http://bit.ly/idl2020

Day 1 Lecture 2

Machine Learning Basics

Xavier Giro-i-Nieto

[@DocXavi](https://twitter.com/DocXavi)
xavier.giro@upc.edu

Associate Professor
Universitat Politècnica de Catalunya

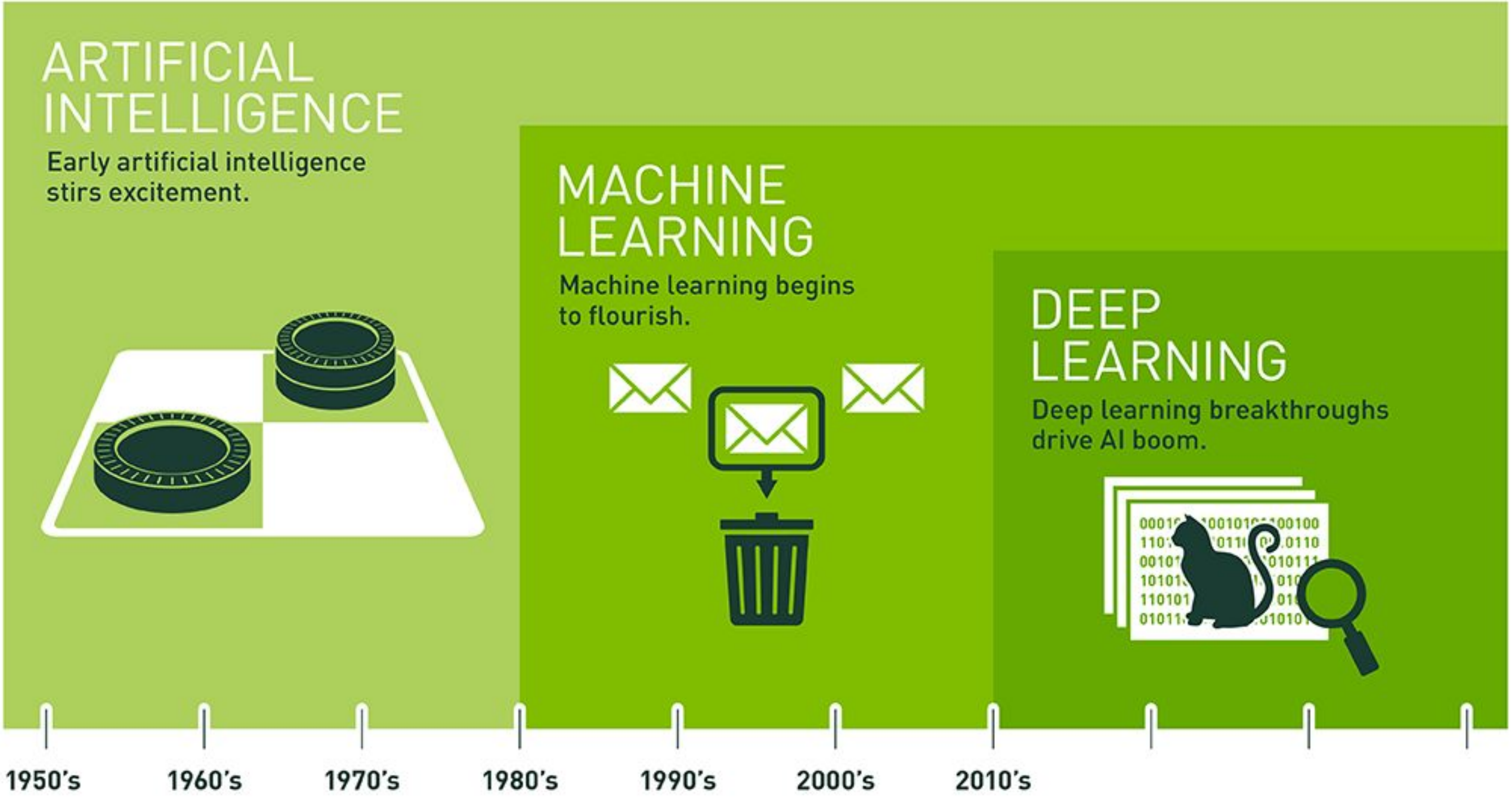
Acknowledgements

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

