Introduction to Machine Learning

Pascal Fua
IC-CVLab

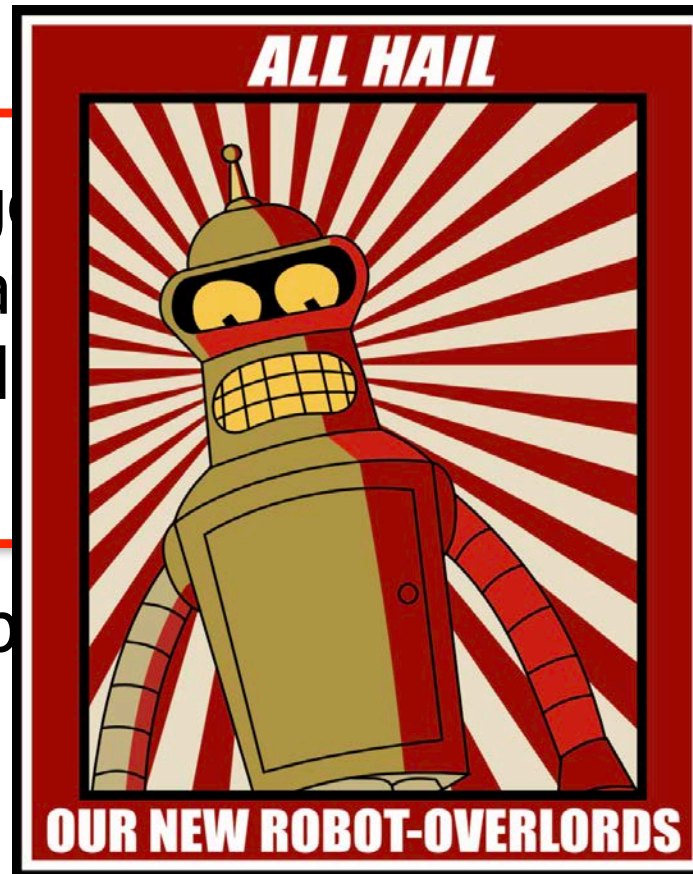# Human vs Machine Learning

Learn from experience

Learn from experience

What is Machine Learning?

- Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding them data and information in the form of observations and real-world interactions.

- Machine learning alg[...]vide knowledge to computers through da[...]interaction with the world. It is then used [...]edictions given new observations.

- Machine learning is ap[...]

