Lecture 1

Machine Learning Basics

Gestió i Distribució de Senyals Audiovisuals (GDSA)

UPC ESEIAAT, Terrassa. Autumn 2020

Xavier Giro-i-Nieto
Associate Professor
Universitat Politècnica de Catalunya

@DocXavi
xavier.giro@upc.edu

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

UPC

ATENeA

Acknowledgements

Kevin McGuinness

kevin.mcguinness@dcu.ie

Research Fellow

Insight Centre for Data Analytics

Dublin City University

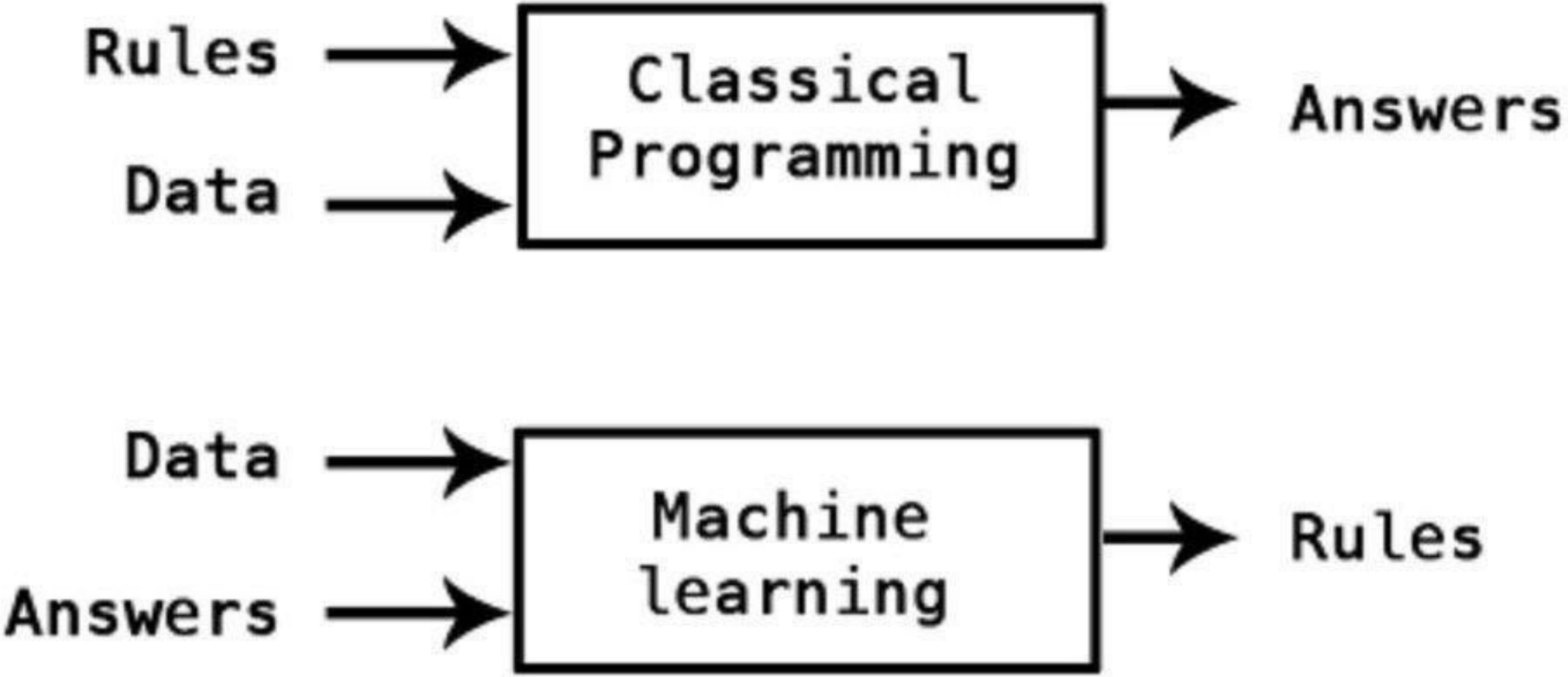
DCU

Insight

Centre for Data Analytics

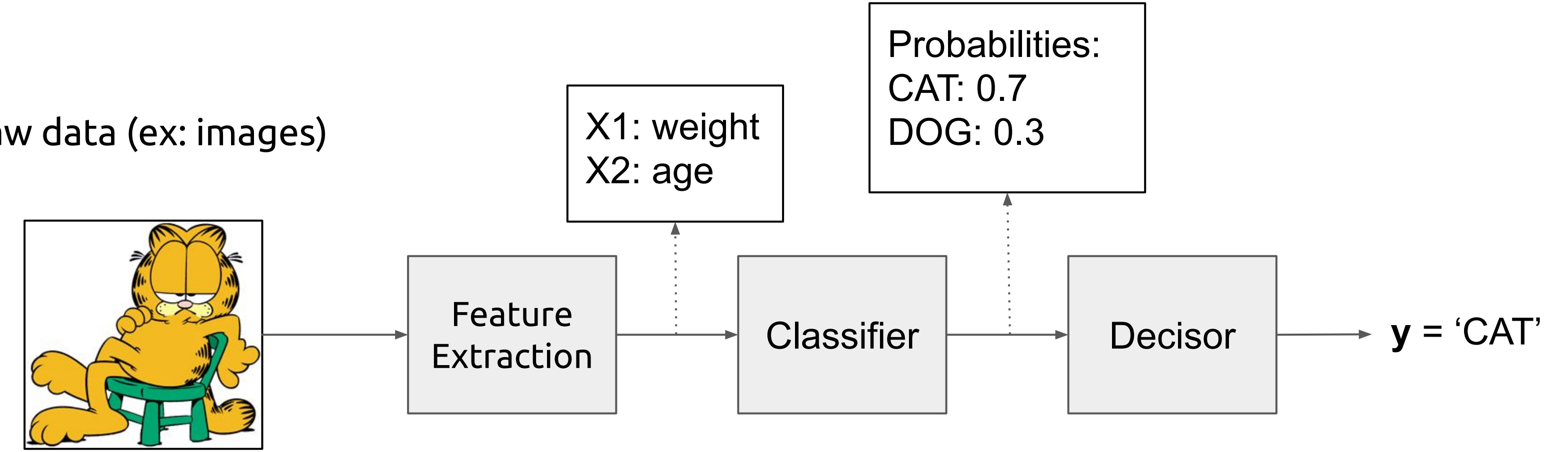
A world leading SFI Research Centre


MACHINE LEARNING
MACHINE LEARNING EVERYWHERE!
memegenerator.net


Rules
Data
Classical Programming
Answers
Data
Answers
Machine learning
Rules

Representation + Learning pipeline

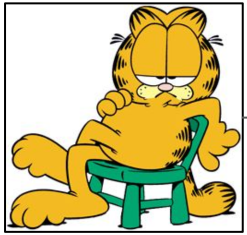
Raw data (ex: images)



Slide concept: Santiago Pascual (UPC 2019). Garfield is a character created by Jim Davis.

Representation + Learning pipeline

Raw data (ex: images)



Feature Extraction

X1: weight
X2: age

Classifier

Probabilities:
CAT: 0.7
DOG: 0.3

Decisor

y = ‘CAT’

Shall we extract features now?

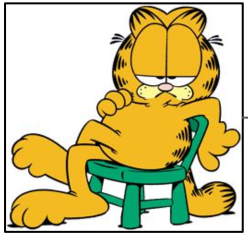
Neural Network

Slide concept: Santiago Pascual (UPC 2019). Garfield is a character created by Jim Davis.