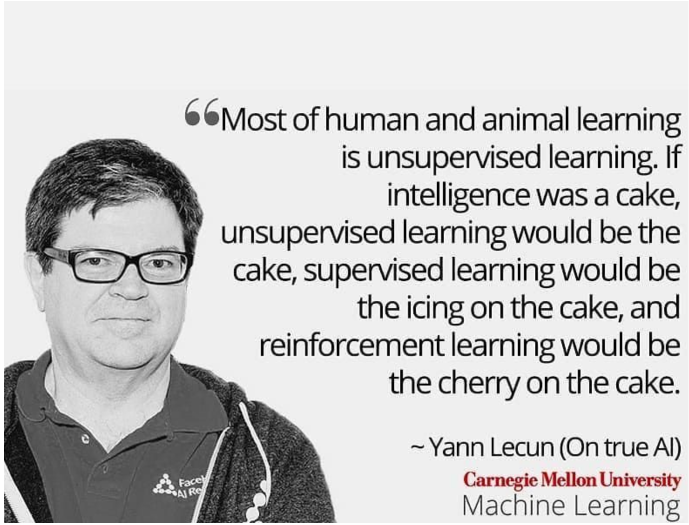


Source: [NVIDIA]

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.

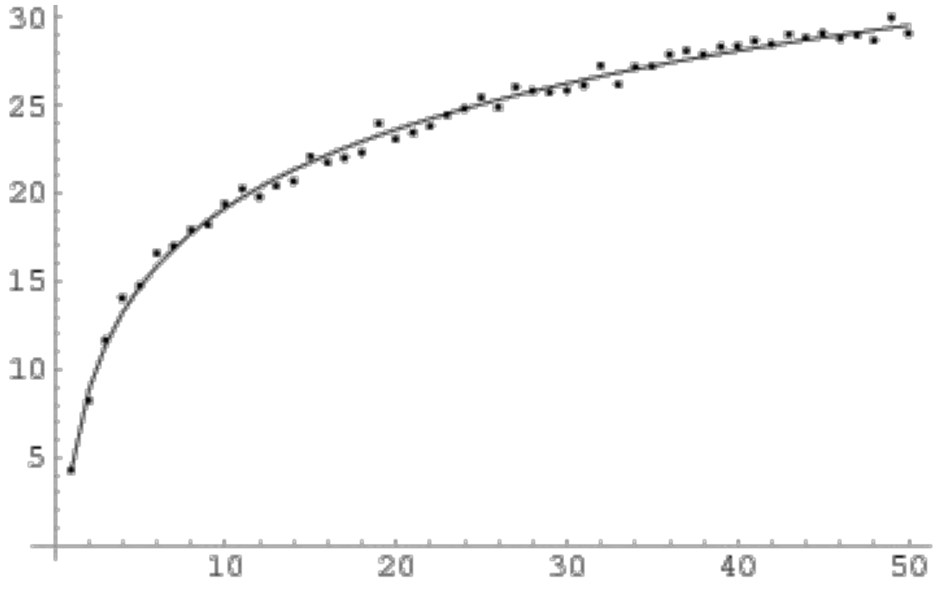
Machine Learning

	...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 $\{(\mathbf{x}_i, \mathbf{y}_i)\}$