Artificial Intelligence/Machine Learning/Deep Learning

Artificial Intelligence

Expert Systems   *A**

                    min-max

Machine Learning

Support Vector Machines

Boosting

            Random Forests

Deep Learning

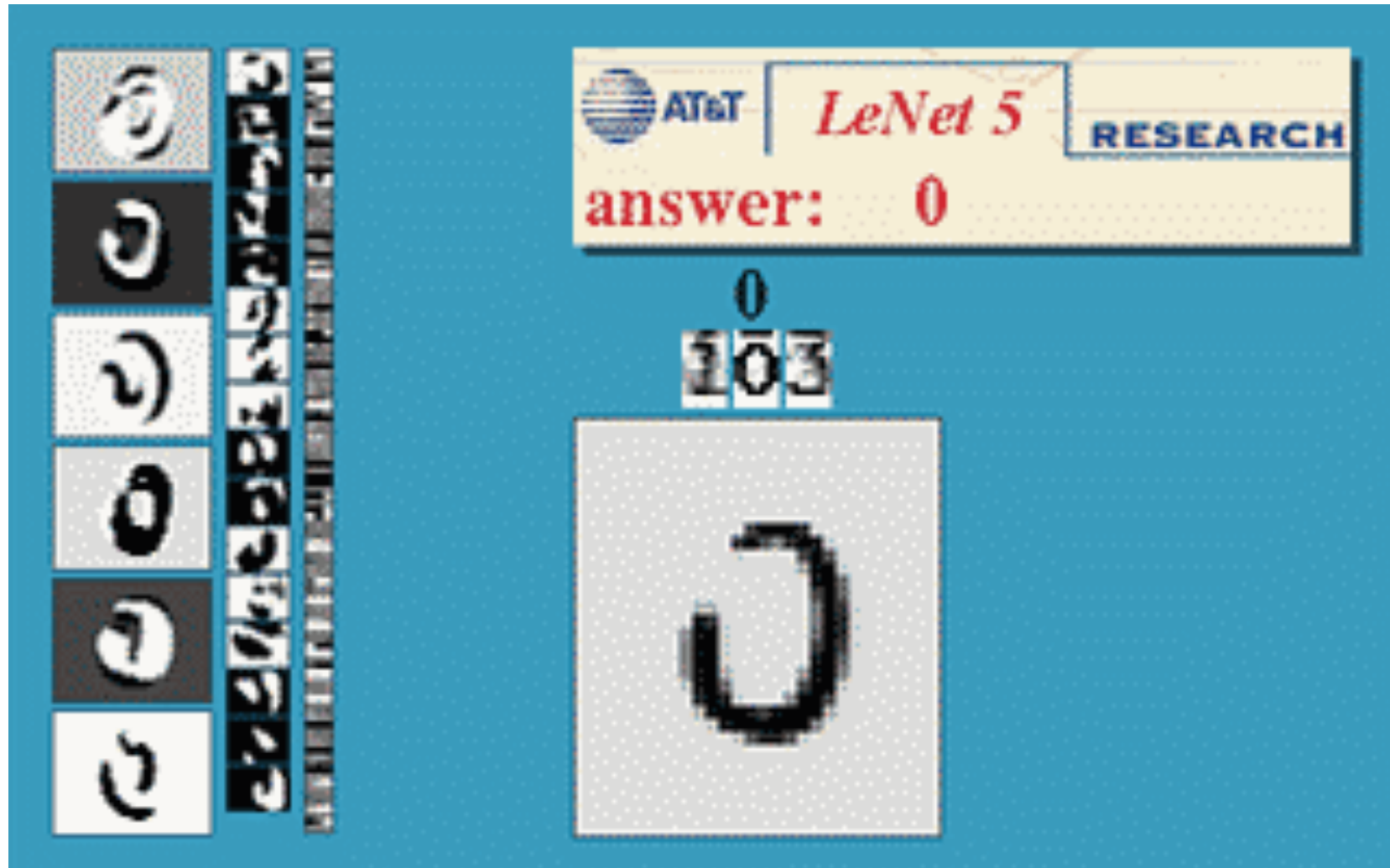*Lenet*   *VGG*

        ResNet

Machine Learning:
