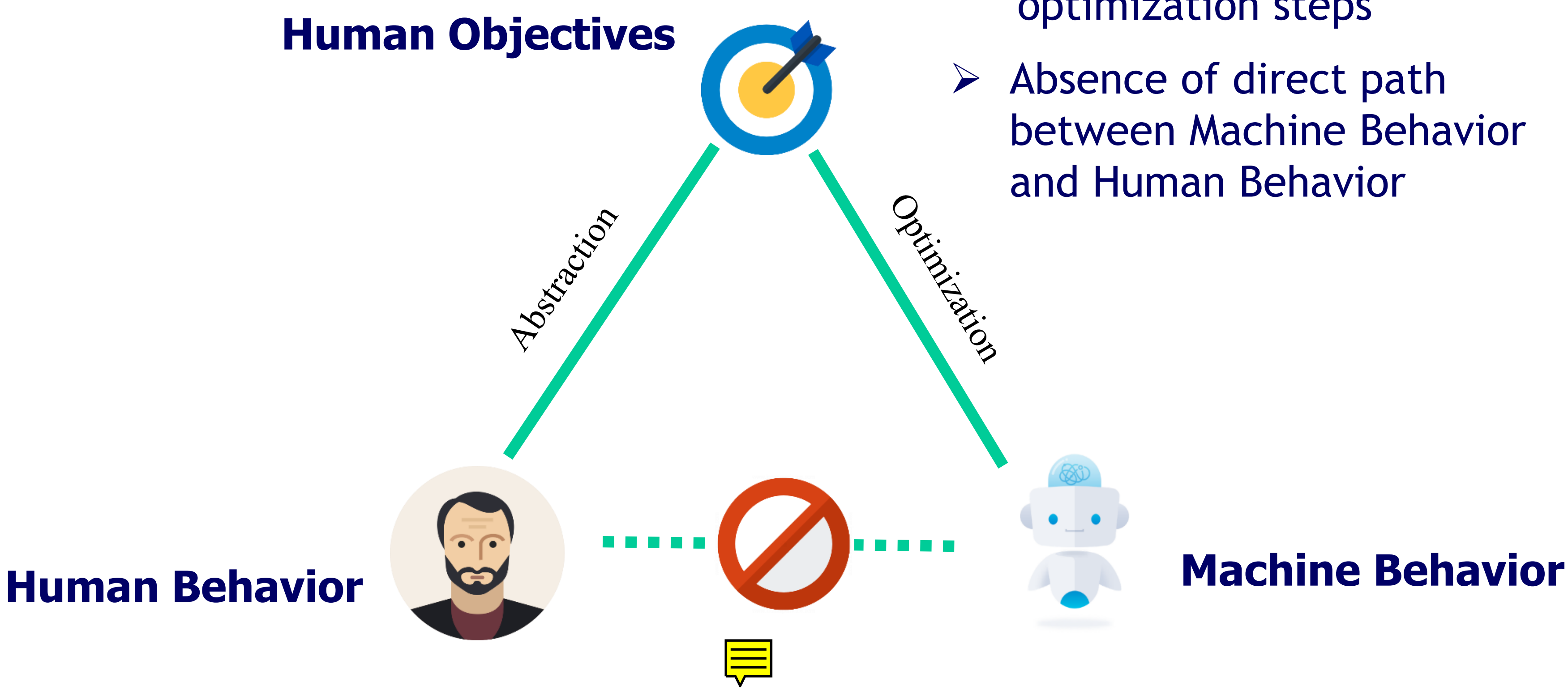
From Human Bevavior to Machine Behavior: The Standard Model

- Learning process is guided by pre-defined objectives (i.e. simplification of complex behaviors).
- Learning process consist of optimization steps



From Human Bevavior to Machine Behavior: The Standard Model

- Learning process is guided by pre-defined objectives (i.e. simplification of complex behaviors).
- Learning process consist of optimization steps
- Absence of direct path between Machine Behavior and Human Behavior



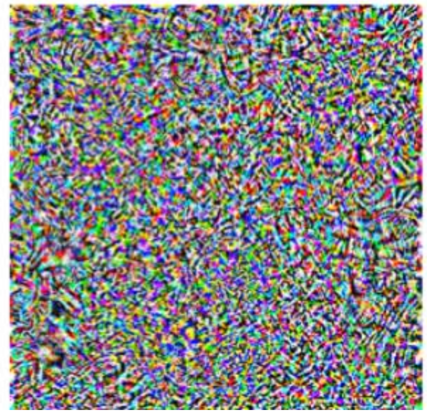
How AI See: An Example with Adversarial Attacks

“pig”

How AI See: An Example with Adversarial Attacks

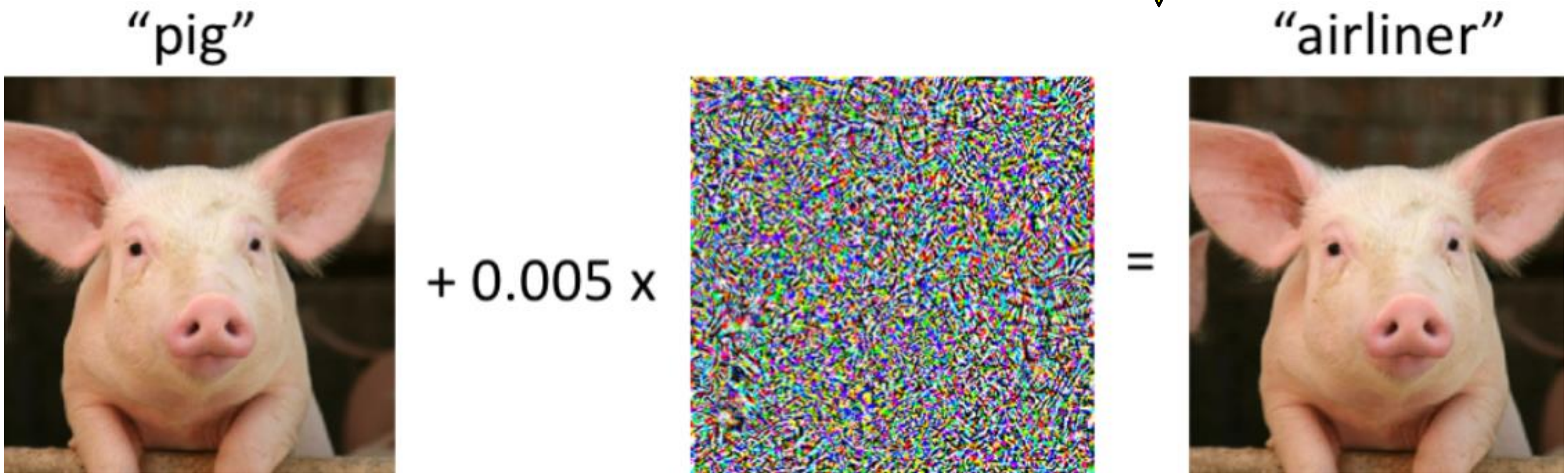
"pig"

+ 0.005 x

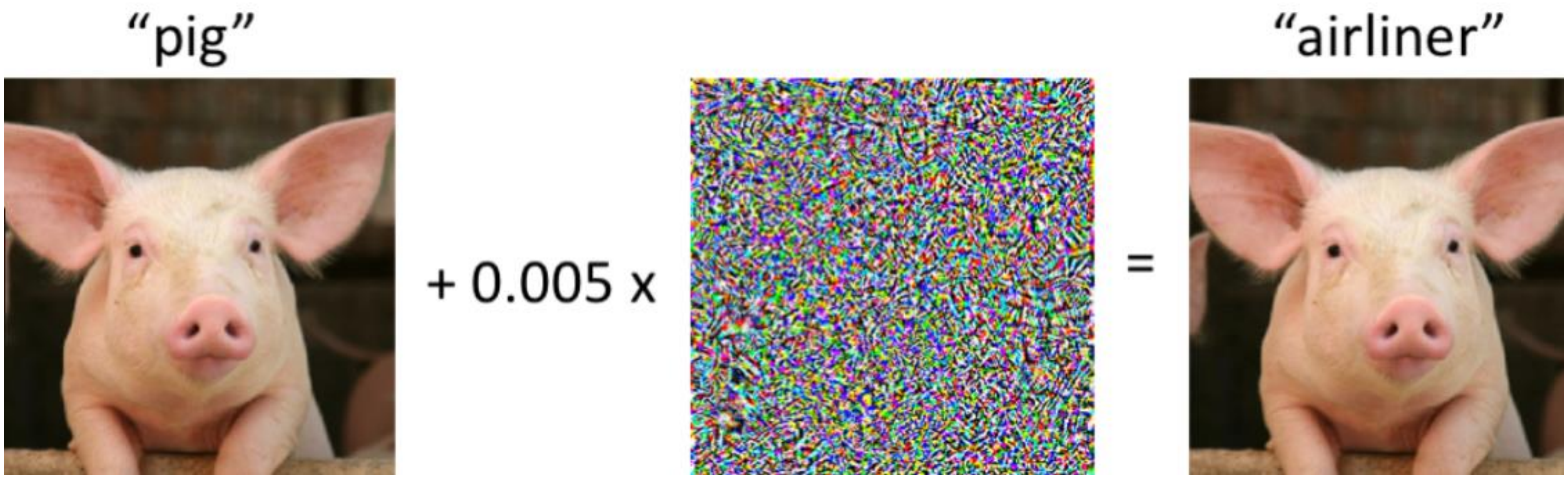


=

How AI See: An Example with Adversarial Attacks



How AI See: An Example with Adversarial Attacks

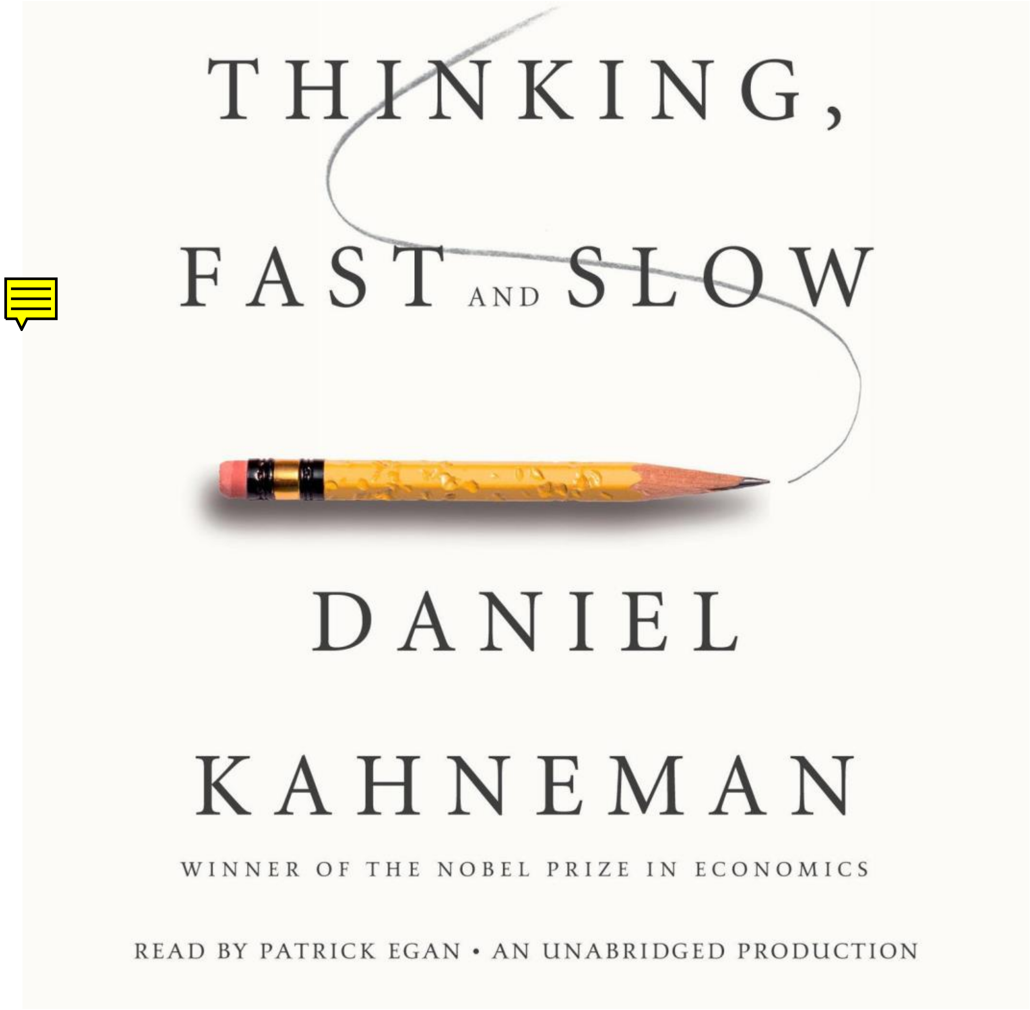


IMPORTANT

Remember Artificial Intelligence is not Human Intelligence

Machine Learning is not equal to Human Learning

Human vs machine learning

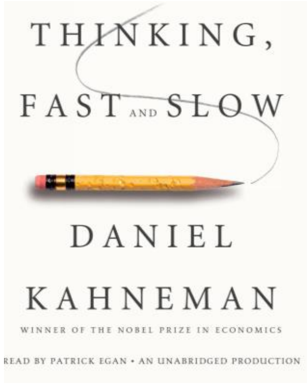


Human Cognition: System 1 vs. System 2

2 systems (and categories of cognitive tasks):

System 1

- Intuitive, fast, **UNCONSCIOUS**, non-linguistic, habitual
- Current DL



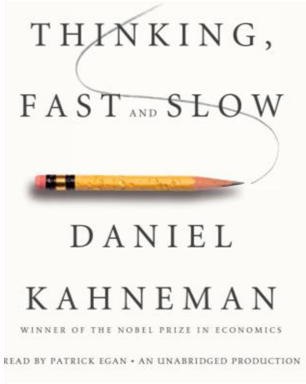
Human Cognition: System 1 vs. System 2

2 systems (and categories of cognitive tasks):

Manipulates high-level / semantic concepts, which can be recombined combinatorially

System 1

- Intuitive, fast, **UNCONSCIOUS**, non-linguistic, habitual
- Current DL



System 2

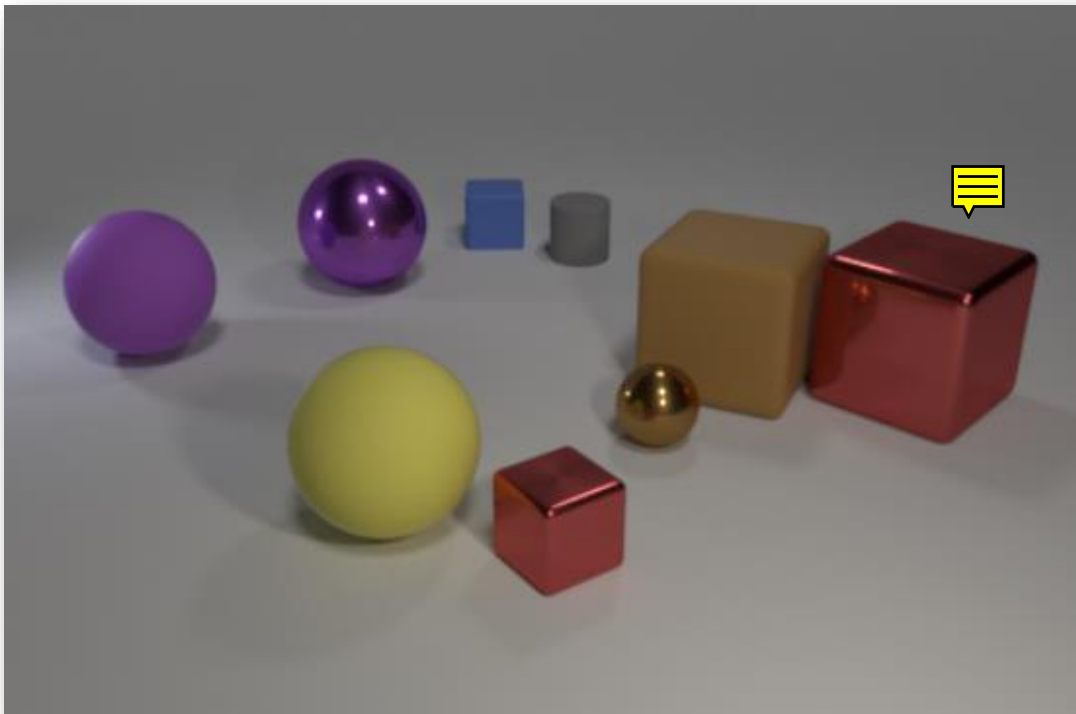
- Slow, logical, sequential, **CONSCIOUS**, linguistic, algorithmic, planning, reasoning
- Future DL

Turing Award 2018 (Hinton, Bengio, LeCunn) + most significant works at AAAI 2020:

Machine Learning of data structures: Capsule Networks, Neuro-Syntactic Machine Learning, Concept Reasoning, Experience Grounds, Logic Rules, ...