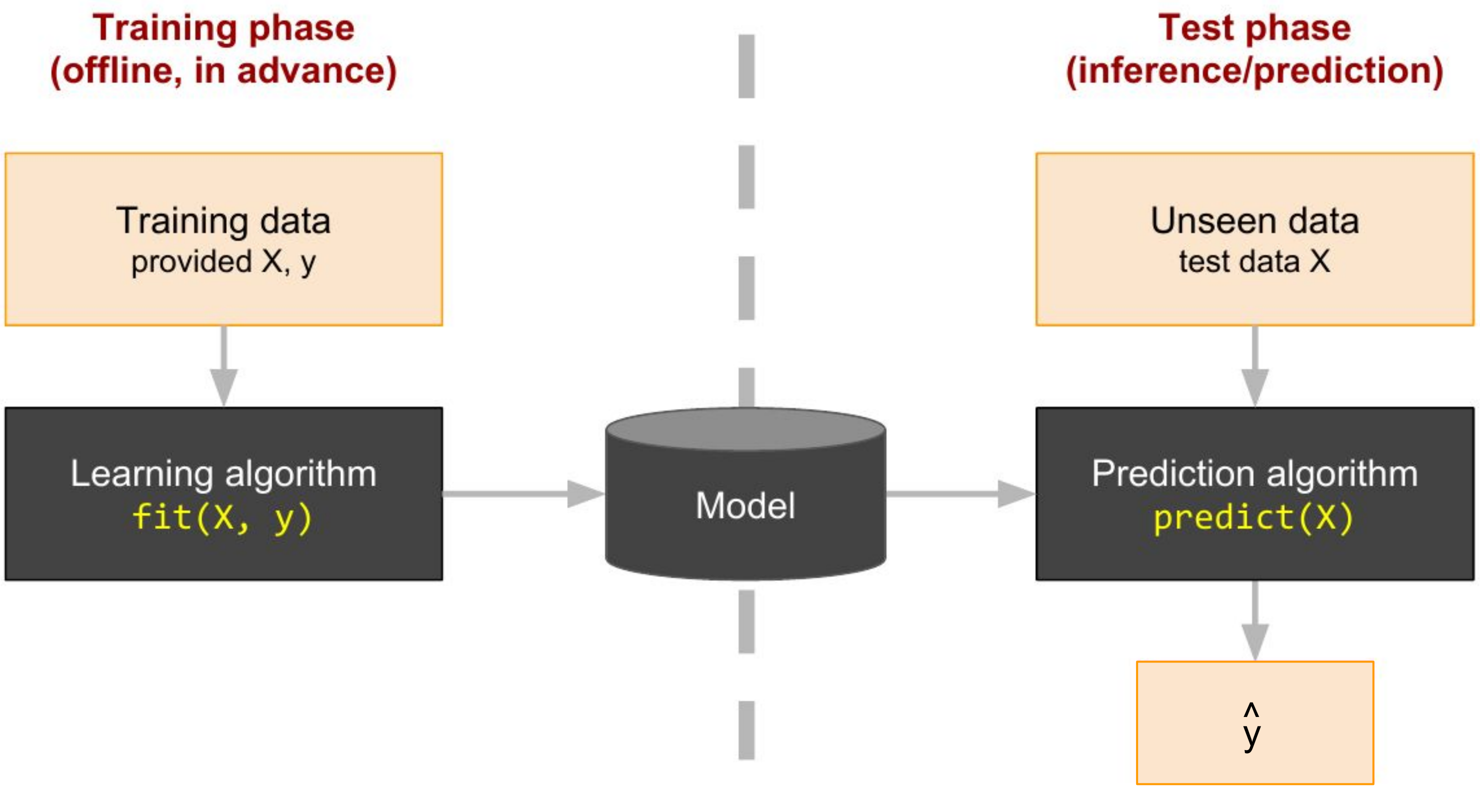
Supervised learning

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, 23º, 22º})