End-to-end Representation Learning

End to End concept

Raw data (ex: images)



We CAN inject the raw data, and features will be learned in the NN !!

Classifier

Neural Network

Probabilities:
CAT: 0.7
DOG: 0.3

Decisor

y = 'CAT'

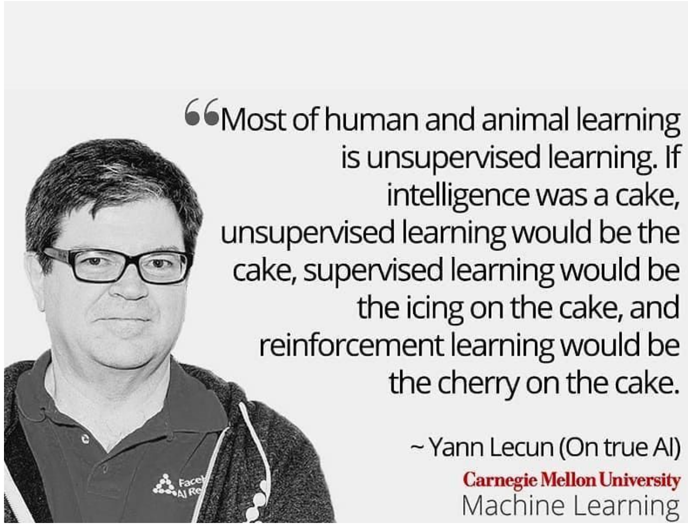
Slide concept: Santiago Pascual (UPC 2019). Garfield is a character created by Jim Davis.

Types of Machine Learning

Yann Lecun's Black Forest cake

- "Pure" Reinforcement Learning (cherry)
  - The machine predicts a scalar reward given once in a while.
  - **A few bits for some samples**
- Supervised Learning (icing)
  - The machine predicts a category or a few numbers for each input
  - Predicting human-supplied data
  - **10→10,000 bits per sample**
- Unsupervised/Predictive Learning (cake)
  - The machine predicts any part of its input for any observed part.
  - Predicts future frames in videos
  - **Millions of bits per sample**

- (Yes, I know, this picture is slightly offensive to RL folks. But I'll make it up)



Machine Learning

	...with a teacher	...without a teacher
Active agent...	Reinforcement learning (with extrinsic reward)	Intrinsic motivation / Exploration.
Passive agent...	Supervised learning	Unsupervised learning

Slide inspired by Alex Graves (Deepmind) at
"Unsupervised Learning Tutorial" @ NeurIPS 2018.