Y. Bengio

Human Cognition: System 1 vs. System 2

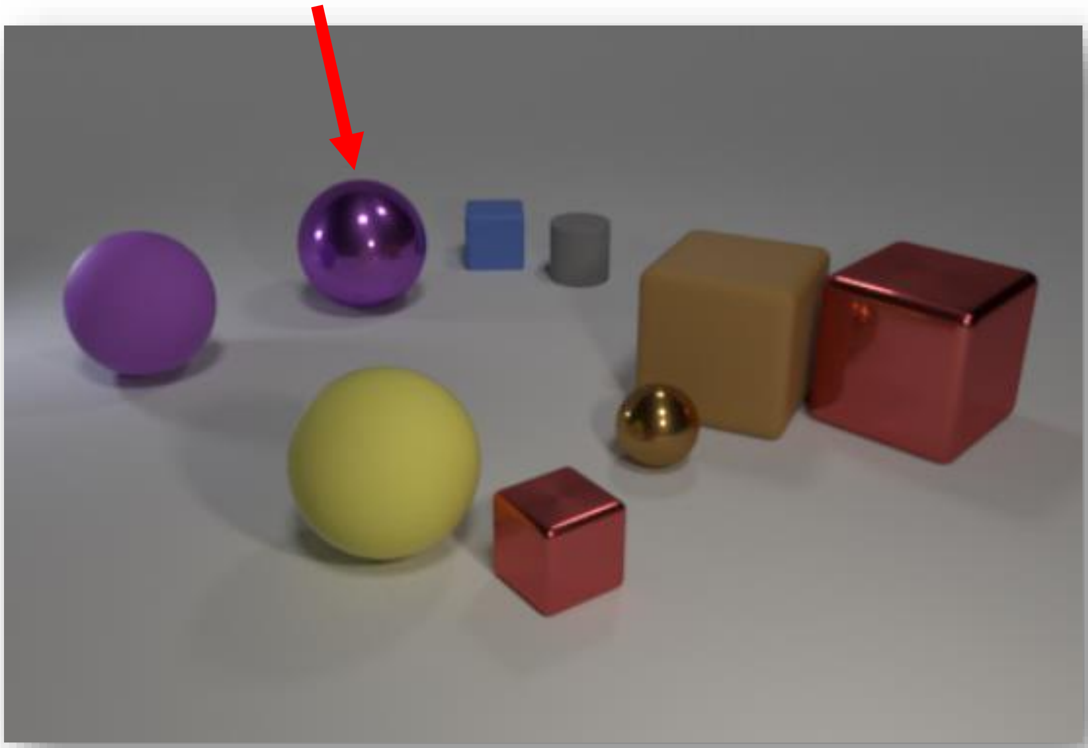


System 1:
Shape recognition, colors, positions

System 2:
Predicting interactions

Human Cognition: System 1 vs. System 2

Can you predict the events after hitting this object?



System 1:
Shape recognition, colors, positions

System 2:
Predicting interactions

Learning Task

<u>DEFINITION:</u> (Tom Mitchel 1998) 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.

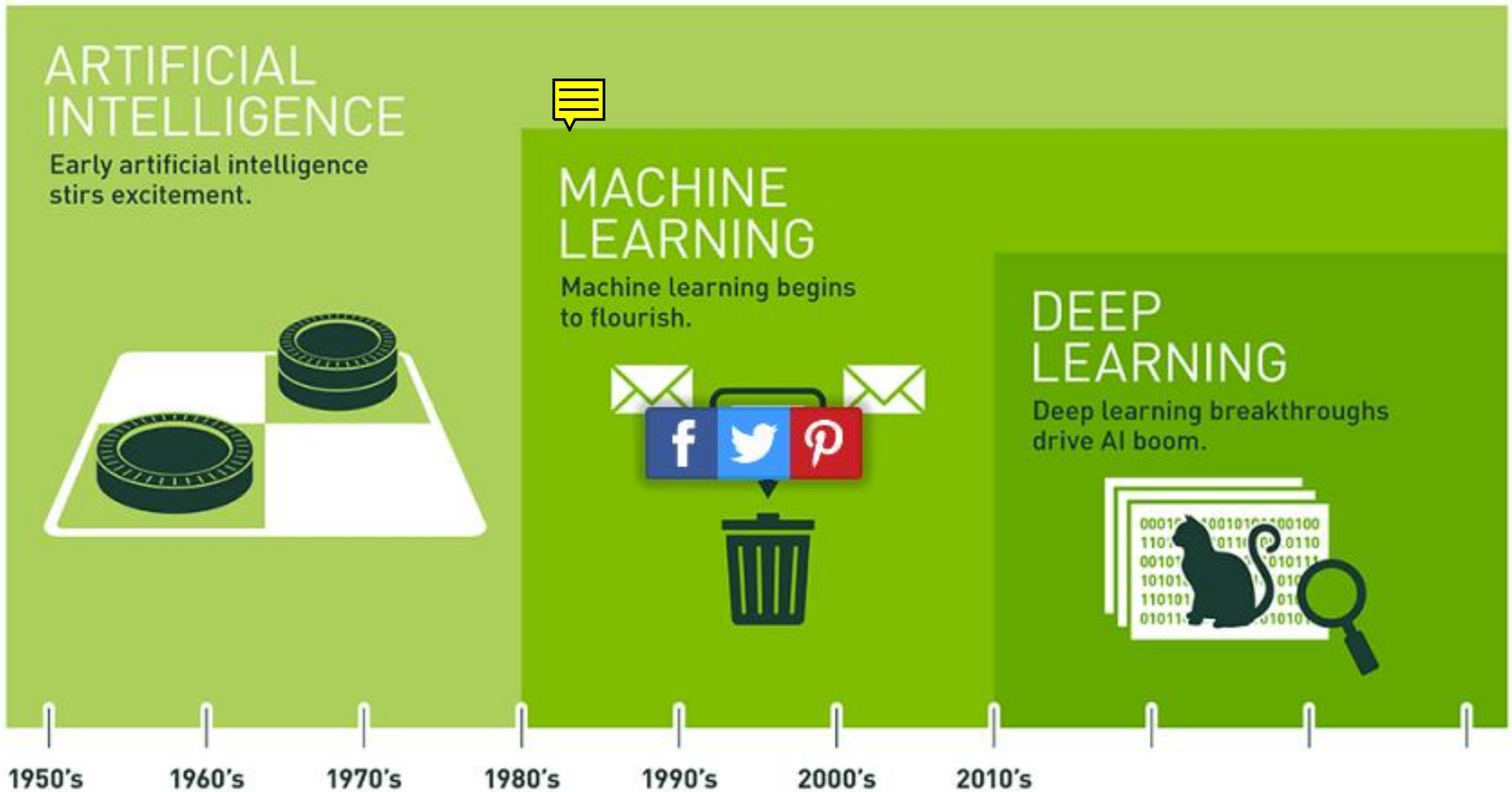
T: Playing checkers
P: Percentage of games won against an arbitrary opponent
E: Playing practice games against itself

T: Recognizing hand-written words
P: Percentage of words correctly classified
E: Database of human-labeled images of handwritten words

T: Driving on four-lane highways using vision sensors
P: Average distance traveled before a human-judged error
E: A sequence of images and steering commands recorded while observing a human driver.

T: Categorize email messages as spam or legitimate.
P: Percentage of email messages correctly classified.
E: Database of emails, some with human-given labels

Machine Learning in Context



- At the core of AI, **machine learning** is simply a way of achieving AI.
- Instead of hard coding SW routines with specific instructions to do a particular task, **ML** is a way of "training" an algorithm so that it can learn how. "Training" involves feeding huge amounts of data to the algorithm and allowing the algorithm to adjust itself and improve.

Human vs machine learning

Designing a Learning System

- Many learning methods involve training
- Training is the acquisition of knowledge, skills, and competencies as a result of the teaching of vocational or practical skills and knowledge that relate to specific useful competence.
- Training requires scenarios or examples (data)

- Unsupervised learning: no feedback
- Supervised learning: uses a series of labelled examples with direct feedback
- Reinforcement learning: indirect feedback, after many examples

Supervised vs Unsupervised

Supervised versus Unsupervised learning:

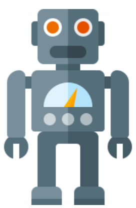
- Learn an unknown function $y_\theta(\mathbf{x}) = h_\theta(\mathbf{x})$, where:
  - $\mathbf{x}$ is an input example and
  - y is the desired output.

- Supervised learning implies we are given a training set of $(\mathbf{x},y)$ pairs by a "teacher.“

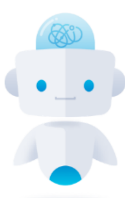
- Unsupervised learning means we are only given the $\mathbf{x}$ and some (ultimate) feedback function on our performance.

From Classical Programming to Machine Learning

Experience → **Learning** → Responses

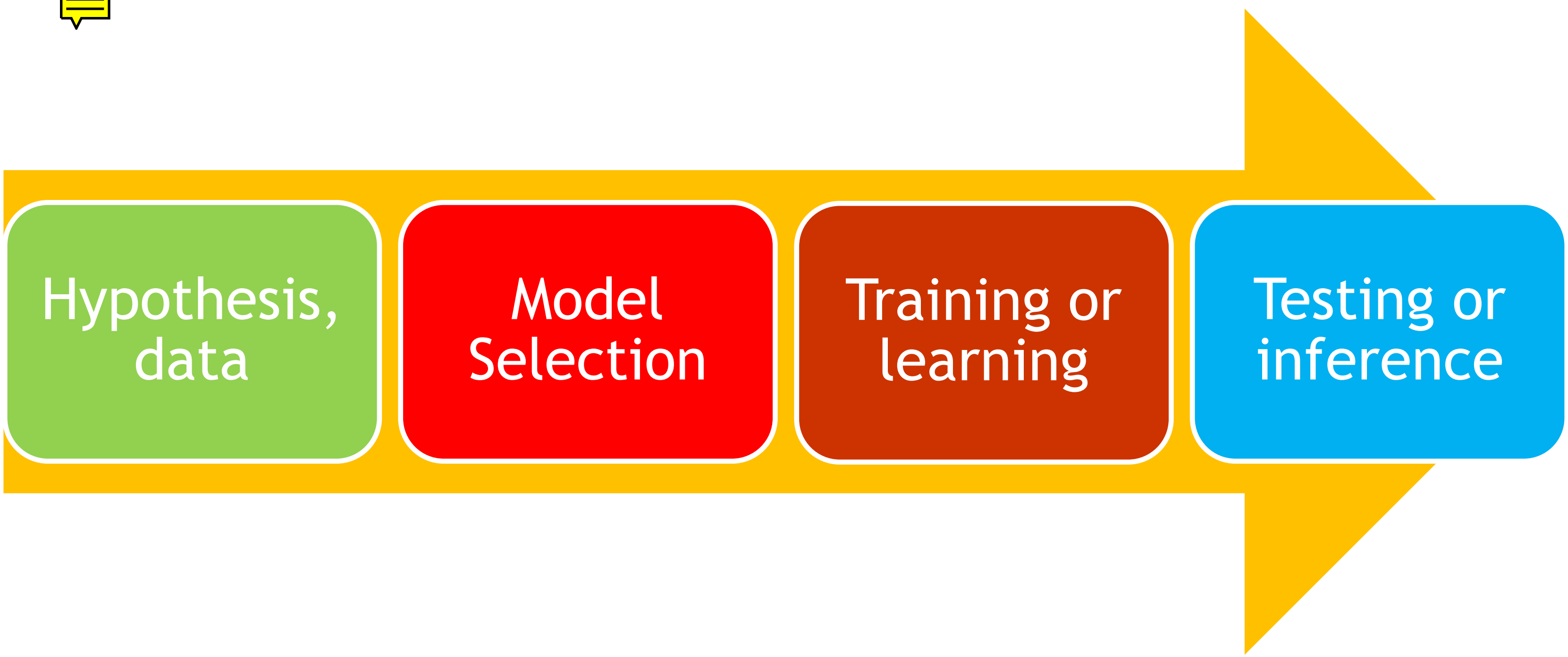


Rules → **Classic programming** → Responses
Data →



Data ($\mathbf{x}$) → **Machine Learning** → Rules ($h_\theta(\mathbf{x})$)
Responses (y) →

Designing a Learning Algorithm: Stages



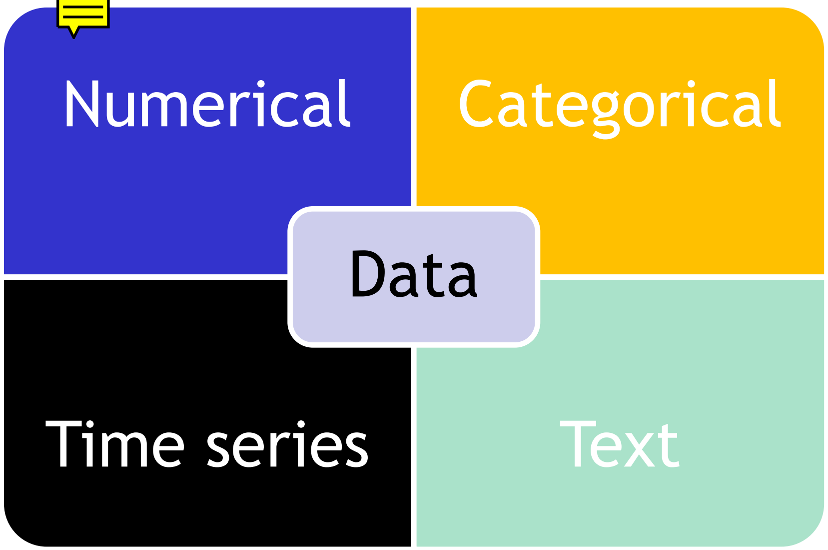
Designing a Learning Algorithm: Stages

Designing a Learning Algorithm

1. **Data:** $\mathbf{x}_n = (x_{n1} \dots x_{nD})^T$ and labels y_n (desirable output, only supervised learning)