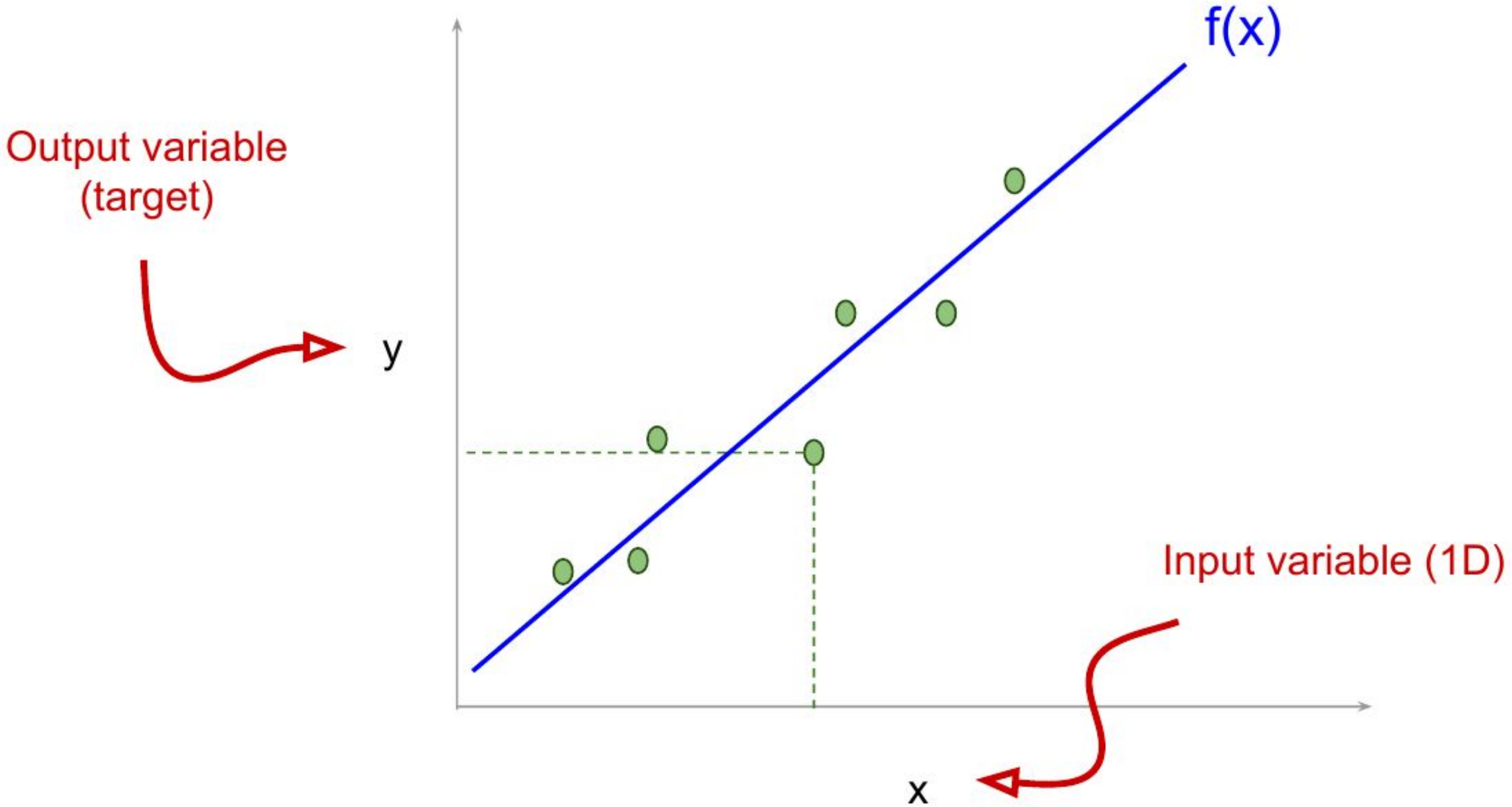
- **Classification:** $\mathbf{y}$ is discrete (e.g. $\mathbf{y}$ = {“dog”,”cat”,”ostrich”}).