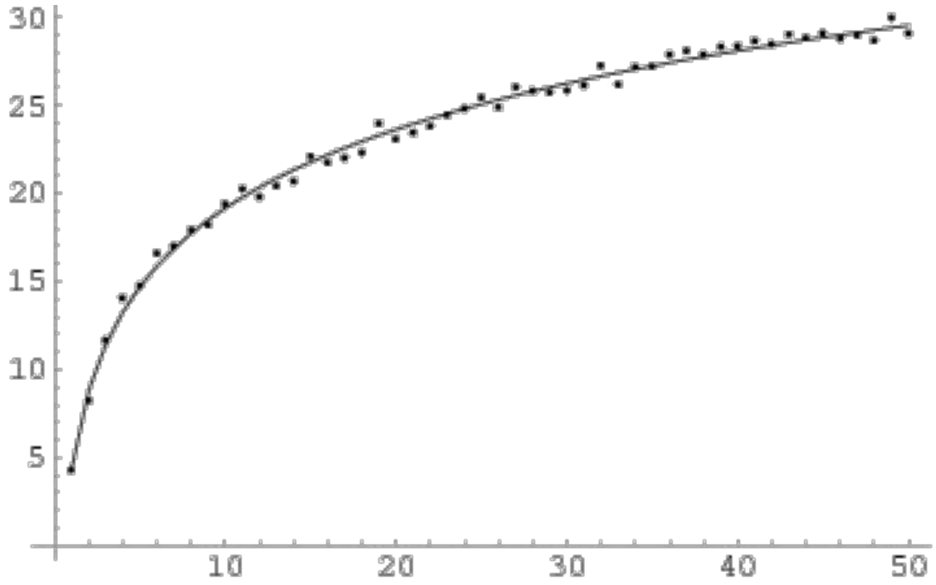
Machine Learning

	...with a teacher	...without a teacher
Active agent...	Reinforcement learning (with extrinsic reward)	Intrinsic motivation / Exploration.
Passive agent...	Supervised learning	Unsupervised learning

Slide inspired by Alex Graves (Deepmind) at [“Unsupervised Learning Tutorial”] @ NeurIPS 2018.

Supervised learning

Fit a function: $\mathbf{y} = f(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^m$