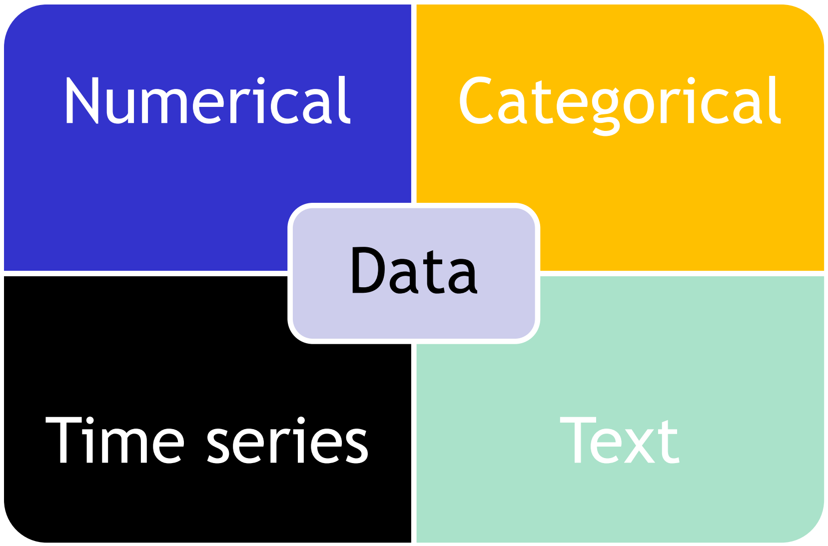
Designing a Learning Algorithm

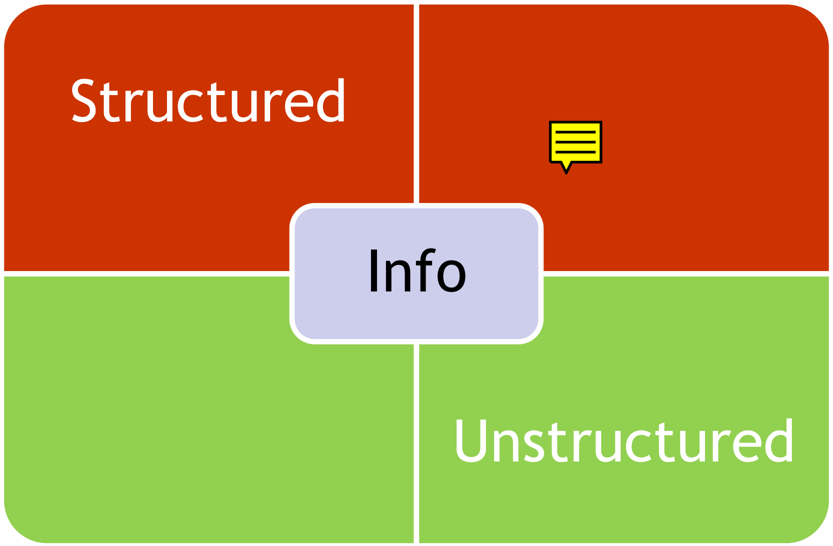
1. **Data:** $\mathbf{x}_n = (x_{n1} \dots x_{nD})^T$ and labels y_n (desirable output, only supervised learning)



Designing a Learning Algorithm

1. **Data:** $\mathbf{x}_n = (x_{n1} \ldots x_{nD})^T$ and labels y_n (desirable output, only supervised learning)

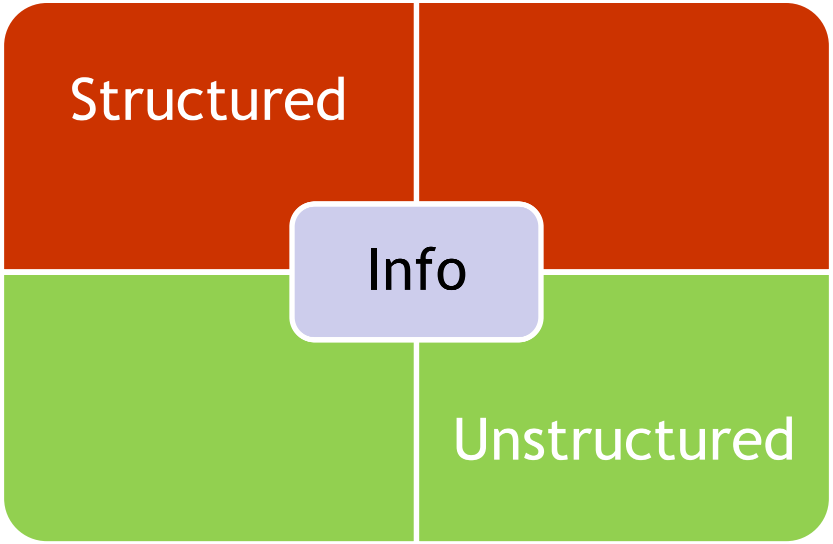




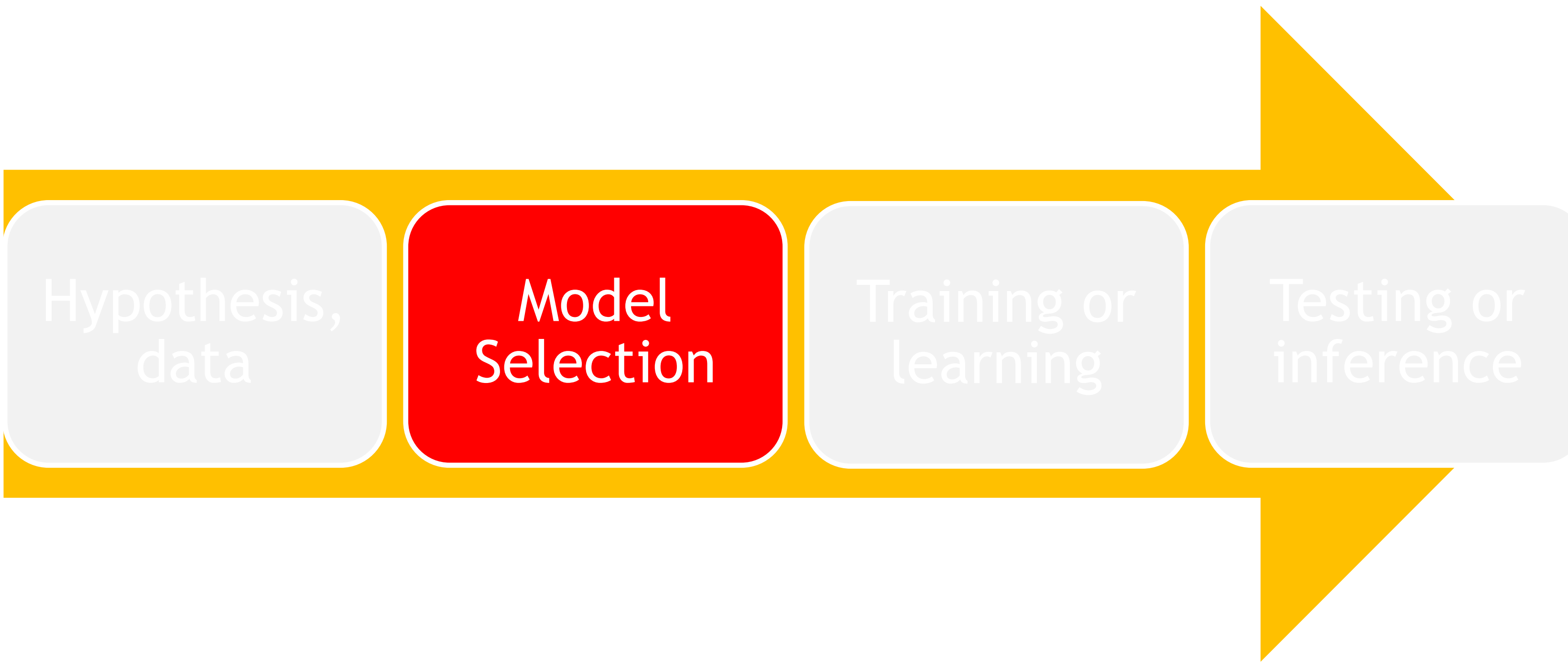
Designing a Learning Algorithm

1. **Data:** $\mathbf{x}_n = (x_{n1} \dots x_{nD})^T$ and labels y_n (desirable output, only supervised learning)

- Structured Information:
  - User profiling
  - Meteorological data
  - Genetic data
- Unstructured Information:
  - Images
  - Audio
  - Text



Designing a Learning Algorithm: Stages



Designing a Learning Algorithm

1. **Data:** $\mathbf{x}_n = (x_{n1} \dots x_{nD})^T$ and labels y_n (desirable output, only supervised learning)
2. Model Selection:
   - Select the model $h_\theta(\mathbf{x})$:

$$y(\mathbf{x}) = h(\mathbf{x}) \to y_\theta(\mathbf{x}) = h_\theta(\mathbf{x}) = h(\mathbf{x}) + \varepsilon$$

Desired output

Ideal Model

Output of our model

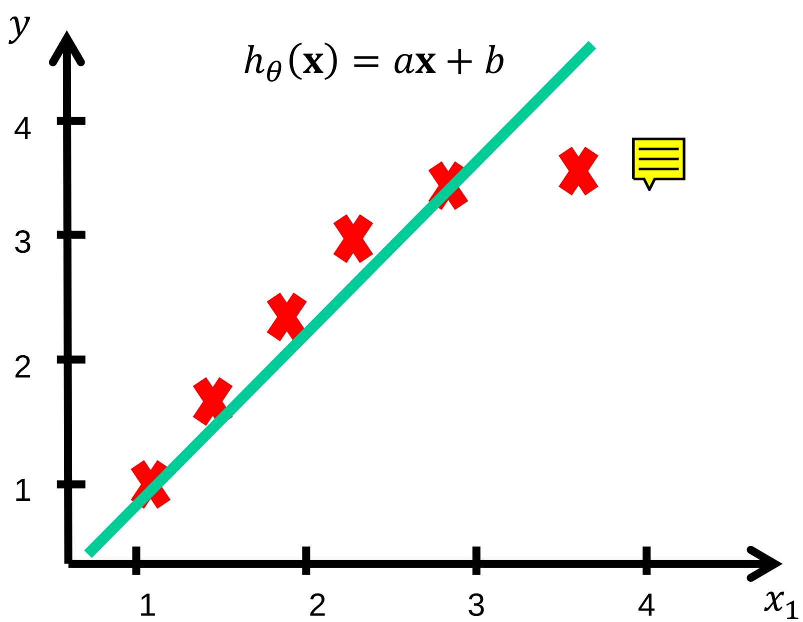
Our Model

Error between real and ideal models

Designing a Learning Algorithm

1. **Data:** $\mathbf{x}_n = (x_{n1} \dots x_{nD})^T$ and labels y_n (desirable output, only supervised learning)
2. Model Selection:
   - Select the model $h_\theta(\mathbf{x})$:

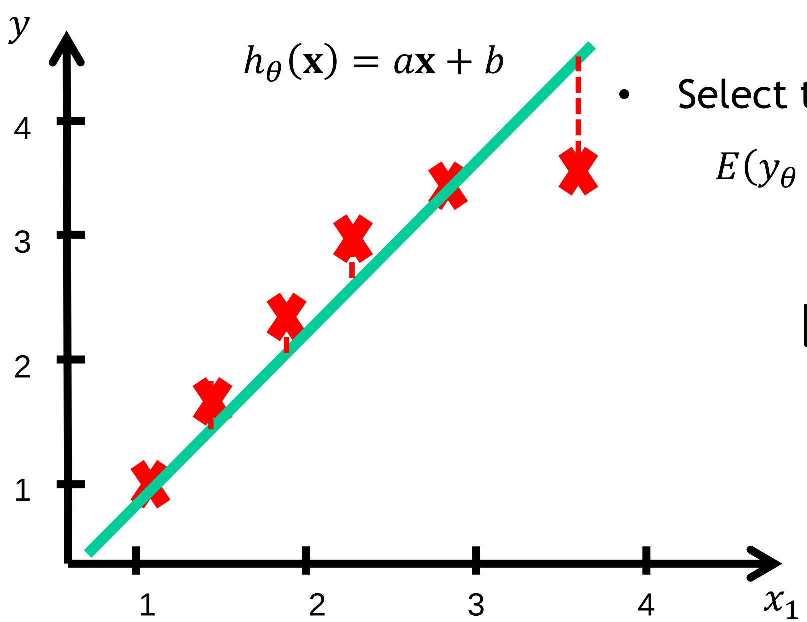
$$y(\mathbf{x}) = h(\mathbf{x}) \to y_\theta(\mathbf{x}) = h_\theta(\mathbf{x}) = h(\mathbf{x}) + \varepsilon$$



Designing a Learning Algorithm

1. **Data:** $\mathbf{x}_n = (x_{n1} \ldots x_{nD})^T$ and labels y_n (desirable output, only supervised learning)
2. Model Selection:
   - Select the model $h_\theta(\mathbf{x})$:

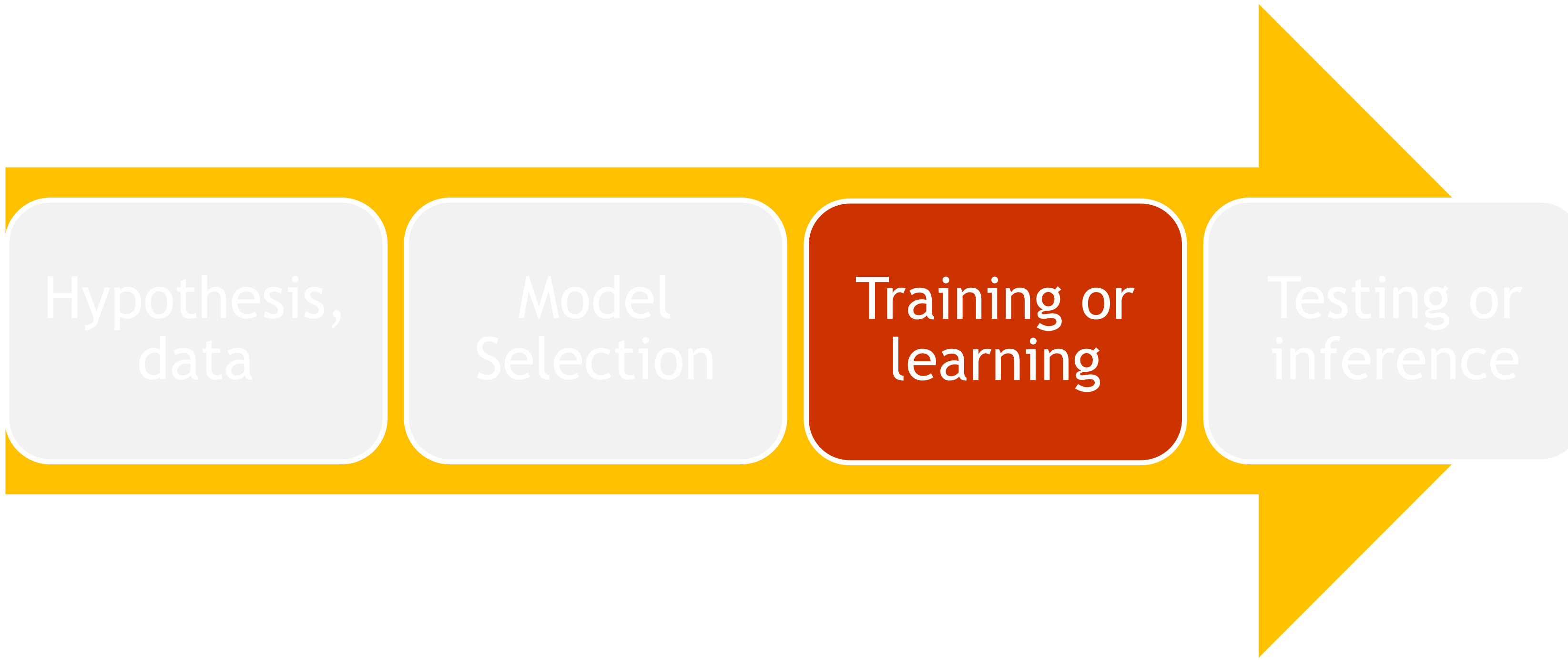
$$y(\mathbf{x}) = h(\mathbf{x}) \rightarrow y_\theta(\mathbf{x}) = h_\theta(\mathbf{x}) = h(\mathbf{x}) + \varepsilon$$