Regression vs Classification

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

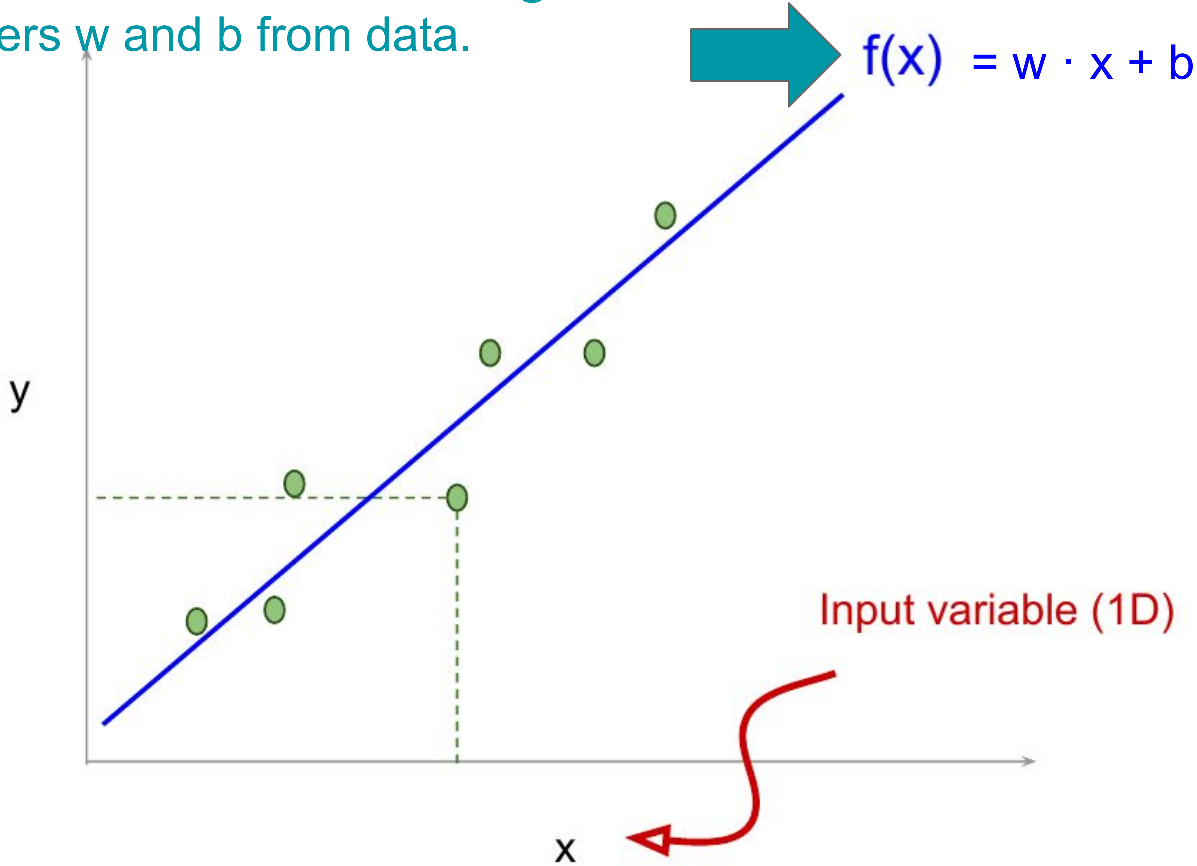
- **Regression:** $\mathbf{y} \in \mathbb{R}^N$ is continuous (e.g. temperatures $\mathbf{y}$ = {19º, 23.2º, 22.8º})
- **Classification:** $\mathbf{y}$ is discrete (e.g. $\mathbf{y}$ = {“dog”, “cat”, “ostrich”}).

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



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

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



Linear Regression (M-D input)

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)

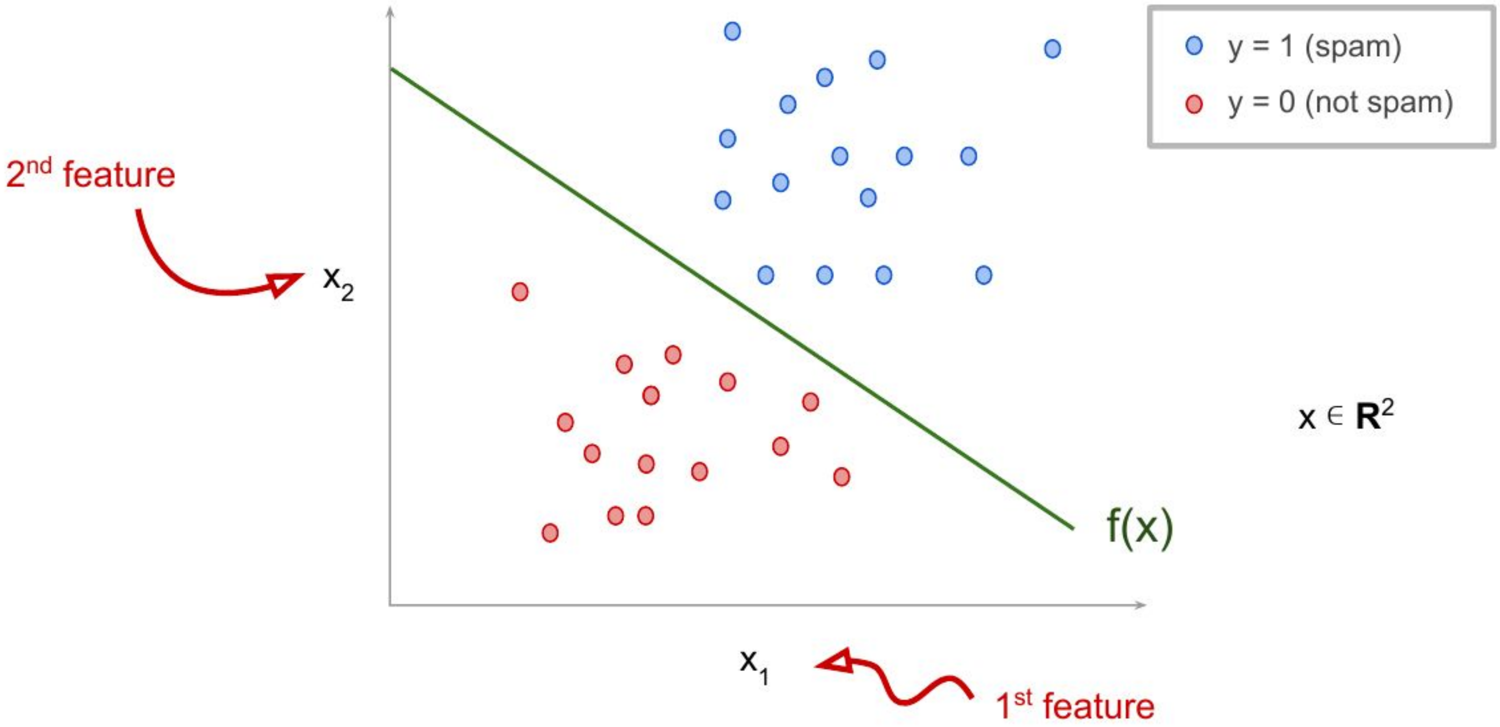
y = price = $w1 \cdot$(sqm) + $w2 \cdot$(lat) + $w3 \cdot$(lon)+ b