Supervised learning

Fit a function: $\mathbf{y} = f(\mathbf{x}),\quad \mathbf{x} \in \mathbb{R}^m$

Given paired training examples $\{(\mathbf{x}_i, \mathbf{y}_i)\}$

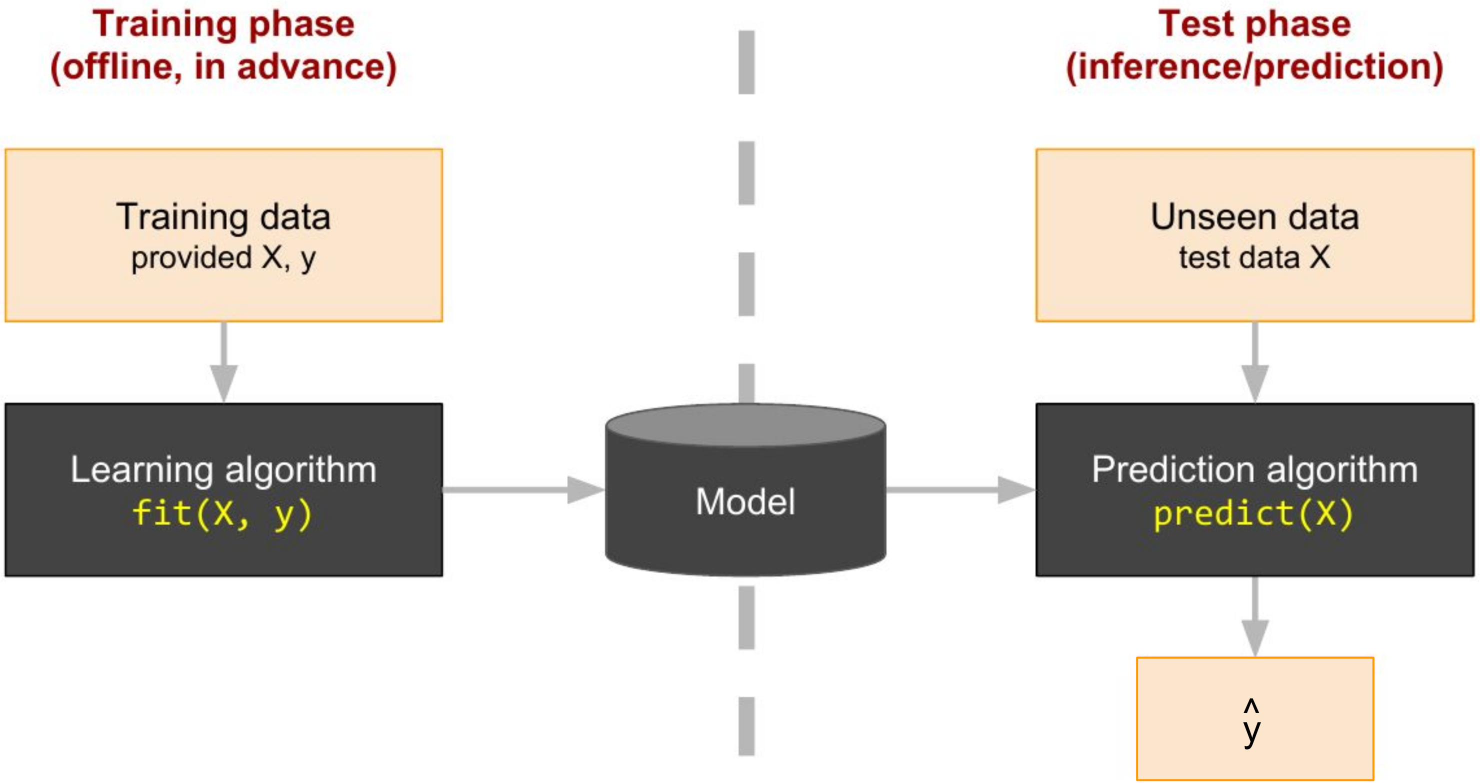
Supervised learning

Fit a function: $\mathbf{y} = f(\mathbf{x}), \quad \mathbf{x} \in \mathbb{R}^m$

Given paired training examples $\{(\mathbf{x}_i, \mathbf{y}_i)\}$

Key point: **generalize well to unseen examples**

Black box abstraction of supervised learning



Regression vs Classification

Depending on the type of target $\mathbf{y}$ we get:

- **Regression:** $\mathbf{y} \in \mathbb{R}^N$ is continuous (e.g. temperatures $\mathbf{y}$ = {19º, 23.2º, 22.8º})

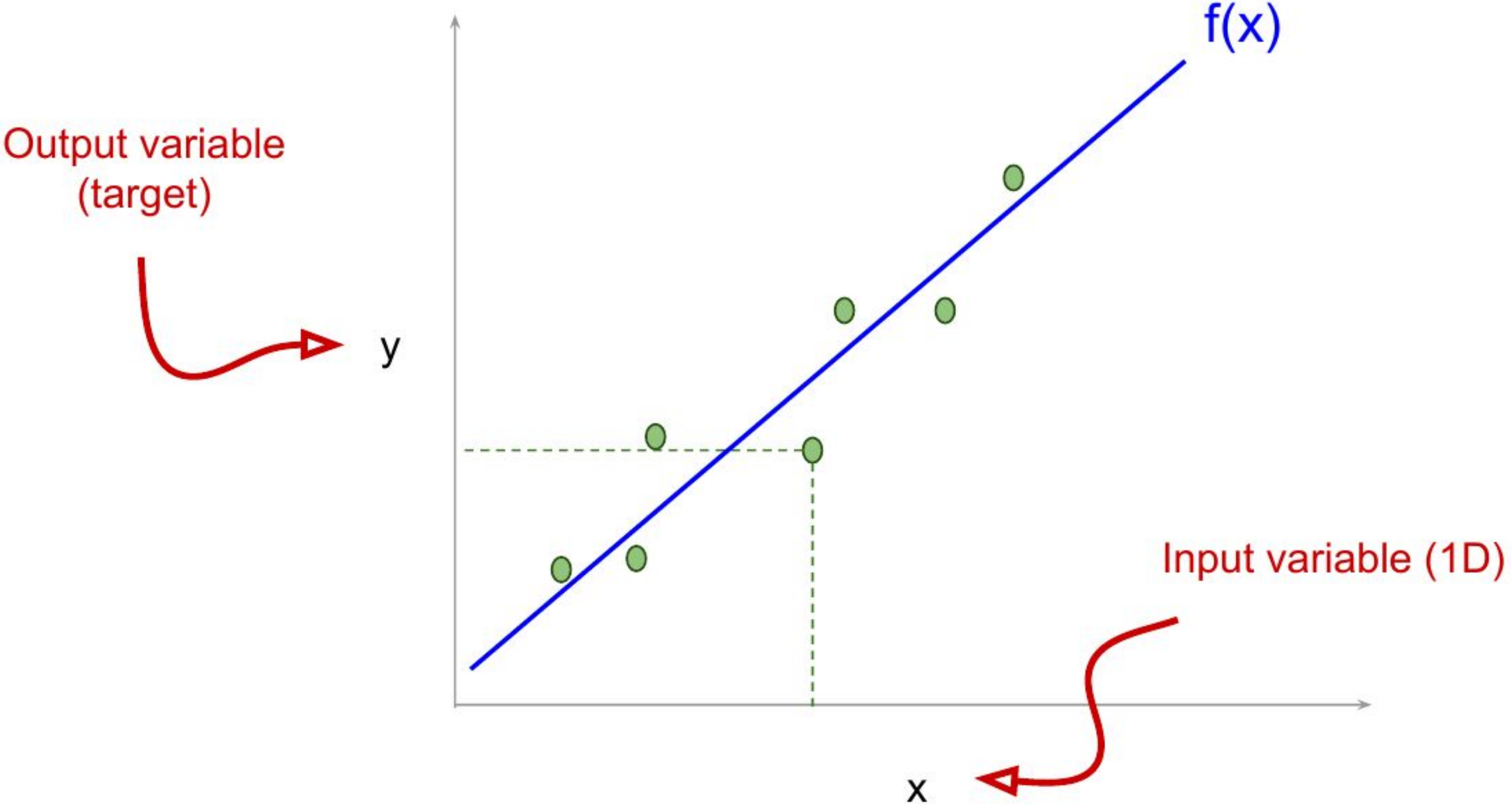
- **Classification:** $\mathbf{y}$ is discrete (e.g. $\mathbf{y}$ = {“dog”,”cat”,”ostrich”}).