Algorithms whose performance can be improved using training data.
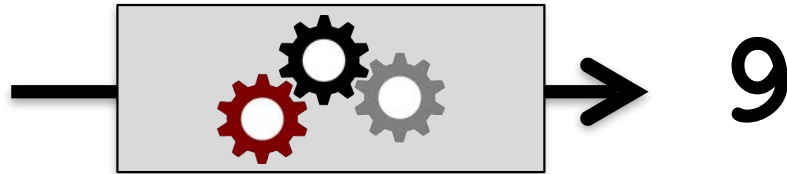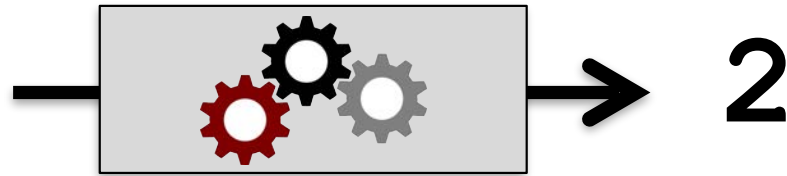
Deep Learning:
Machine Learning Algorithms that rely on neural nets and well suited to cases where there is a lot of training data.
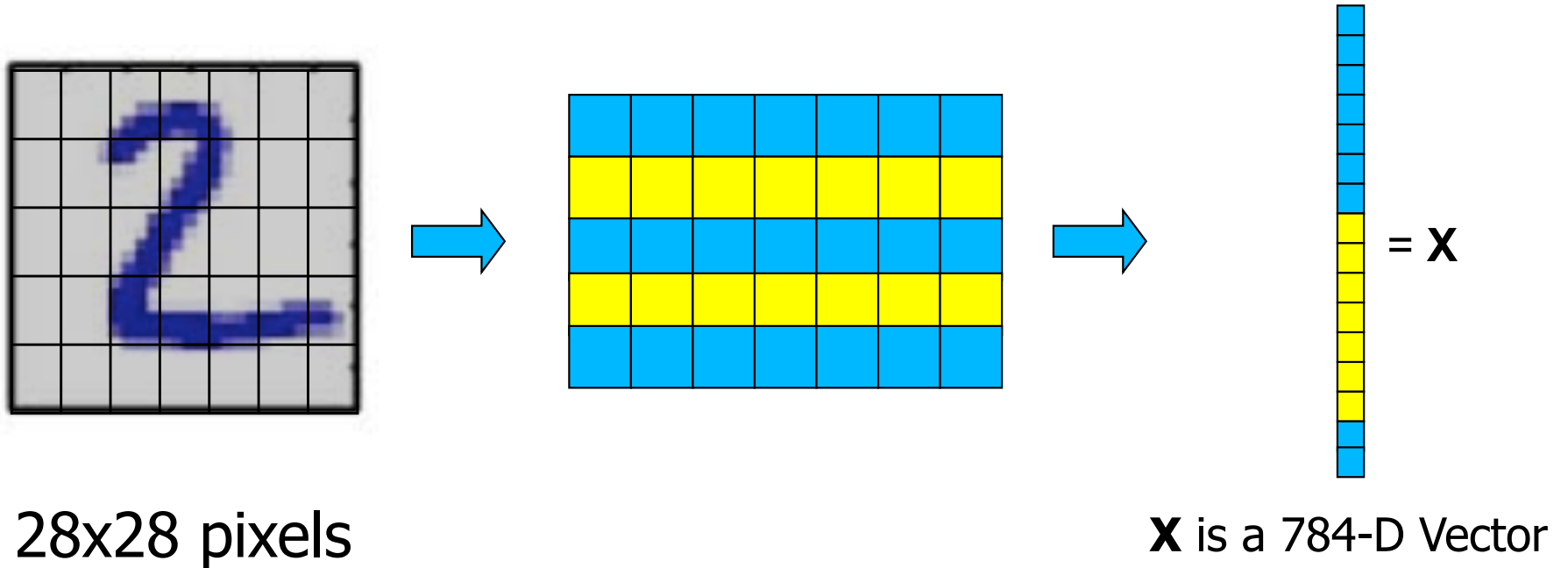
Recognizing Hand-Written Digits