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, 23º, 22º})
- **<u>Classification</u>:** $\mathbf{y}$ is discrete (e.g. $\mathbf{y}$ = {"dog","cat","ostrich"}).

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



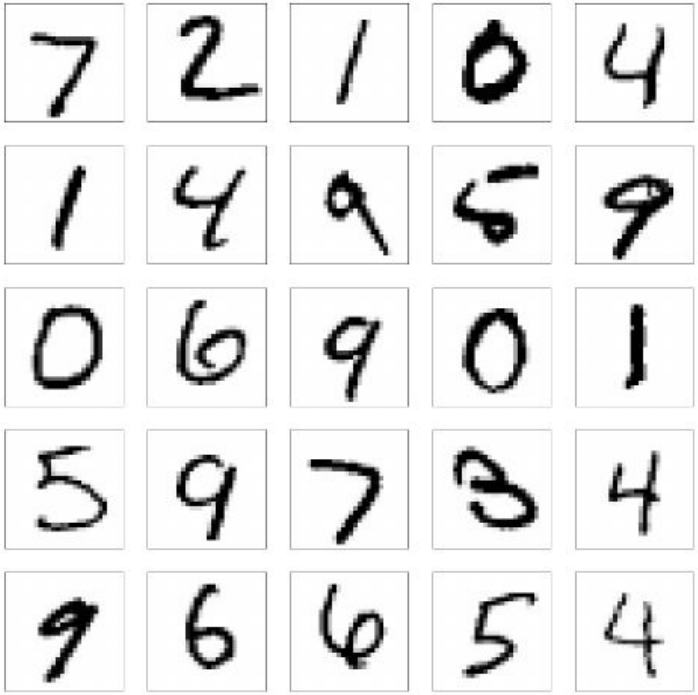
$x \in \mathbf{R}^2$

Multi-class Classification

Produce a classifier to map from pixels to the digit.