   - Select the cost function:

$$E(y_\theta - y)^2 = E[h(\mathbf{x}) + \varepsilon - h_\theta(\mathbf{x})]^2 = \underbrace{[h(\mathbf{x}) - h_\theta(\mathbf{x})]^2}_{\text{Reducible}} + \underbrace{Var(\varepsilon)}_{\text{Irreducible}}$$

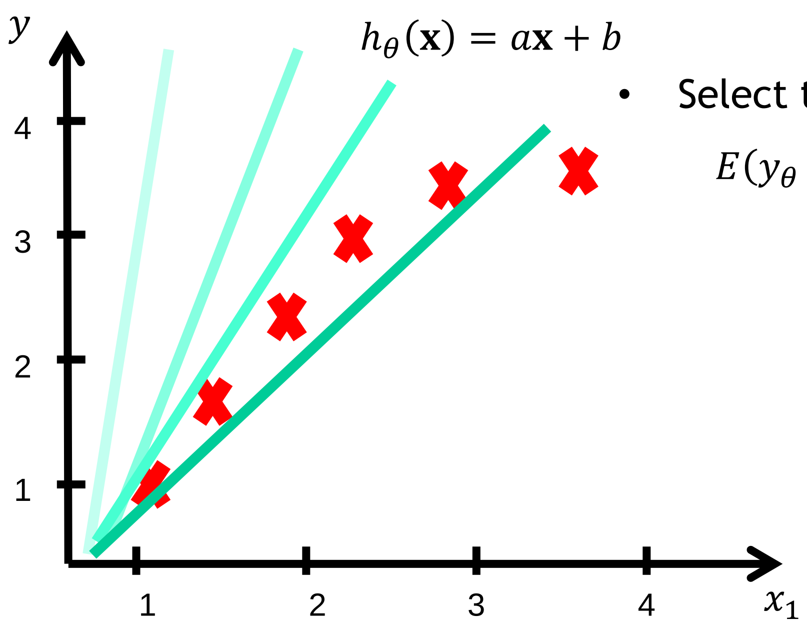
Designing a Learning Algorithm: Stages



Designing a Learning Algorithm

1. **Data:** $\mathbf{x}_n = (x_{n1} \dots x_{nD})^T$ and labels y_n (desirable output, only supervised learning)
2. Model Selection:
   - Select the model $h_\theta(\mathbf{x})$:

$$y(\mathbf{x}) = h(\mathbf{x}) \rightarrow y_\theta(\mathbf{x}) = h_\theta(\mathbf{x}) = h(\mathbf{x}) + \varepsilon$$



   - Select the cost function:

$$E(y_\theta - y)^2 = E[h(\mathbf{x}) + \varepsilon - h_\theta(\mathbf{x})]^2 = \underbrace{[h(\mathbf{x}) - h_\theta(\mathbf{x})]^2}_{\text{Reducible}} + \underbrace{Var(\varepsilon)}_{\text{Irreducible}}$$

3. Training or Learning
   - Find the set of parameters θ that optimize the cost function.

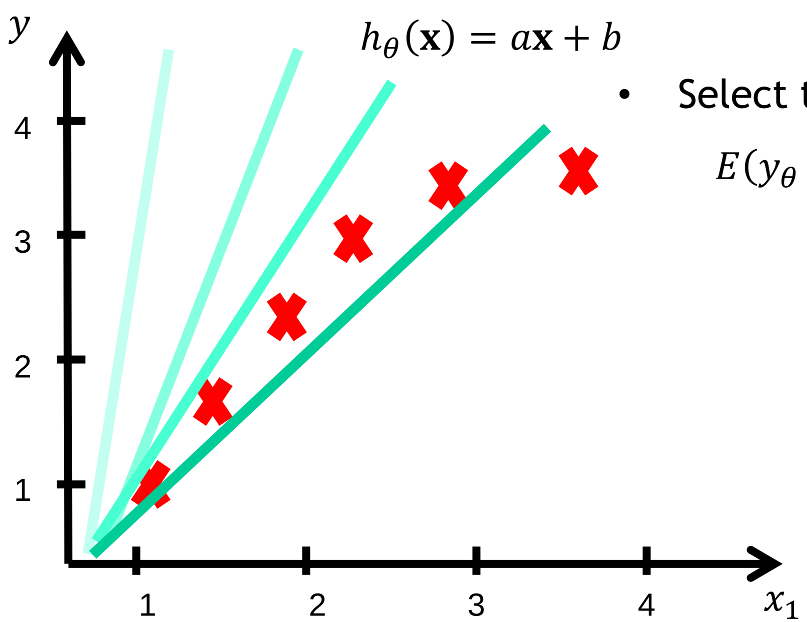
Designing a Learning Algorithm: Stages

Designing a Learning Algorithm

1. **Data:** $\mathbf{x}_n = (x_{n1} \dots x_{nD})^T$ and labels y_n (desirable output, only supervised learning)
2. Model Selection:
   - Select the model $h_\theta(\mathbf{x})$:

$$y(\mathbf{x}) = h(\mathbf{x}) \rightarrow y_\theta(\mathbf{x}) = h_\theta(\mathbf{x}) = h(\mathbf{x}) + \varepsilon$$

$$h_\theta(\mathbf{x}) = a\mathbf{x} + b$$

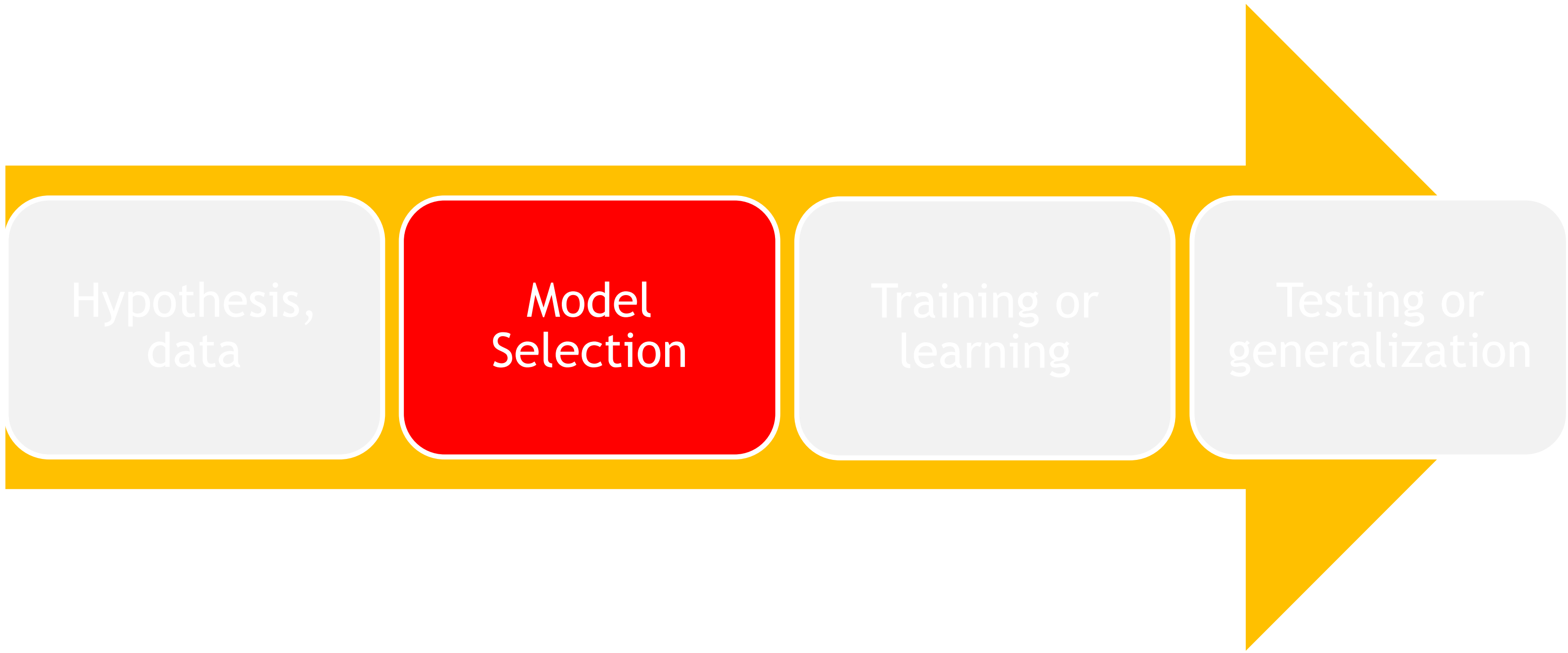


   - Select the cost function:

$$E(y_\theta - y)^2 = E[h(\mathbf{x}) + \varepsilon - h_\theta(\mathbf{x})]^2 = \underbrace{[h(\mathbf{x}) - h_\theta(\mathbf{x})]^2}_{\text{Reducible}} + \underbrace{Var(\varepsilon)}_{\text{Irreducible}}$$

3. Training or Learning
   - Find the set of parameters θ that optimize the cost function.
4. Test or Inference
   - Apply the learned model h_θ to unseen data during the training. Predict new $\hat{y}(\mathbf{x})$ for new $\mathbf{x}$.

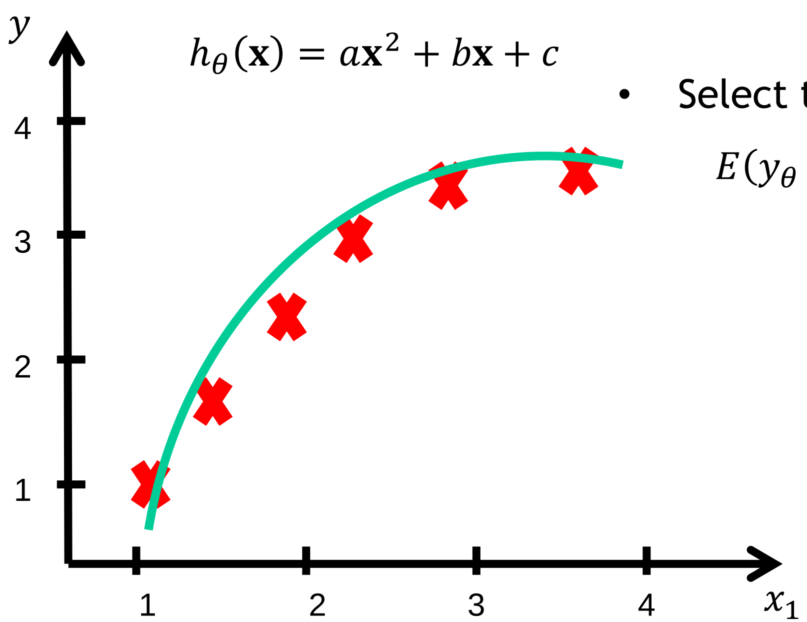
Designing a Learning Algorithm: Stages



Designing a Learning Algorithm

1. **Data:** $\mathbf{x}_n = (x_{n1} \dots x_{nD})^T$ and labels y_n (desirable output, only supervised learning)
2. Model Selection:
   - Select the model $h_\theta(\mathbf{x})$:

$$y(\mathbf{x}) = h(\mathbf{x}) \rightarrow y_\theta(\mathbf{x}) = h_\theta(\mathbf{x}) = h(\mathbf{x}) + \varepsilon$$

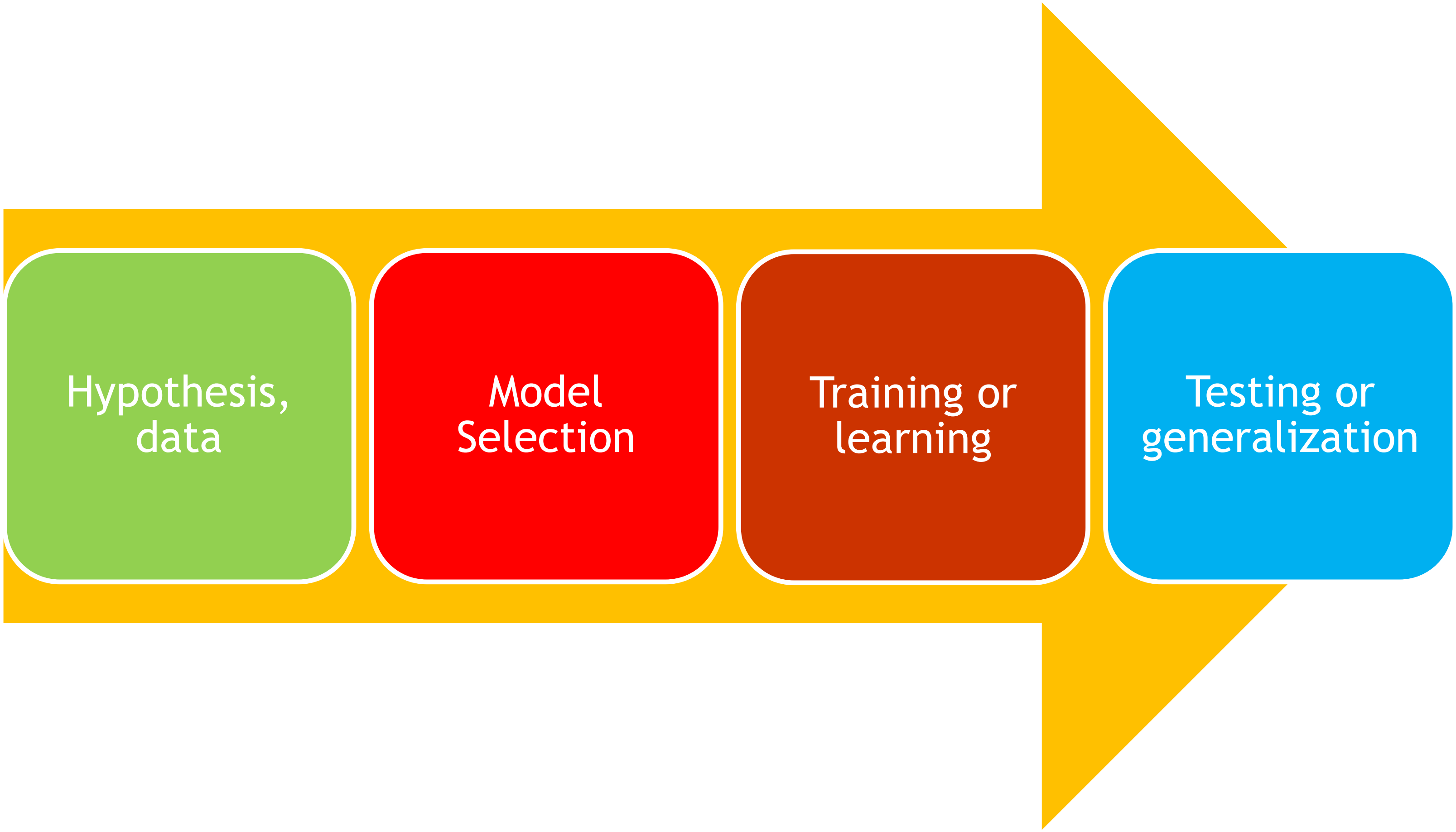


- Select the cost function:

$$E(y_\theta - y)^2 = E[h(\mathbf{x}) + \varepsilon - h_\theta(\mathbf{x})]^2 = \underbrace{[h(\mathbf{x}) - h_\theta(\mathbf{x})]^2}_{\text{Reducible}} + \underbrace{Var(\varepsilon)}_{\text{Irreducible}}$$