LeNet (1989-1999)
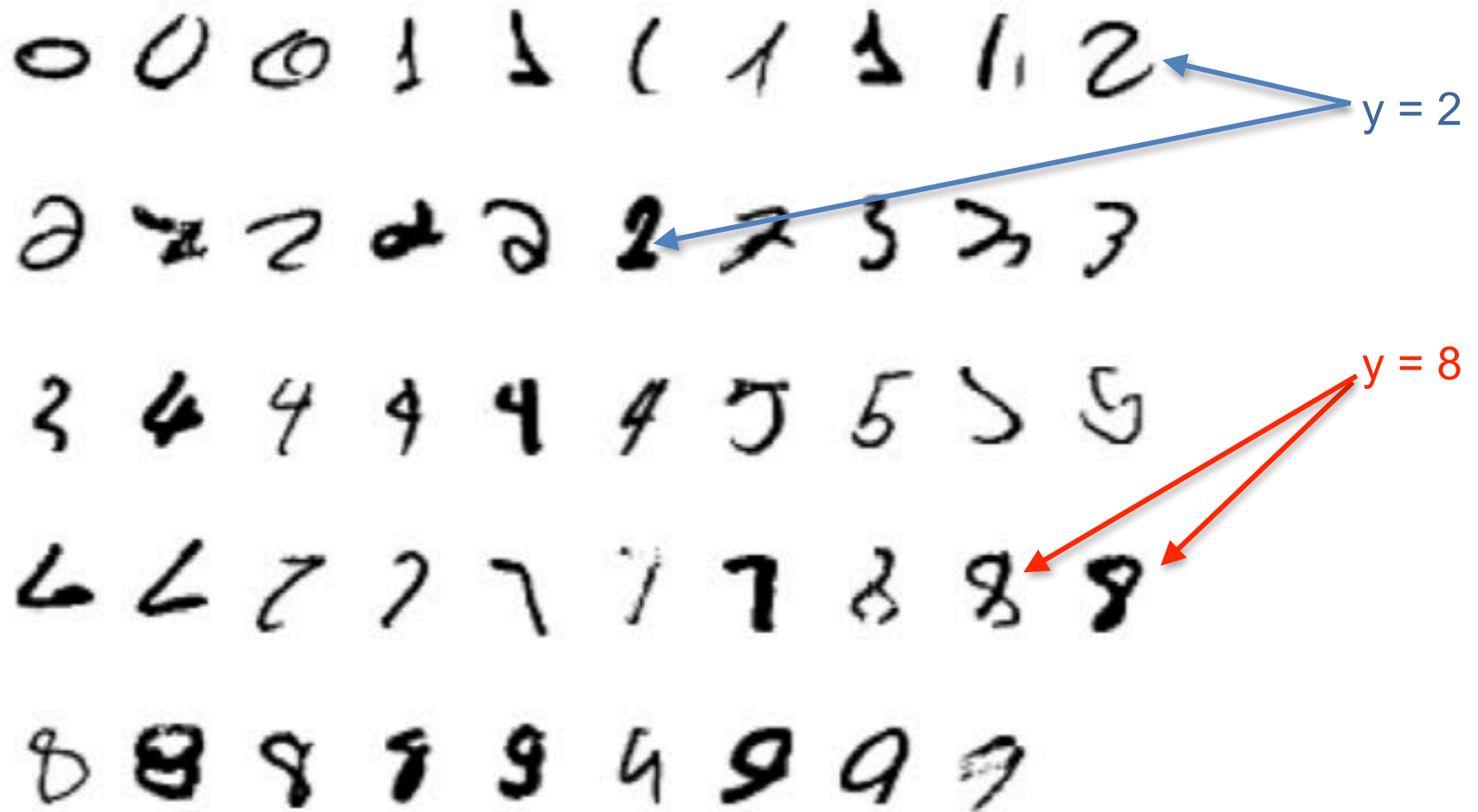
Recognizing Hand-Written Digits

Predictor and Labels

28x28 pixels

$\mathbf{x}$ is a 784-D Vector

$$y : \mathbf{x} \in \mathbb{R}^{784} \rightarrow \{0, 1, 2, \cdots, 9\} \quad ?$$