Regression vs Classification

Depending on the type of target $\mathbf{y}$ we get:

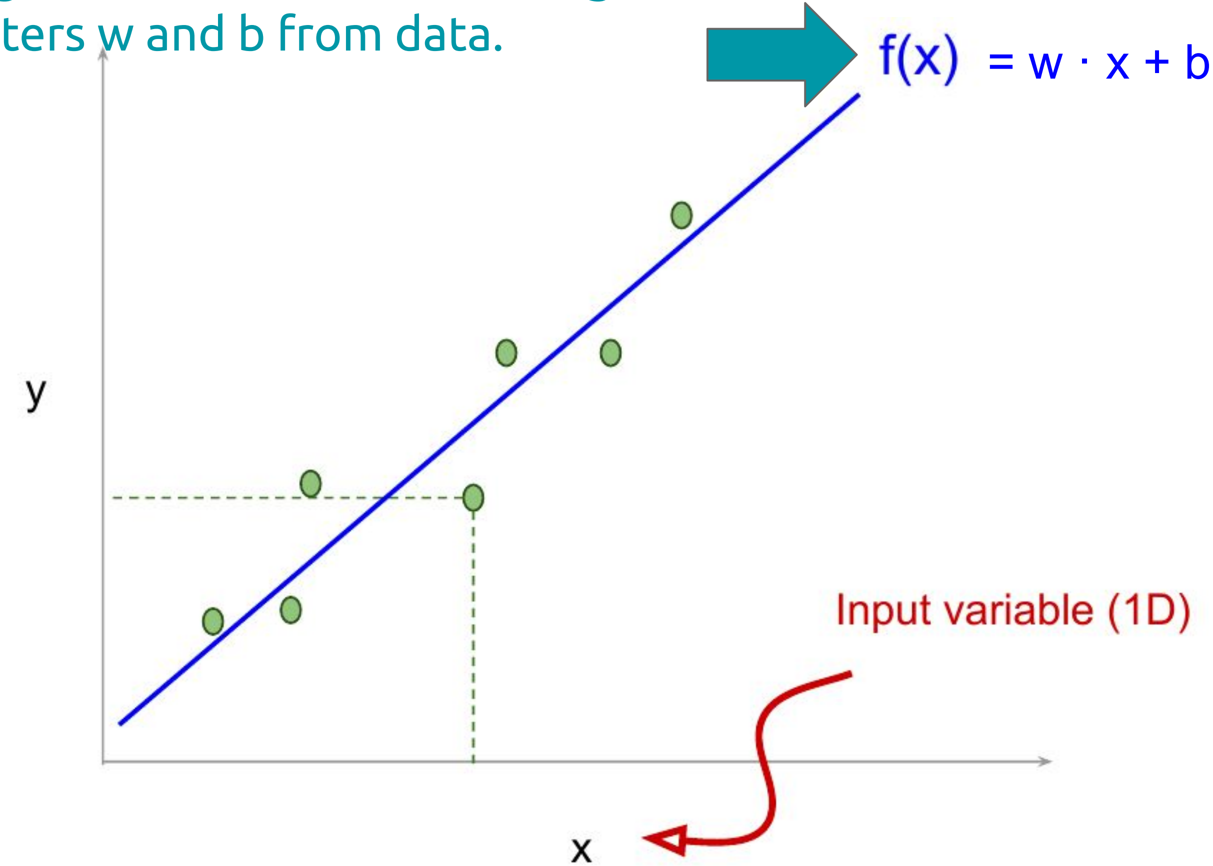
- **<u>Regression</u>:** $\mathbf{y} \in \mathbb{R}^N$ is continuous (e.g. temperatures $\mathbf{y}$ = {19º, 23.2º, 22.8º})
- **<u>Classification</u>:** $\mathbf{y}$ is discrete (e.g. $\mathbf{y}$ = {“dog”, “cat”, “ostrich”}).

Linear Regression (eg. 1D input - 1D ouput)



Linear Regression (eg. 1D input - 1D ouput)

Training this model means **learning** parameters w and b from data.