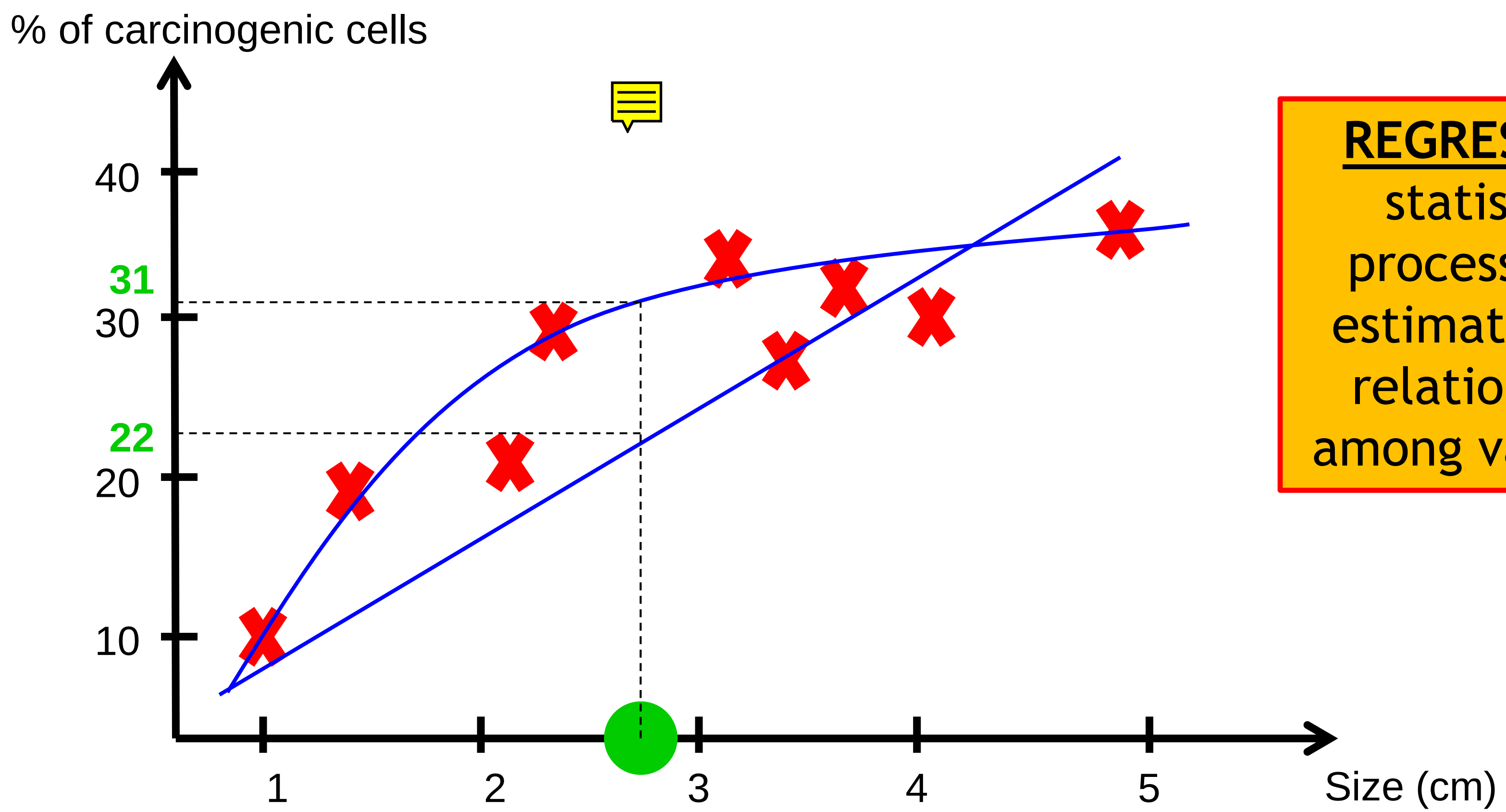
3. Training or Learning
   - Find the set of parameters θ that optimize the cost function.
4. Test or Inference
   - Apply the learned model h_θ to unseen data during the training. Predict new $\hat{y}(\mathbf{x})$ for new $\mathbf{x}$.

Designing a Learning Algorithm: Stages



Supervised Learning: Regression

Problem: tumor prediction. We want to predict the % of carcinogenic cells based on the size of a tumor

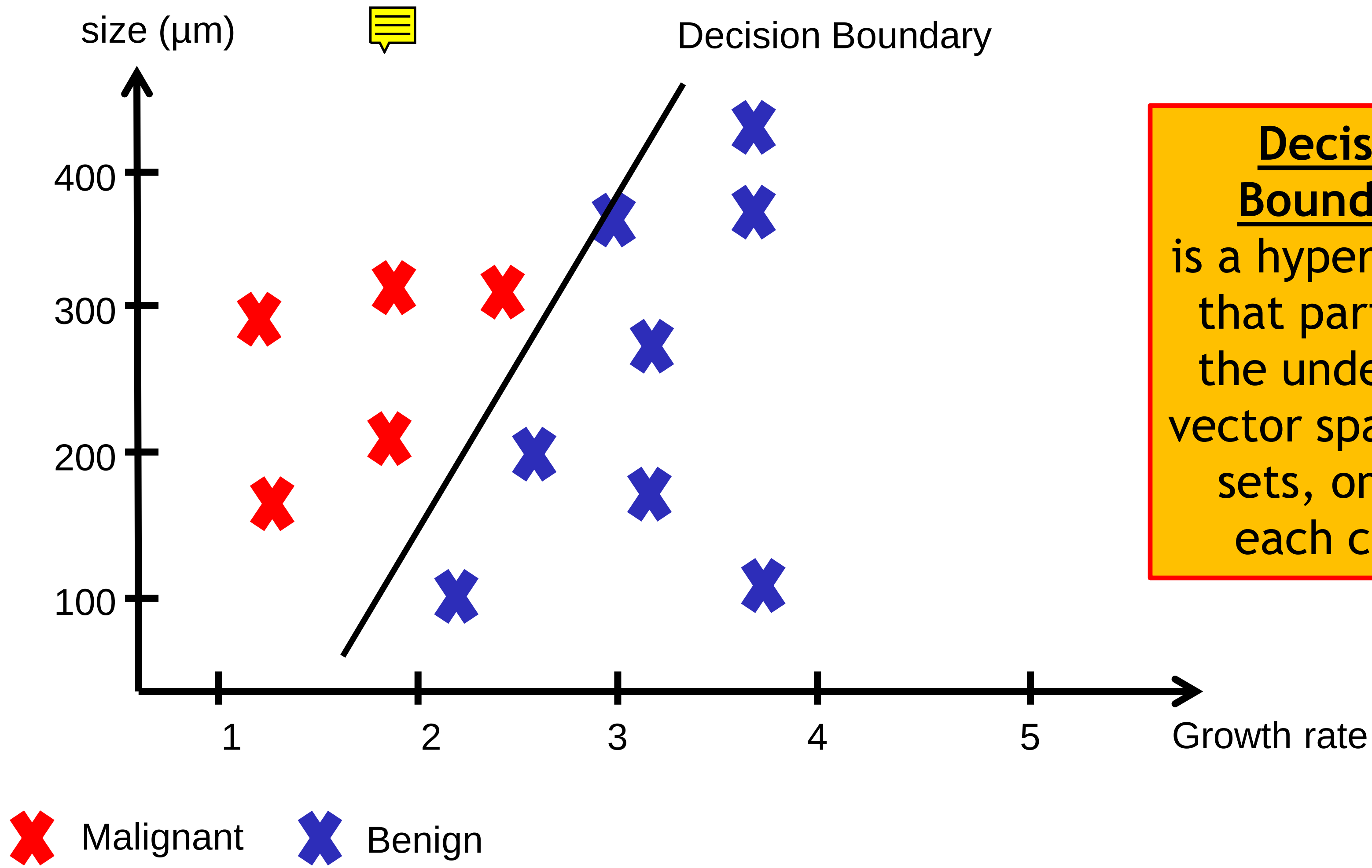


REGRESSION: statistical processes for estimating the relationships among variables

Output: a continuous value

Supervised Learning: Classification

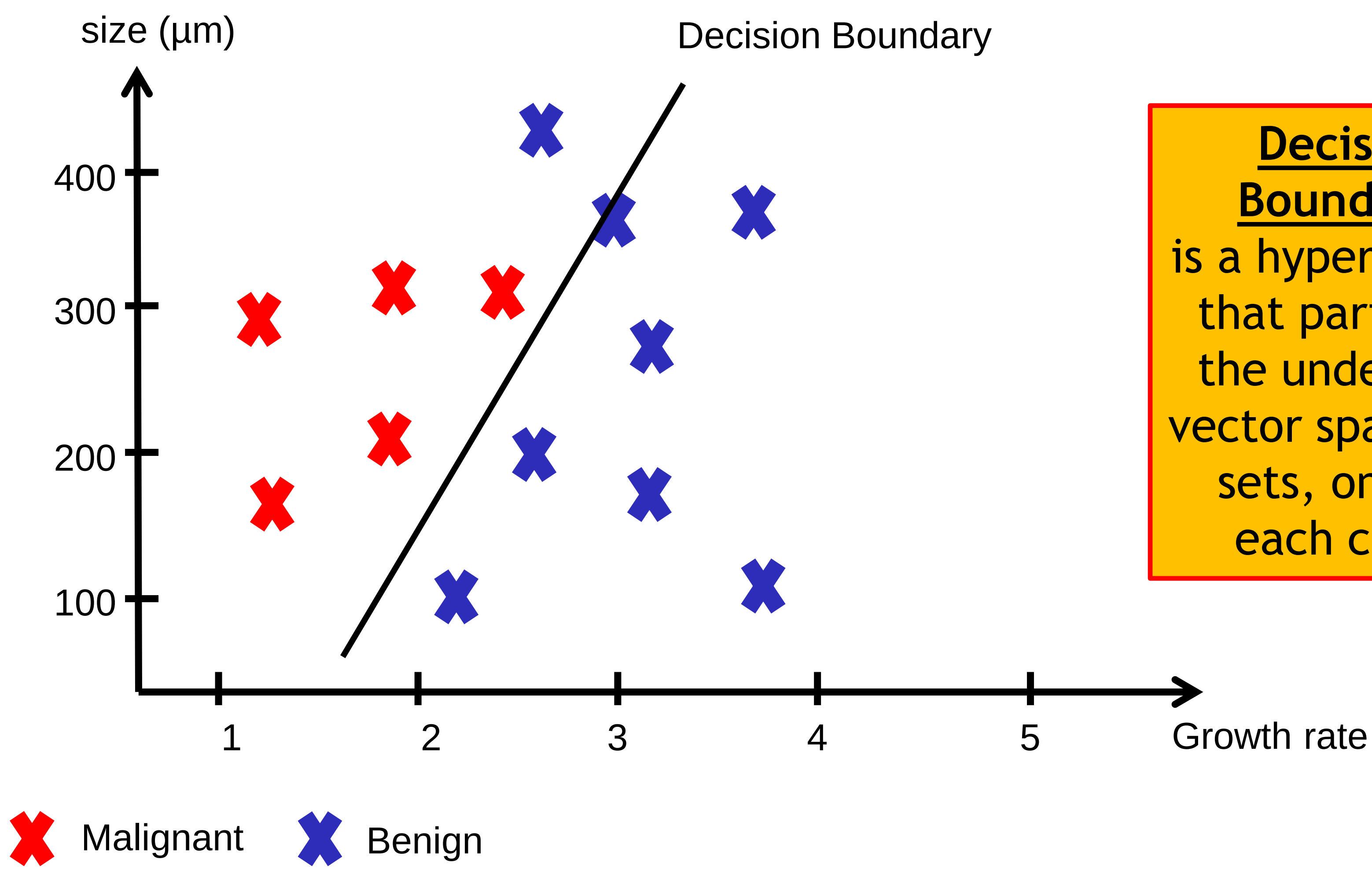
Problem: multivariate classification of tumor



Decision Boundary:
is a hypersurface that partitions the underlying vector space into sets, one for each class.

Supervised Learning: Classification

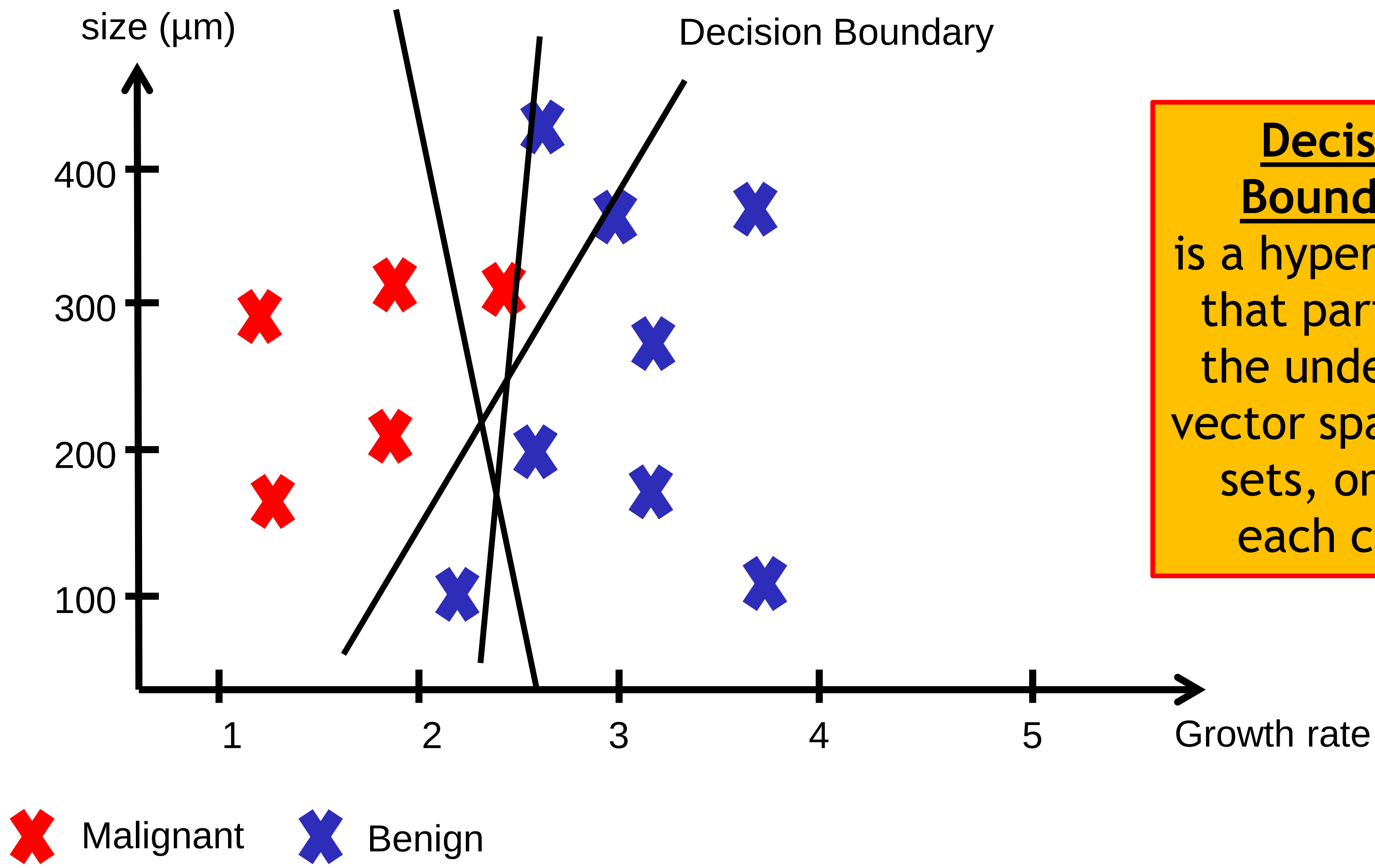
Problem: multivariate classification of tumor



<u>Decision Boundary:</u>
is a hypersurface that partitions the underlying vector space into sets, one for each class.