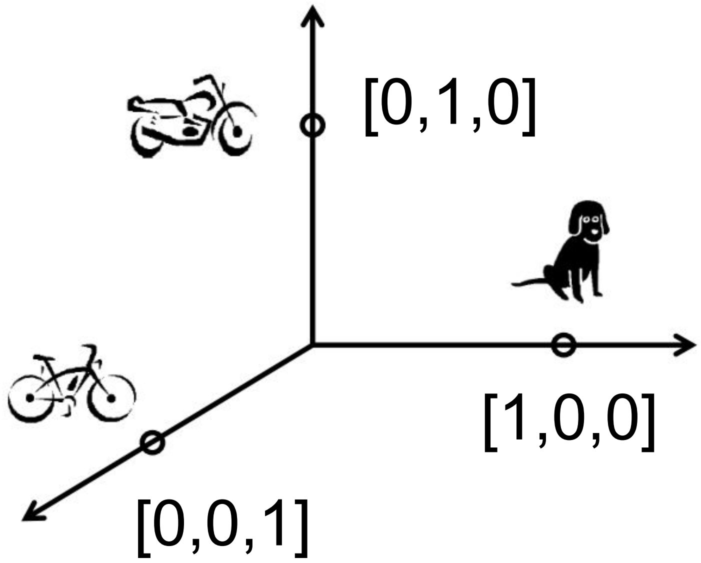
- If images are grayscale and 28×28 pixels in size, then $\mathbf{x}_i \in \mathbb{R}^{784}$
- $y_i \in \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$

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



Multi-class Classification

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

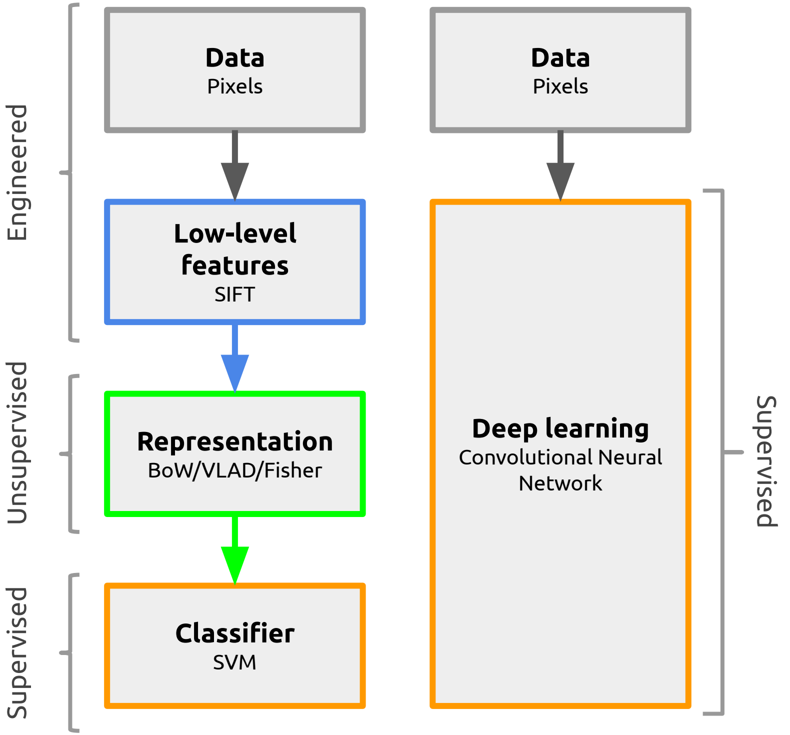


One-hot representations

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

End-to-end Learning

- Old style machine learning:
  - Engineer features (by some unspecified method)
  - Create a representation (descriptor)
  - Train shallow classifier on representation
- Example:
  - SIFT features (engineered)
  - BoW representation (engineered + unsupervised learning)
  - SVM classifier (convex optimization)
- **Deep learning**
  - Learn layers of features, representation, and classifier in one go based on the data alone
  - Primary methodology: deep neural networks (non-convex)



Multi-class Classification

What is the dimensionality of a one-hot representation of the MNIST classes ?

A. 1
B. 28
C. 10
D. 784

Multi-class Classification

What is the dimensionality of a one-hot representation of the MNIST classes ?

A. 1
B. 28
C. 10
D. 784

Regression vs Classification

Should you treat these three problems as classification or as regression problems?

Problem	**Regression ?**	**Classification ?**
Predicting whether stock price of a company will increase tomorrow		
Predict the number of copies a music album will be sold next month		
Predicting the gender of a person by his/her handwriting style		

Regression vs Classification

Should you treat these three problems as classification or as regression problems?