Linear Regression (eg. M-D input - 1D output)

Input data can also be M-dimensional with vector **x**:

$$\boxed{y = \mathbf{w}^T \cdot \mathbf{x} + b} = \ w1 \cdot x1 + w2 \cdot x2 + w3 \cdot x3 + \ldots + wM \cdot xM + b$$

e.g. we want to predict the **price of a house (y)** based on:

x1 = square-meters (sqm)

x2,3 = location (lat, lon)

y = price = $w1 \cdot (\text{sqm}) + w2 \cdot (\text{lat}) + w3 \cdot (\text{lon}) + b$

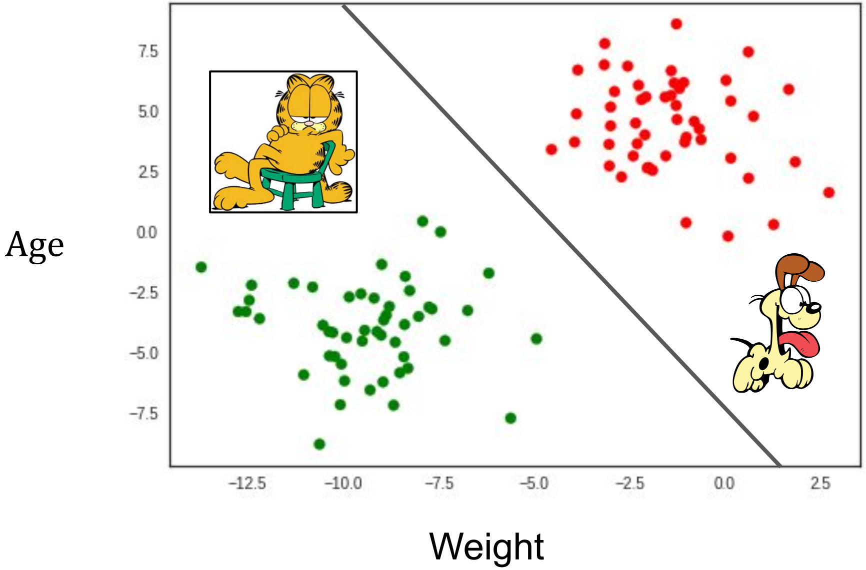
Regression vs Classification

Depending on the type of target $\mathbf{y}$ we get:

- **Regression:** $\mathbf{y} \in \mathbb{R}^N$ is continuous (e.g. temperatures $\mathbf{y}$ = {19º, 23.2º, 22.8º})
- **Classification:** $\mathbf{y}$ is discrete (e.g. $\mathbf{y}$ = {”dog”,”cat”,”ostrich”}).

Binary Classification (eg. 2D input, 1D ouput)