Supervised Learning: Classification

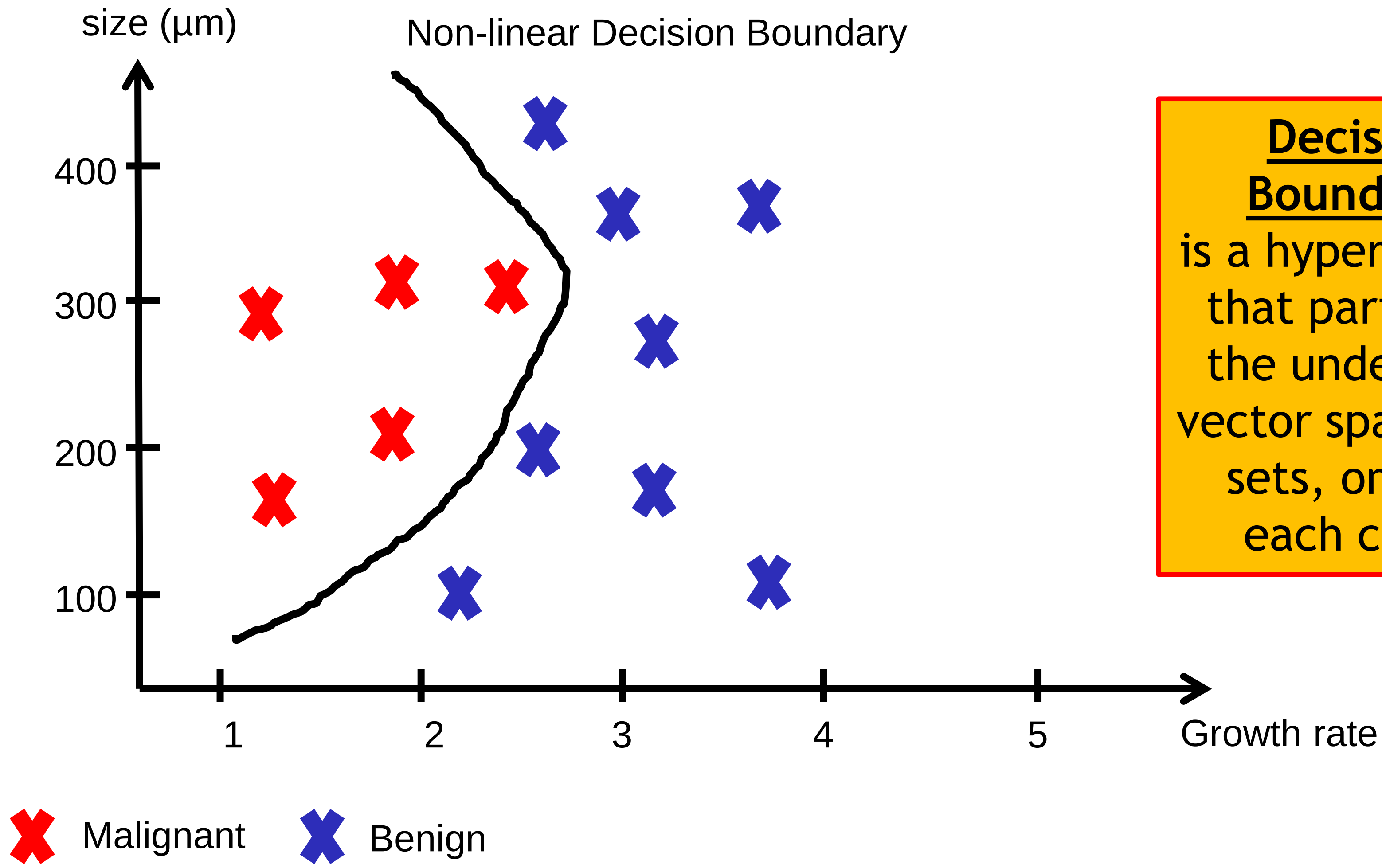
Problem: multivariate classification of tumor



<u>Decision Boundary:</u>
is a hypersurface that partitions the underlying vector space into sets, one for each class.

Supervised Learning: Classification

Problem: multivariate classification of tumor



Decision Boundary:
is a hypersurface that partitions the underlying vector space into sets, one for each class.

Supervised Learning: Classification

Example: Cancer diagnosis

Training Set

Patient ID	# of Tumors	Avg Area	Avg Density	Diagnosis
1	5	20	118	Malignant
2	3	15	130	Benign
3	7	10	52	Benign
4	2	30	100	Malignant

- Use this training set to learn how to classify patients where diagnosis is not known:

Test Set

Patient ID	# of Tumors	Avg Area	Avg Density	Diagnosis
101	4	16	95	?
102	9	22	125	?
103	1	14	80	?

Input Data / Features

Labels

- The input data is often easily obtained, whereas the labels is not.

Classification Problem

Goal: Use training set + some learning method to produce a predictive model.

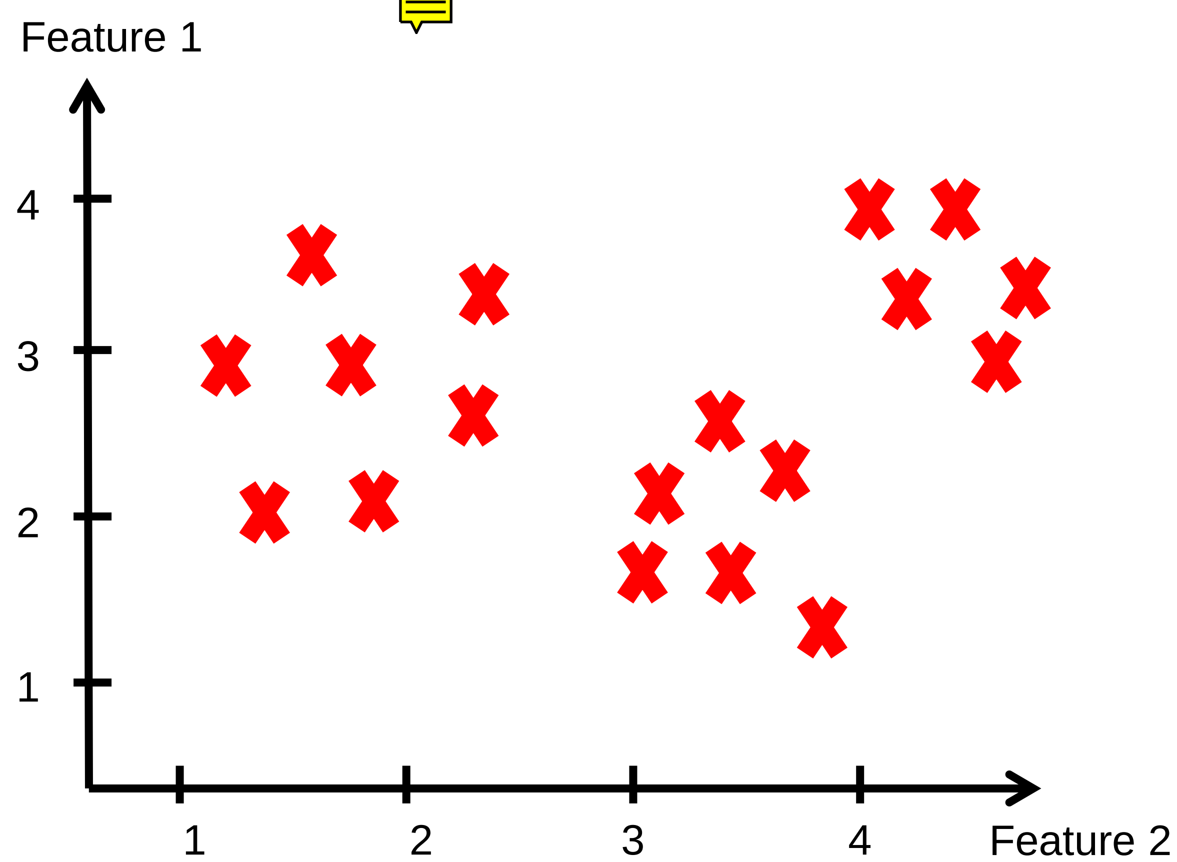
Use this predictive model to classify new data.

Sample applications:

Application	Input Data	Classification
Medical Diagnosis	Noninvasive tests	Results from invasive measurements
Optical Character Recognition	Scanned bitmaps	Letter A-Z
Protein Folding	Amino acid construction	Protein shape (helices, loops, sheets)
Research Paper Acceptance	Words in paper title	Paper accepted or rejected

Unsupervised Learning: Clustering

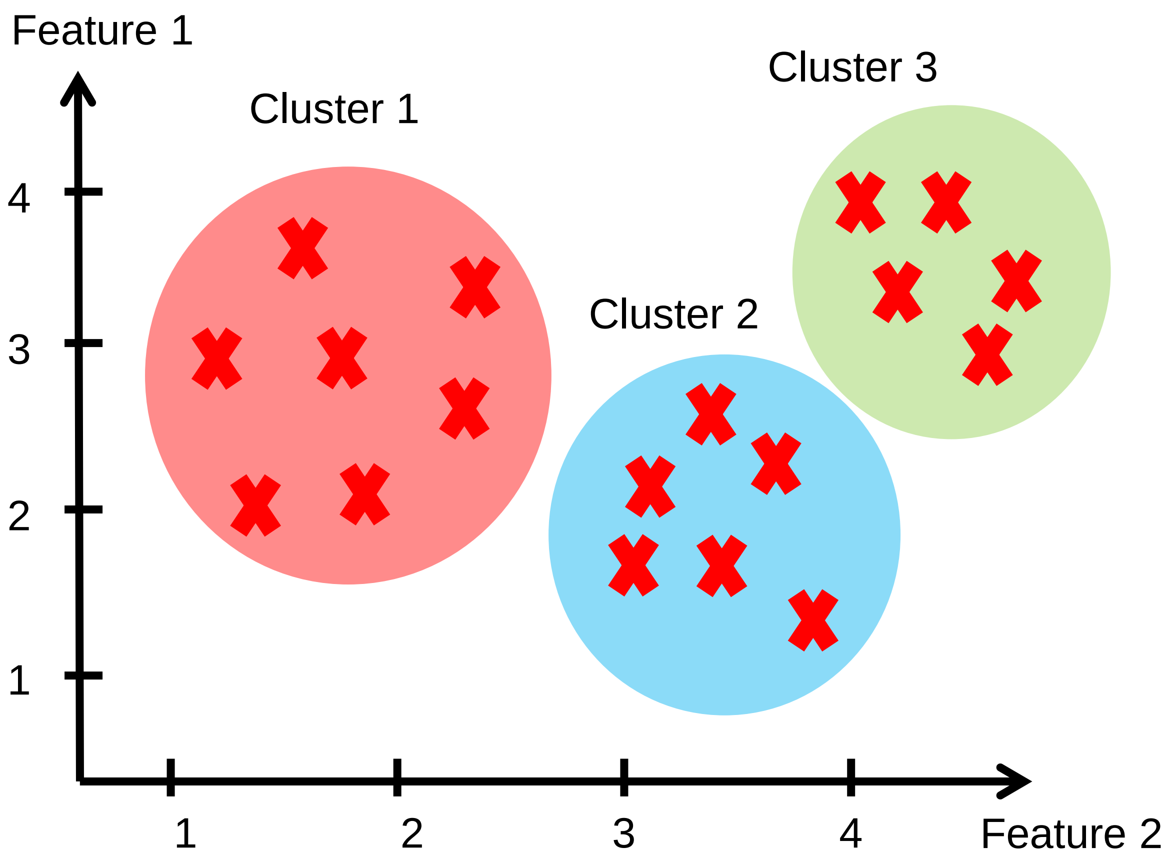
Problem: genotype-phenotype associations



Clustering:
grouping a set of objects in such a way that objects in the same group are more similar (in some sense or another) to each other than to those in other groups.

Unsupervised Learning: Clustering

Problem: genotype-phenotype associations

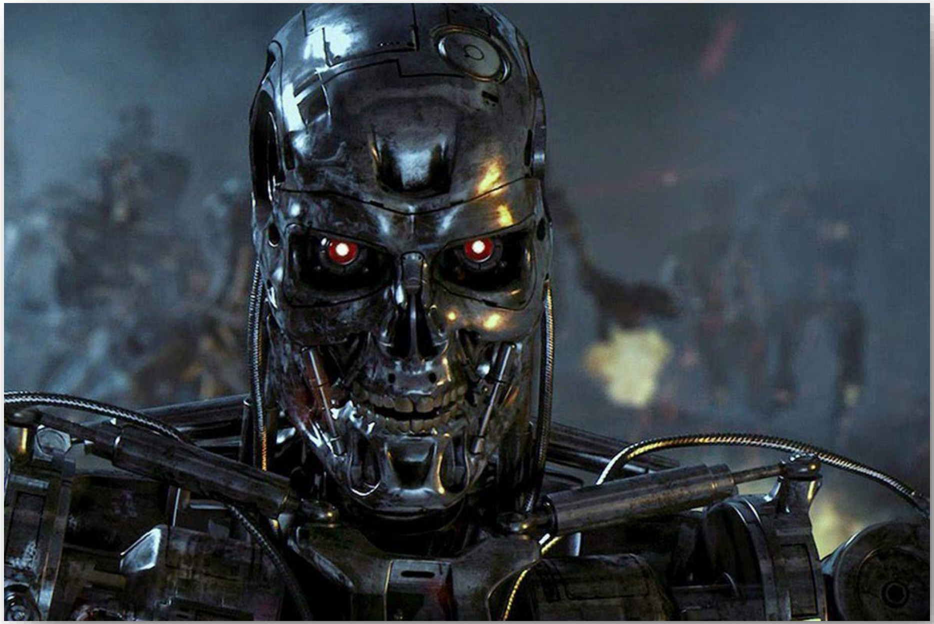


<u>Clustering:</u> grouping a set of objects in such a way that objects in the same group are more similar (in some sense or another) to each other than to those in other groups.

Fear the Machine Learning?