Problem	Regression ?	Classification ?
Predicting whether stock price of a company will increase tomorrow		✓
Predict the number of copies a music album will be sold next month	✓	
Predicting the gender of a person by his/her handwriting style		✓

Discussion

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

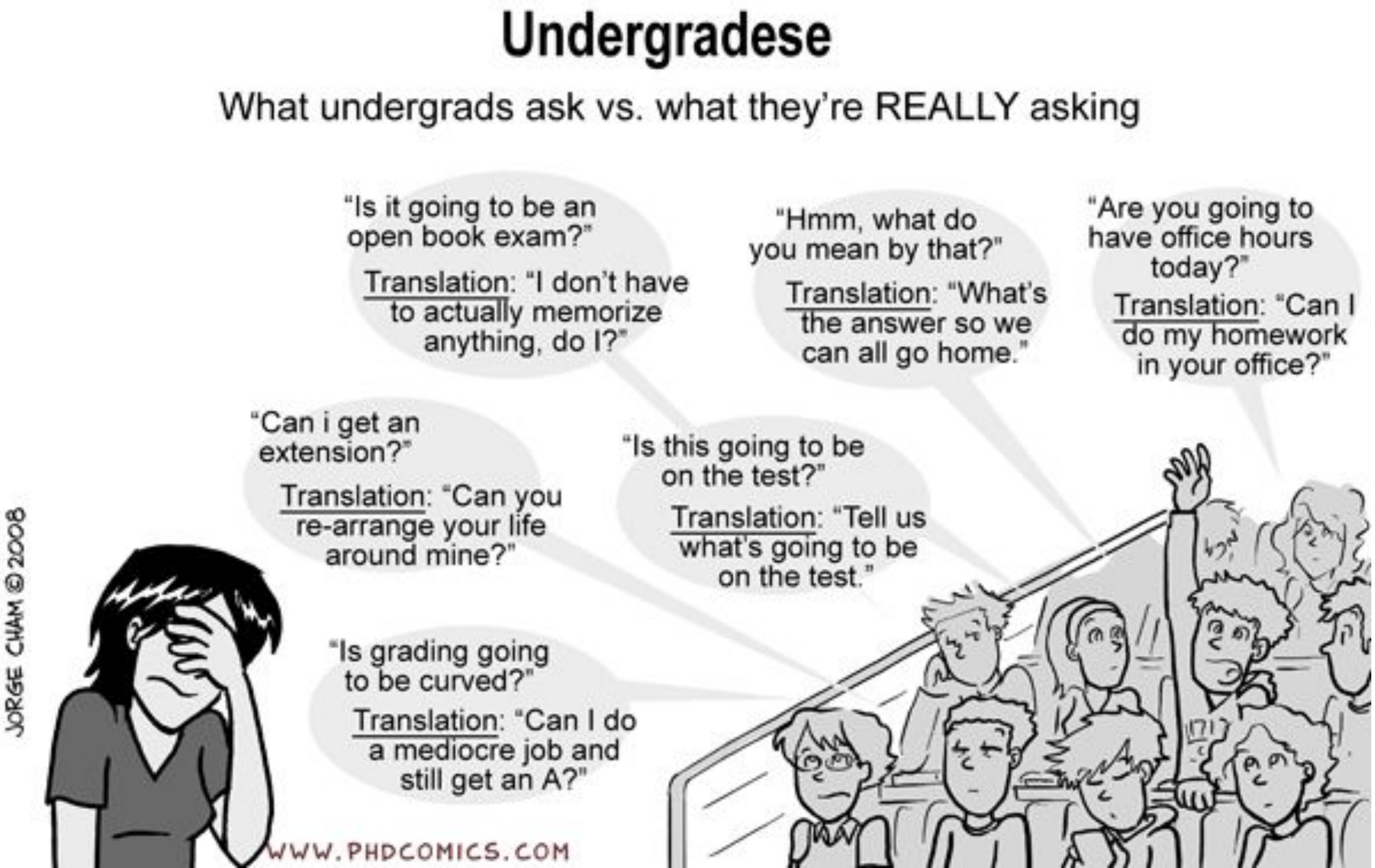
michael_nielsen
@michael_nielsen

Whenever I see this kind of headline, I always think "But what if intelligence is mostly about curve-fitting, and we're merely too un-self-aware to notice?"

AI today and tomorrow is mostly about curve fitting, not intelligence

9:16 PM · Nov 23, 2019 · Twitter Web App

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