2D feature space



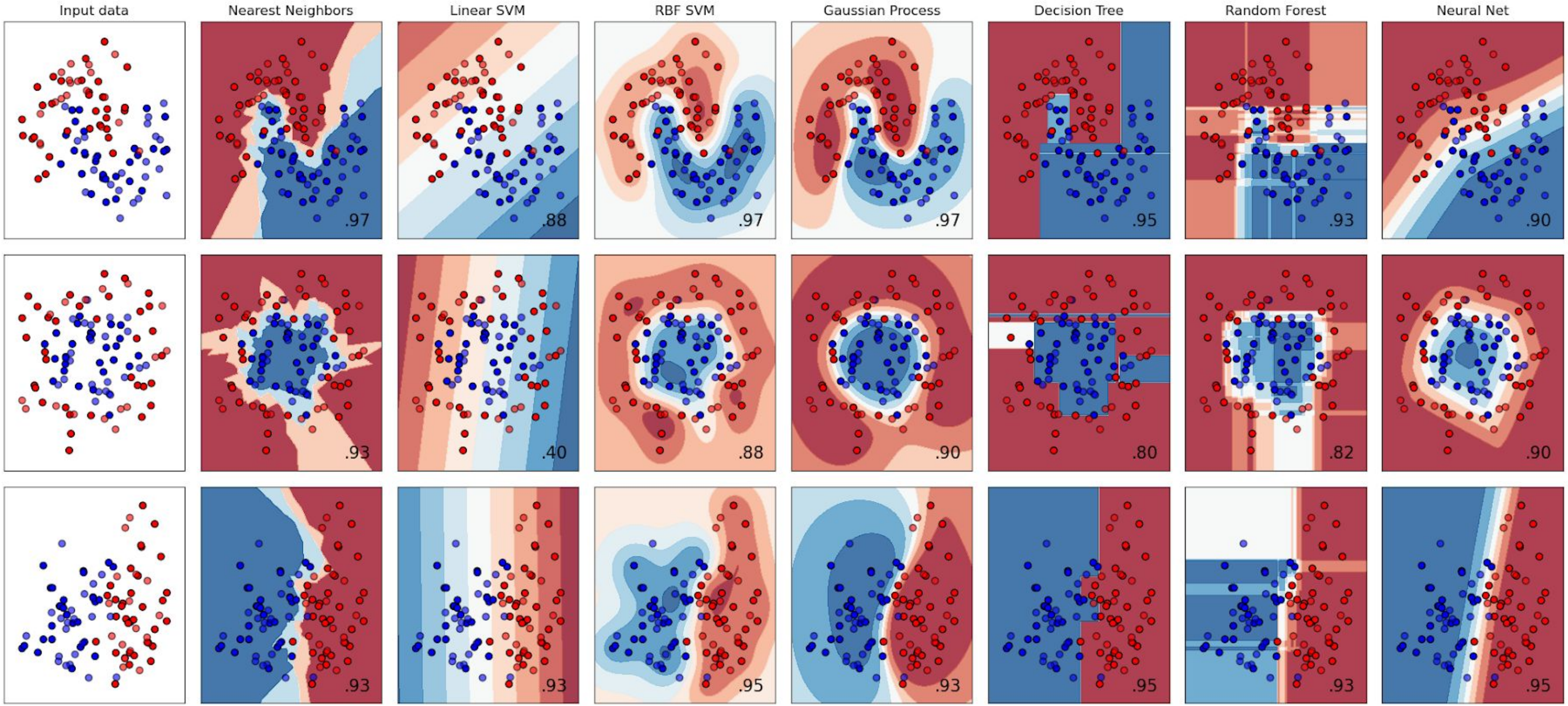
$f(x) = \begin{cases} \\ \end{cases}$

 $\text{if } w \cdot x + b > 0$

 otherwise

Slide inspiration: Santiago Pascual (UPC TelecomBCN 2019). Garfield and Odie are characters created by Jim Davis.

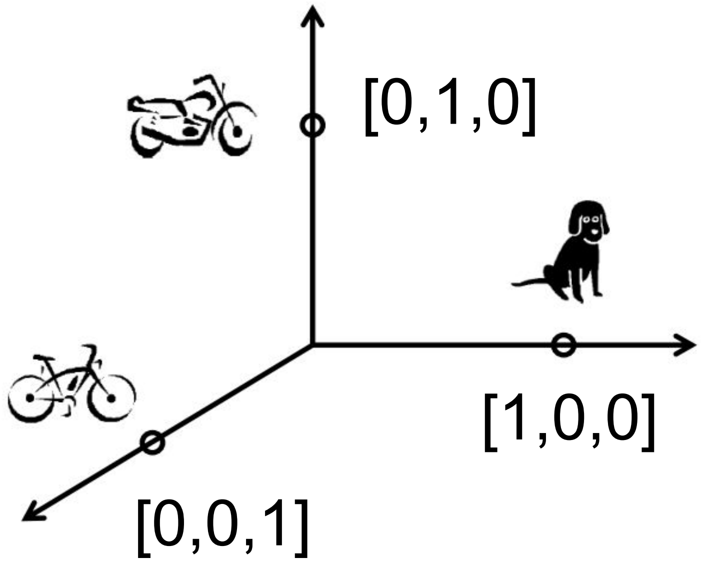
Binary Classification (eg. 2D input, 1D ouput)



Source: Classifier comparison in scikit-learn [related tweet]

Multi-class Classification (N-D output)

- **<u>Classification</u>:** $\mathbf{y}$ is discrete (e.g. $\mathbf{y}$ = {"dog","bike","motorbike"}.
  - Classes are often coded as **one-hot vector** (each class corresponds to a different dimension of the output space)

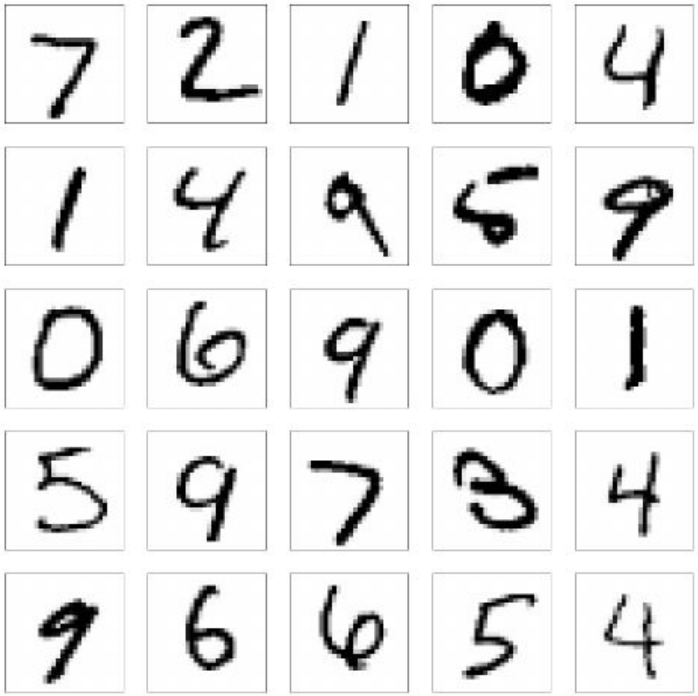


One-hot representations

Perronin, F., CVPR Tutorial on LSVR @ CVPR'14, Output embedding for LSVR

Multi-class Classification (N-D output)

Handwritten Digits from the **MNIST** dataset



Produce a classifier to map from pixels to the digit.

- If images are grayscale and 28×28 pixels in size, then $\mathbf{x}_i \in \mathbb{R}^{784}$
- $y_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

Example of a **multi-class classification** task.