"The real fear of machine learning and artificial intelligence should be its ability to reflect and amplify our biases and the lack of diversity of the people creating it"

Mike King's presentation at The Inbounder New York, May 22 2017.

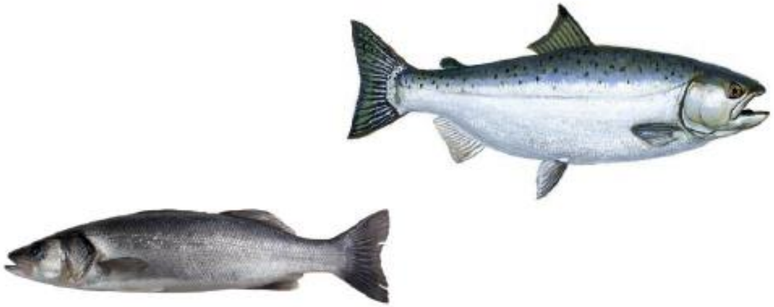


EXTRA MATERIAL

Example

Problem: sorting (classification) incoming fish on a conveyor belt according to species

- Assume that we have only two kinds of fish:
  - –Salmon
  - –Sea bass



***Adapted from Duda, Hart and Stork, Pattern Classification, 2nd Ed.**

Example

- What kind of information can distinguish one species from the other?

**Adapted from Duda, Hart and Stork, Pattern Classification, 2nd Ed.*

Example

- What kind of information can distinguish one species from the other?
  - Length, width, weight, number and shape of fins, tail shape, etc.

Example

- What kind of information can distinguish one species from the other?
  - Length, width, weight, number and shape of fins, tail shape, etc.
- What can cause problems during sensing?

**Adapted from Duda, Hart and Stork, Pattern Classification, 2nd Ed.*

Example

- What kind of information can distinguish one species from the other?
  - Length, width, weight, number and shape of fins, tail shape, etc.
- What can cause problems during sensing?
  - Lighting conditions, position of fish on the conveyor belt, camera noise, etc.

**Adapted from Duda, Hart and Stork, Pattern Classification, 2nd Ed.*

Example

- What kind of information can distinguish one species from the other?
  - Length, width, weight, number and shape of fins, tail shape, etc.
- What can cause problems during sensing?
  - Lighting conditions, position of fish on the conveyor belt, camera noise, etc.
- What are the steps in the process?

**Adapted from Duda, Hart and Stork, Pattern Classification, 2nd Ed.*

Example

- What kind of information can distinguish one species from the other?
  - Length, width, weight, number and shape of fins, tail shape, etc.
- What can cause problems during sensing?
  - Lighting conditions, position of fish on the conveyor belt, camera noise, etc.
- What are the steps in the process?
  - Capture image -> isolate fish -> take measurements -> make decision

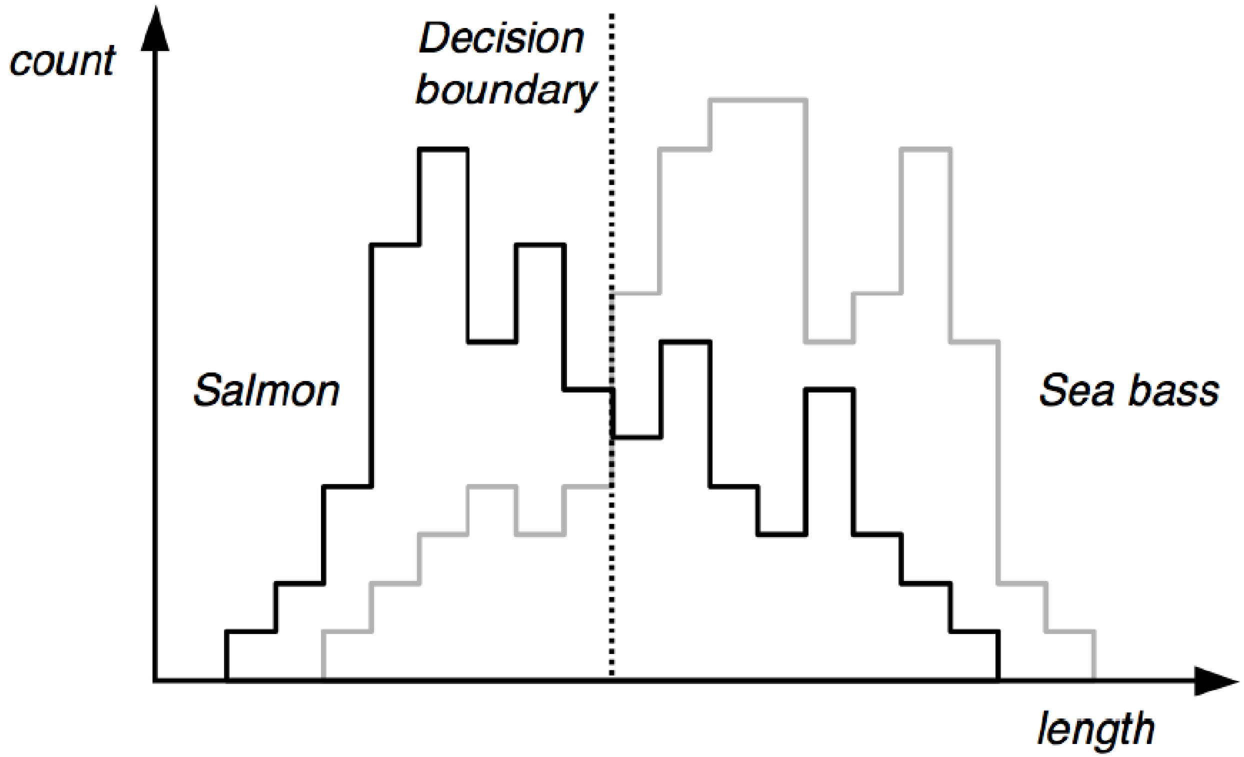
**Adapted from Duda, Hart and Stork, Pattern Classification, 2nd Ed.*

Example

- **Sensor**
  - The camera captures an image as a new fish enters the sorting area
- **Preprocessing**
  - Adjustments for average intensity levels
  - Segmentation to separate fish from background
- **Feature Extraction**
  - Assume a fisherman told us that a sea bass is generally longer than a salmon. We can use **length** as a feature and decide between sea bass and salmon according to a threshold on length.

**Adapted from Duda, Hart and Stork, Pattern Classification, 2nd Ed.*

Example



We estimate the system's probability of error and obtain a discouraging result of 40%. Can we improve this result?

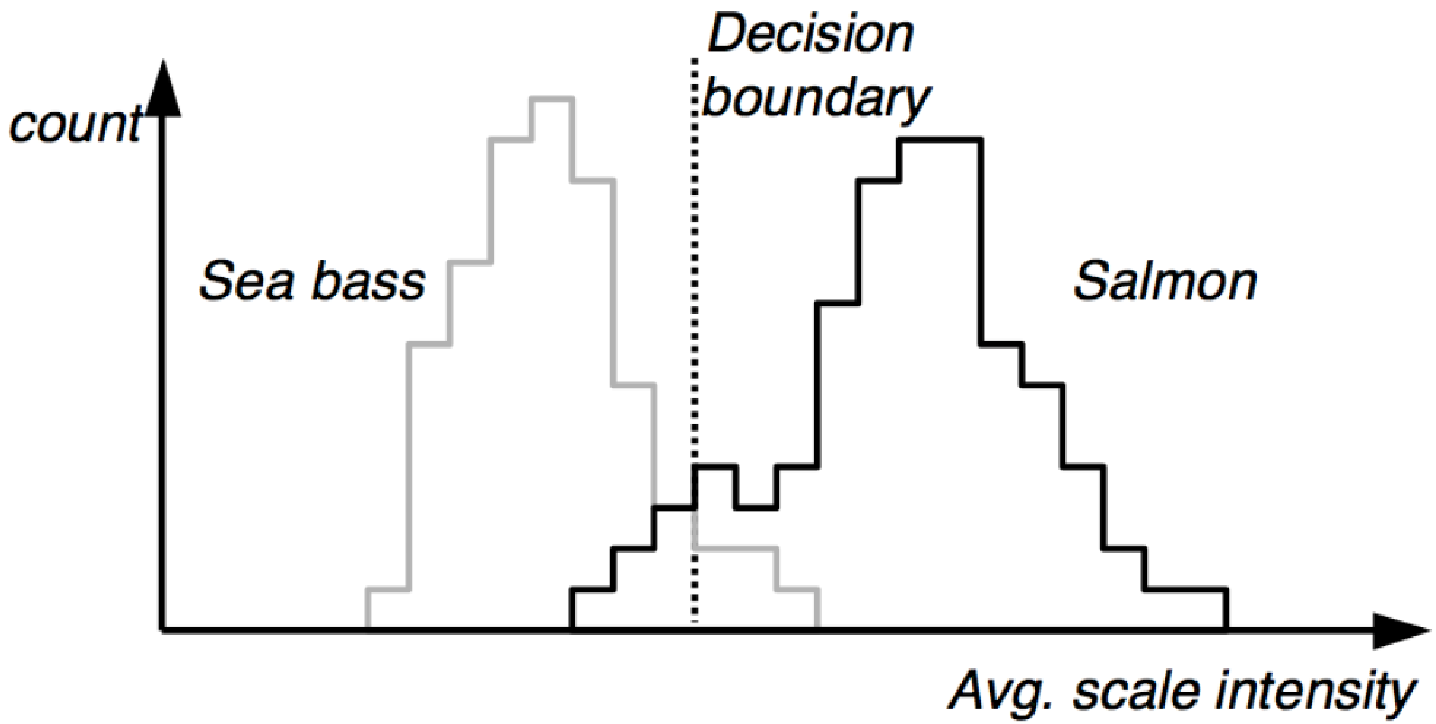
***Adapted from Duda, Hart and Stork, Pattern Classification, 2nd Ed.**

Example

- Even though sea bass is longer than salmon on the average, there are many examples of fish where this observation does not hold
- Committed to achieve a higher recognition rate, we try a number of features
  - Width, Area, Position of the eyes w.r.t. mouth...
  - only to find out that these features contain no discriminatory information
- Finally we find a "good" feature: **average intensity of the fish scales**

**Adapted from Duda, Hart and Stork, Pattern Classification, 2nd Ed.*

Example



Histogram of the lightness feature for two types of fish in **training samples**. It looks easier to choose the threshold but we still can not make a perfect decision.

**Adapted from Duda, Hart and Stork, Pattern Classification, 2nd Ed.*

Example

An example: multiple features

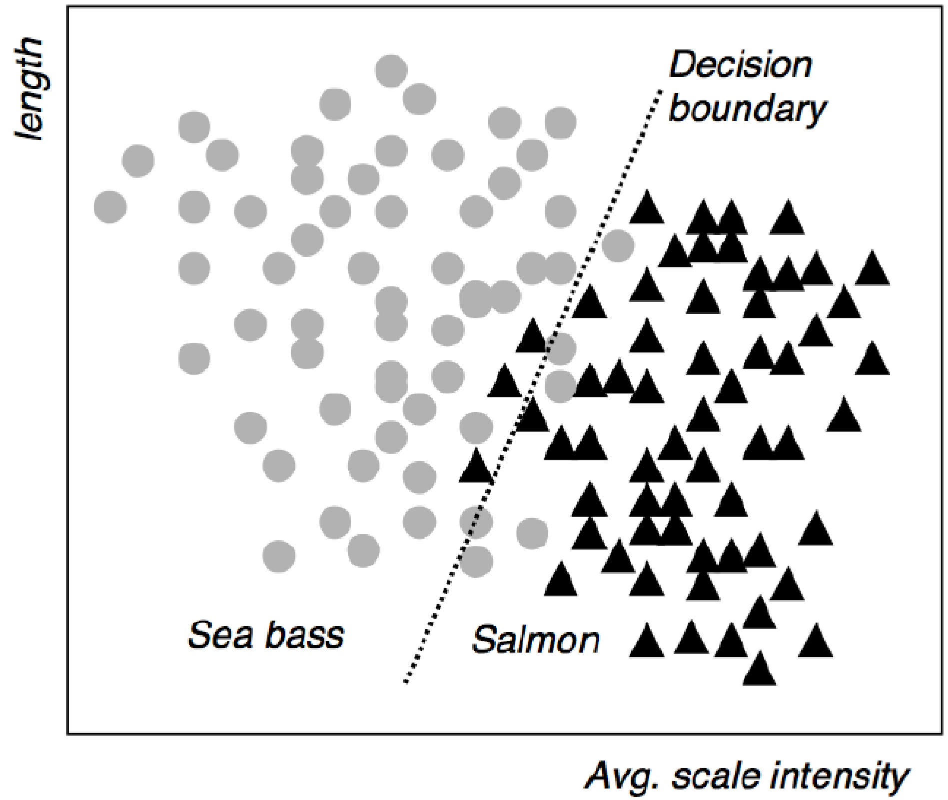
- We can use two features in our decision:
  - lightness: x_1
  - length: x_2
- Each fish image is now represented as a point (feature vector)

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

in a two-dimensional **feature space**.

**Adapted from Duda, Hart and Stork, Pattern Classification, 2nd Ed.*

Example



Scatter plot of lightness and length features for training samples. We can compute a **decision boundary** to divide the feature space into two regions with a classification rate of 95.7%.

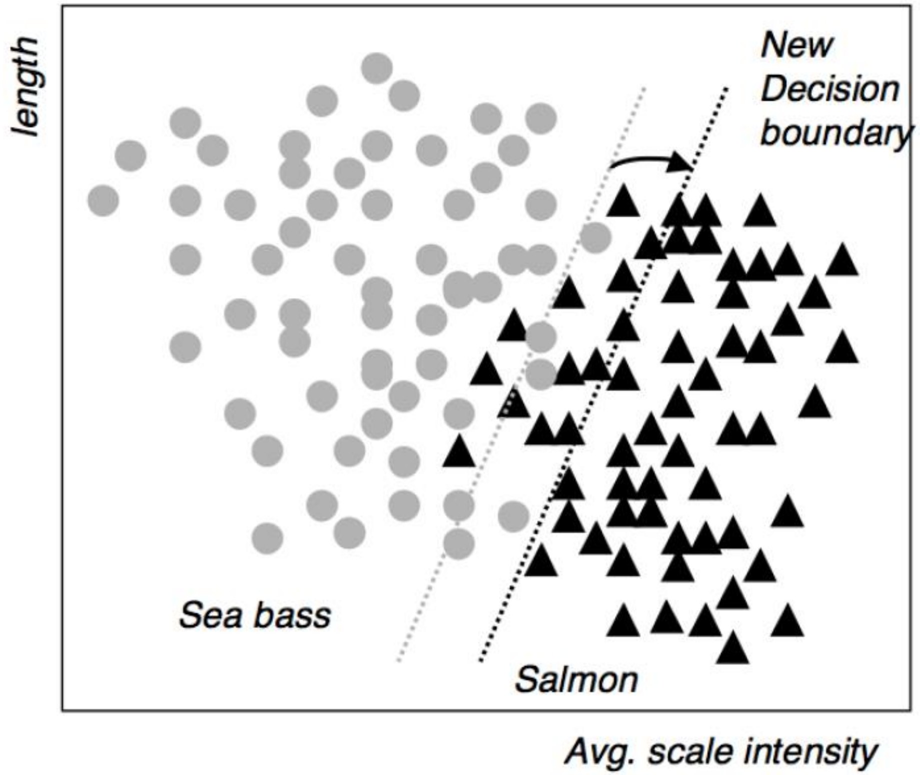
**Adapted from Duda, Hart and Stork, Pattern Classification, 2nd Ed.*

Example

- We should also consider **costs of different errors** we make in our decisions.
- For example, if the fish packing company knows that:
  – Customers who buy salmon will object vigorously if they see sea bass in their cans.
  – Customers who buy sea bass will not be unhappy if they occasionally see some expensive salmon in their cans.
- How does this knowledge affect our decision?