Multi-class Classification (N-D output)

```
cat: xT = [1,0,0, ..., 0]
dog: xT = [0,1,0, ..., 0]
.
.
house: xT = [0,0,0, …,0,1,0,...,0]
.
.
.
```

Number of words, |V| ?

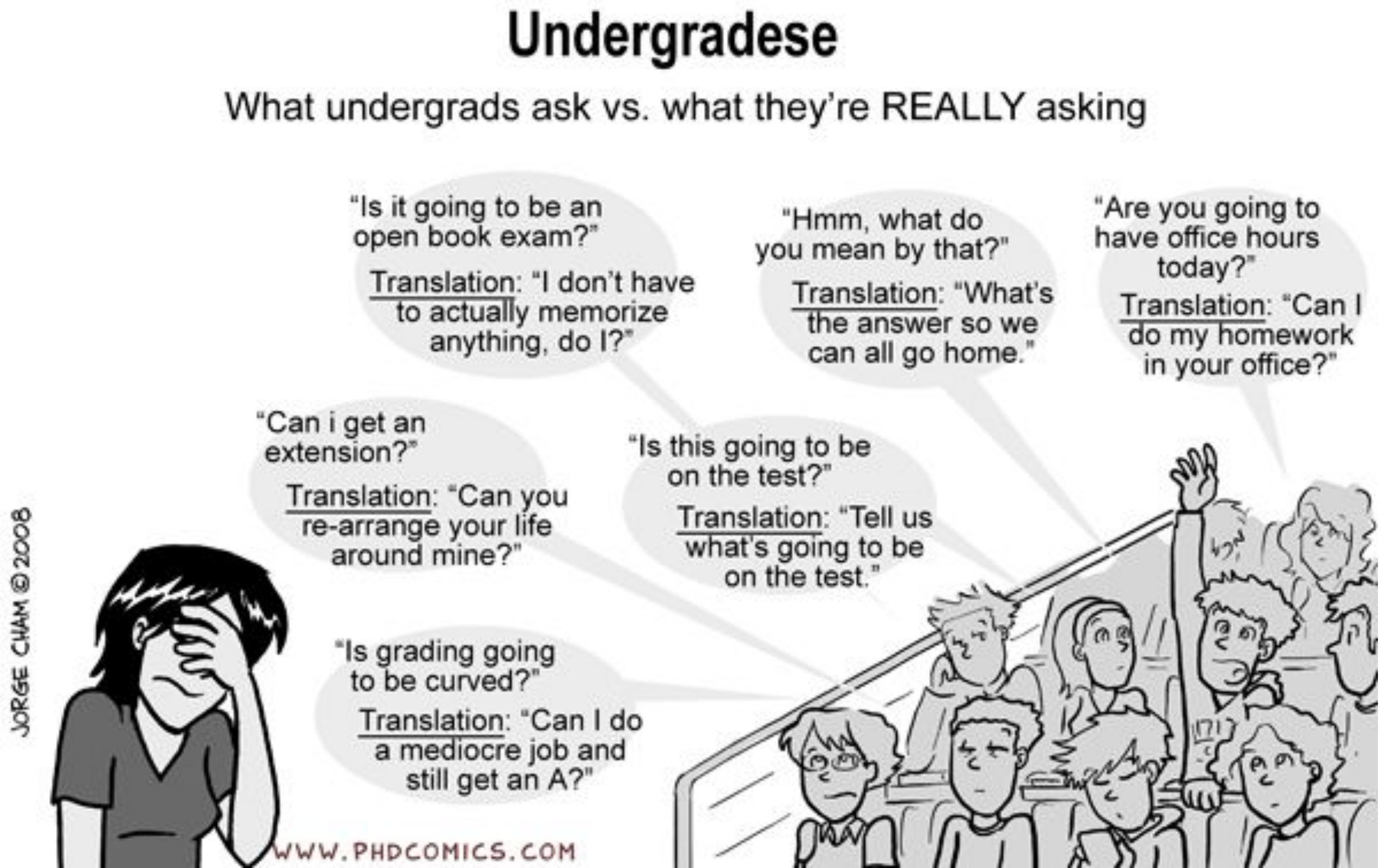
B2: 5K
C2: 18K
LVSR: 50-100K
Wikipedia (1.6B): 400K
Crawl data (42B): 2M

One-hot representations

Discussion

Can intelligence be modelled by curve fitting ?
(click on the tweet to read the thread discussion for arguments)

Questions ?


Undergradese
What undergrads ask vs. what they're REALLY asking
"Is it going to be an open book exam?"
Translation: "I don't have to actually memorize anything, do I?"
"Hmm, what do you mean by that?"
Translation: "What's the answer so we can all go home."
"Are you going to have office hours today?"
Translation: "Can I do my homework in your office?"
"Can i get an extension?"
Translation: "Can you re-arrange your life around mine?"
"Is this going to be on the test?"
Translation: "Tell us what's going to be on the test."
"Is grading going to be curved?"
Translation: "Can I do a mediocre job and still get an A?"
JORGE CHAM © 2008
WWW.PHDCOMICS.COM