**Adapted from Duda, Hart and Stork, Pattern Classification, 2nd Ed.*

Example

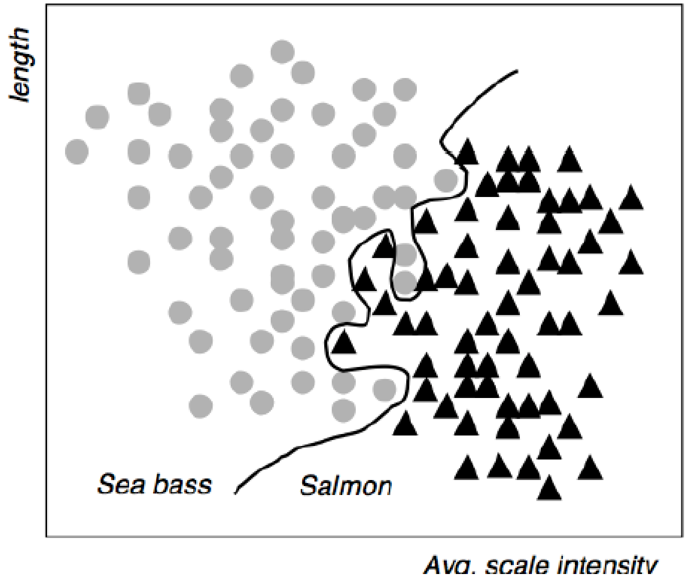


We could intuitively shift the decision boundary to minimize an alternative cost function

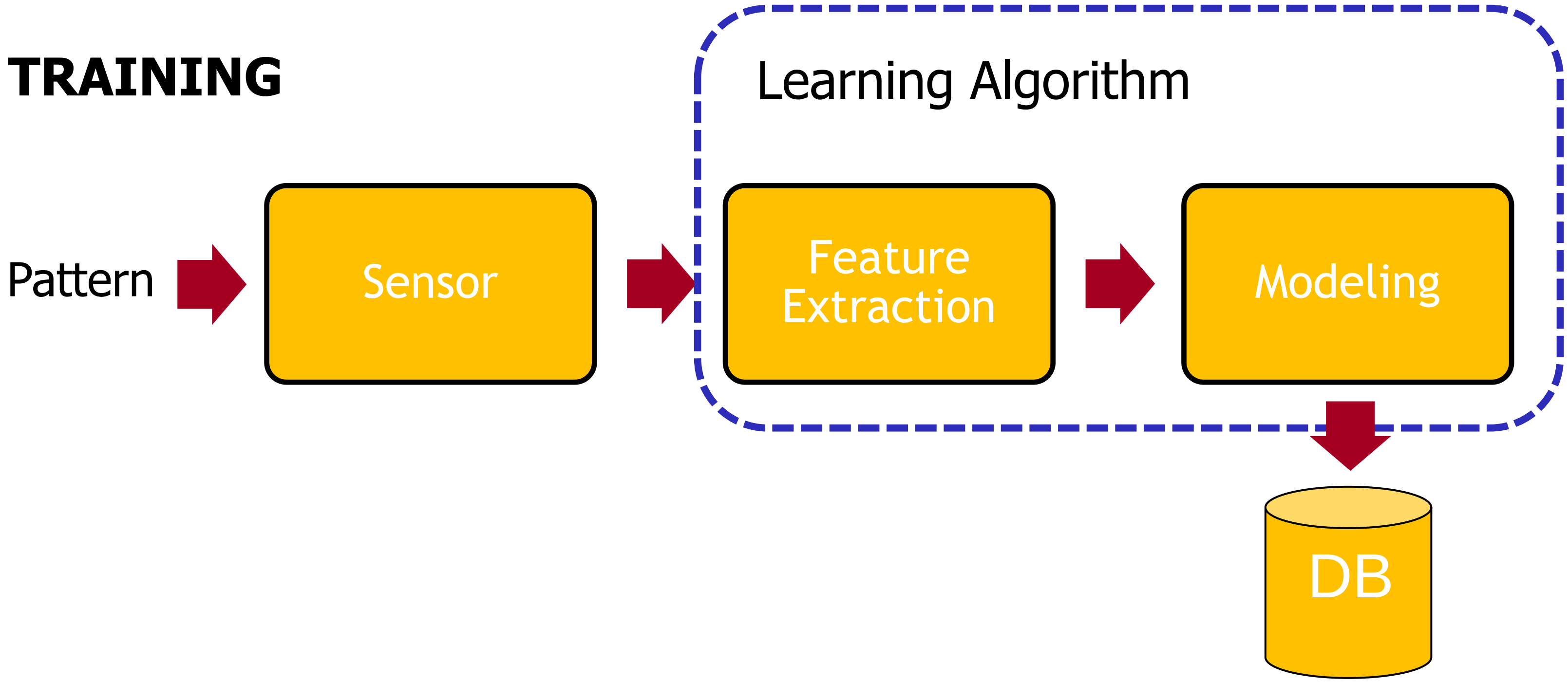
**Adapted from Duda, Hart and Stork, Pattern Classification, 2nd Ed.*

Example

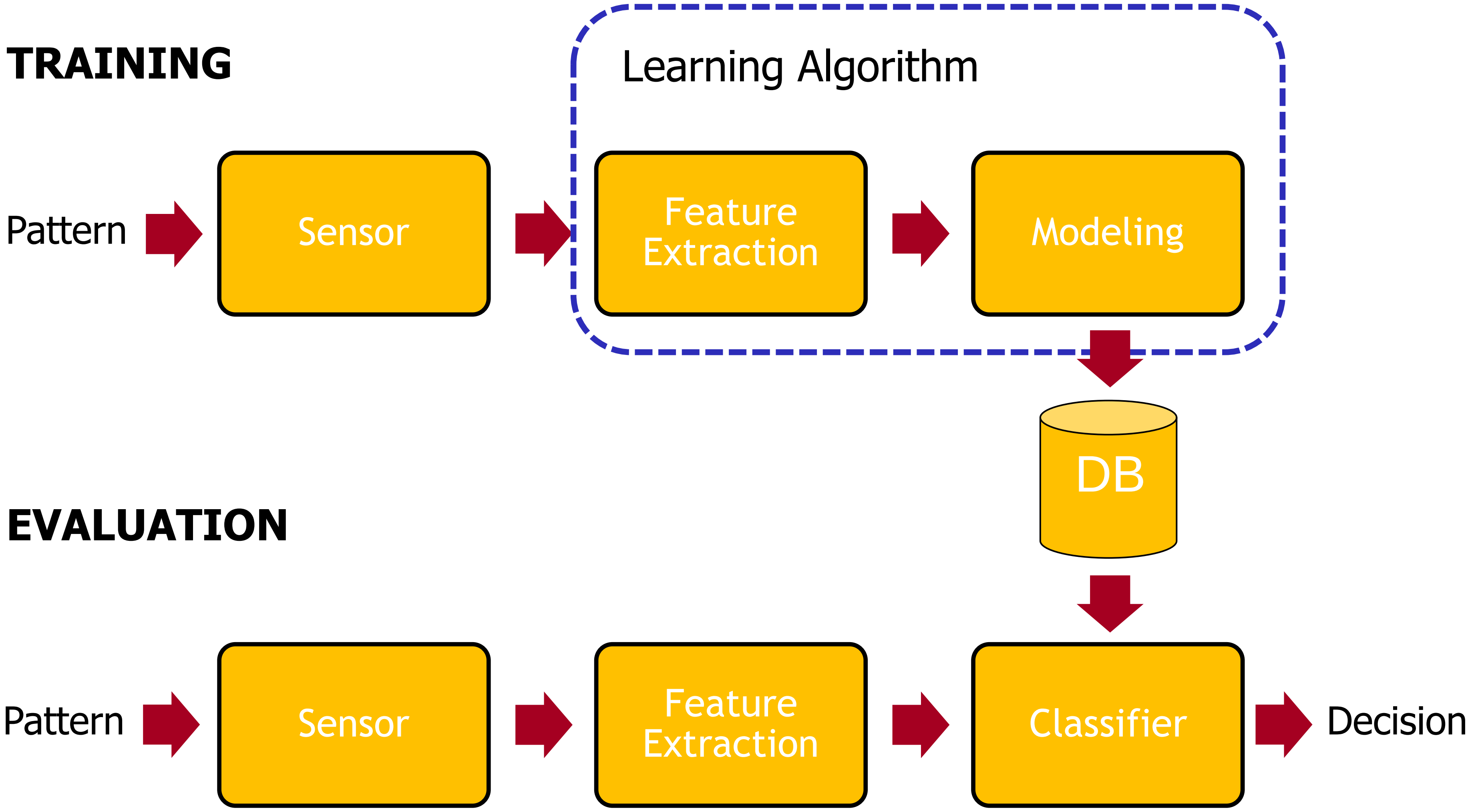
- **The issue of generalization**
  - The recognition rate of our linear classifier (95.7%) met the design specifications, but we still think we can improve the performance of the system
  - We then design a classifier that obtains an impressive classification rate of 99.9975% with the following decision boundary



Pattern Recognition Systems

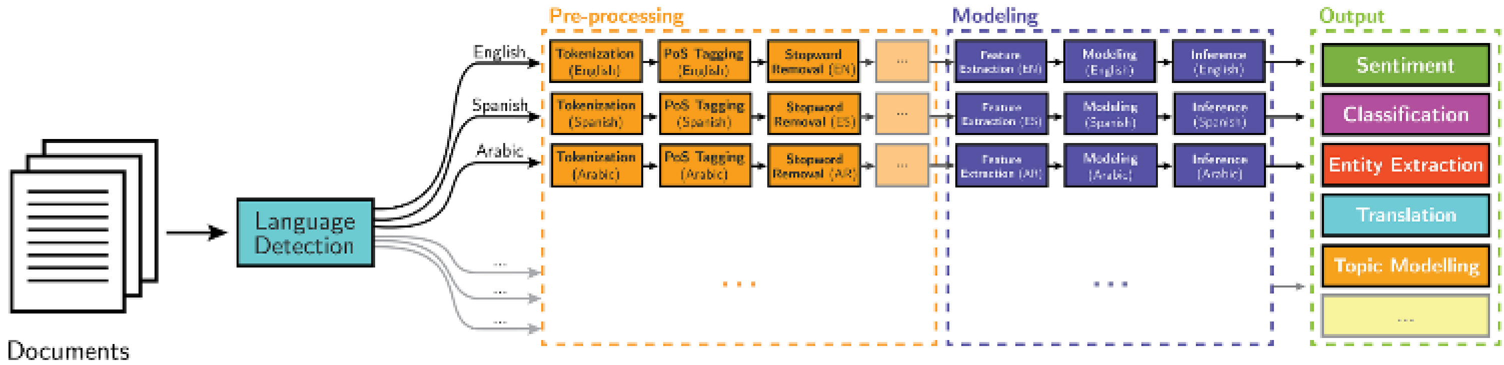


Pattern Recognition Systems



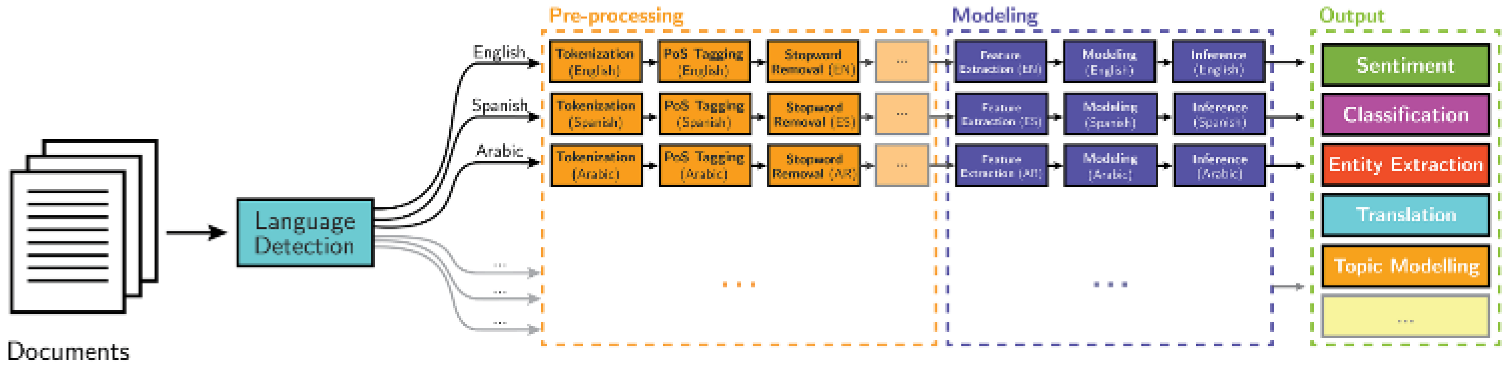
Architectures

Classical NLP

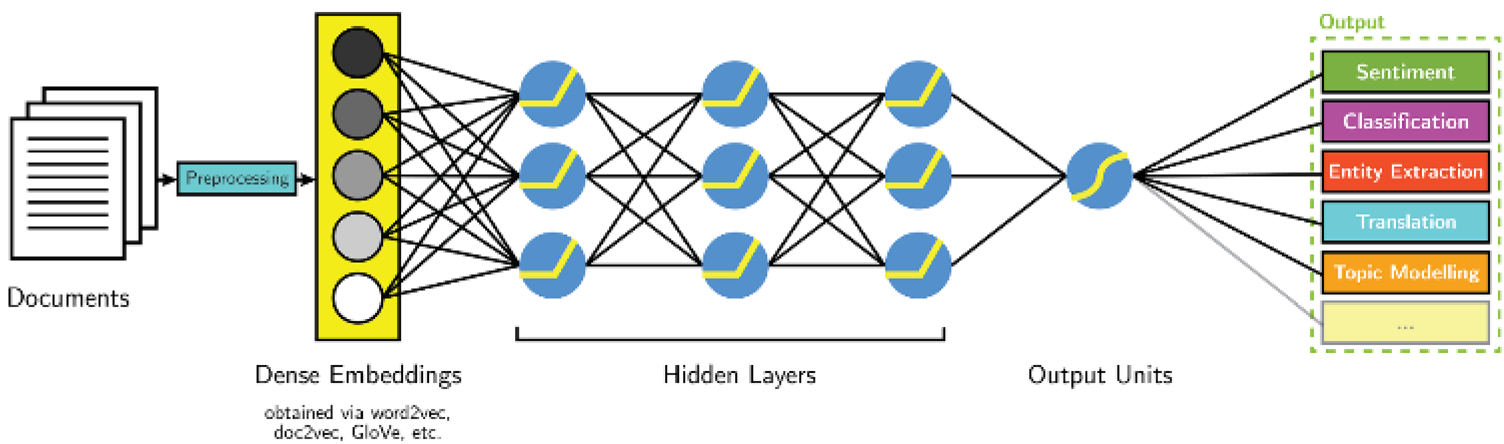


Architectures

Classical NLP



Deep Learning-based NLP



Aprendizaje Automático

Introduction to Machine Learning

Máster en Bioinformática y Biología Computacional