Predictor   **Labels**

Labeled Training Set

y = 2
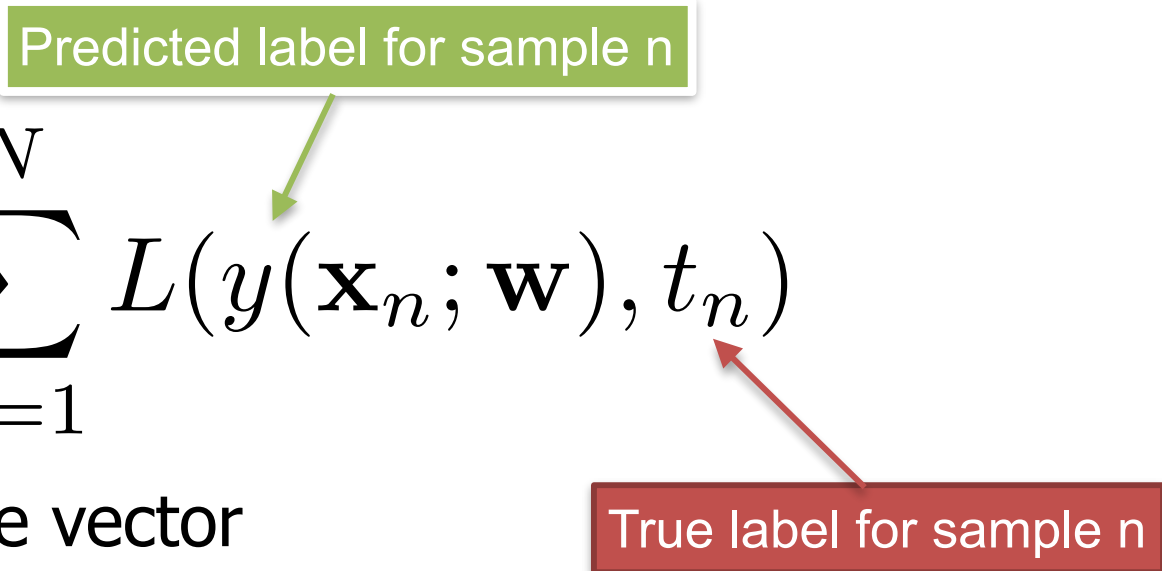
y = 8

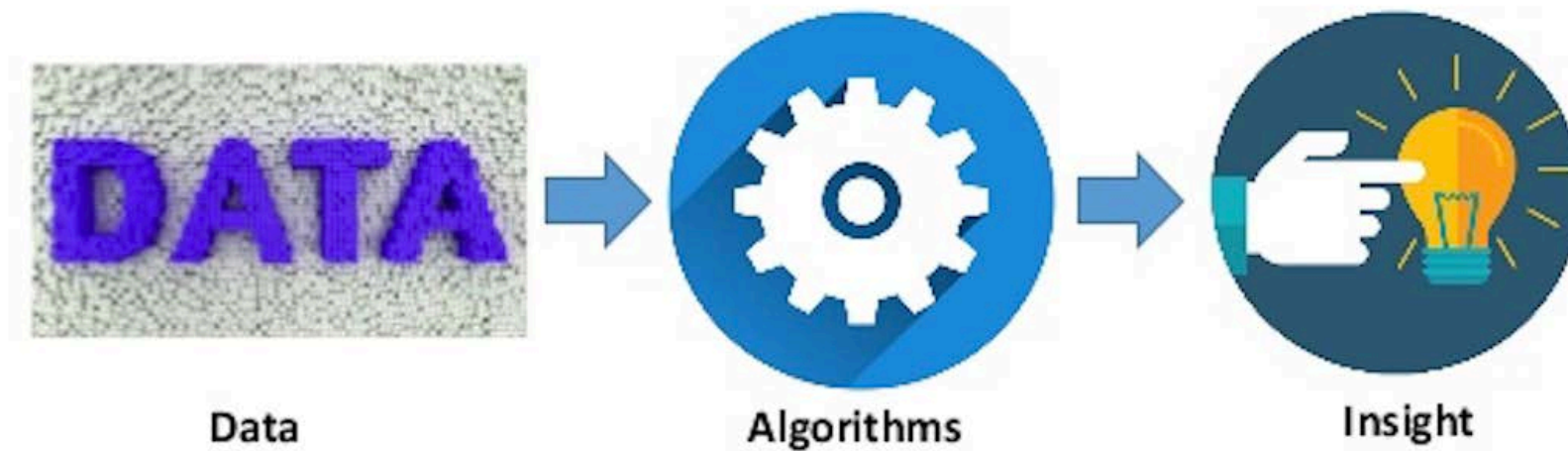$$T = \{(\mathbf{x}_n, t_n) \quad \text{for} \quad 1 \leq n \leq N\}$$

Supervised Classification

Minimize:

$$E(\mathbf{w}) = \sum_{n=1}^{N} L(y(\mathbf{x}_n; \mathbf{w}), t_n)$$

True label for sample n

- **x:** Feature vector
- **w:** Model parameters
- t: Label
- y: Predictor
- L: Loss Function
- E: Error Function

—> ML is an optimization problem

Generic Scheme

Data → Algorithms → Insight

Medical Research

Data: Feature vectors that characterize mothers.

	Age at delivery	Weight prior to pregnancy (pounds)	Smoker	Doctor visits during 1st trimester	Race	Birth Weight (grams)
Patient 1	29	140	Yes	2	Caucasian	2977
Patient 2	32	132	No	4	Caucasian	3080
Patient 3	36	175	No	0	African-Am	3600
*	*	*	*	*	*	*
*	*	*	*	*	*	*
Patient 189	30	95	Yes	2	Asian	3147

Feature vector

Image from Lumen Learning

Insight: What characteristics of a mother contribute most to low birth weight.

Spam Detection
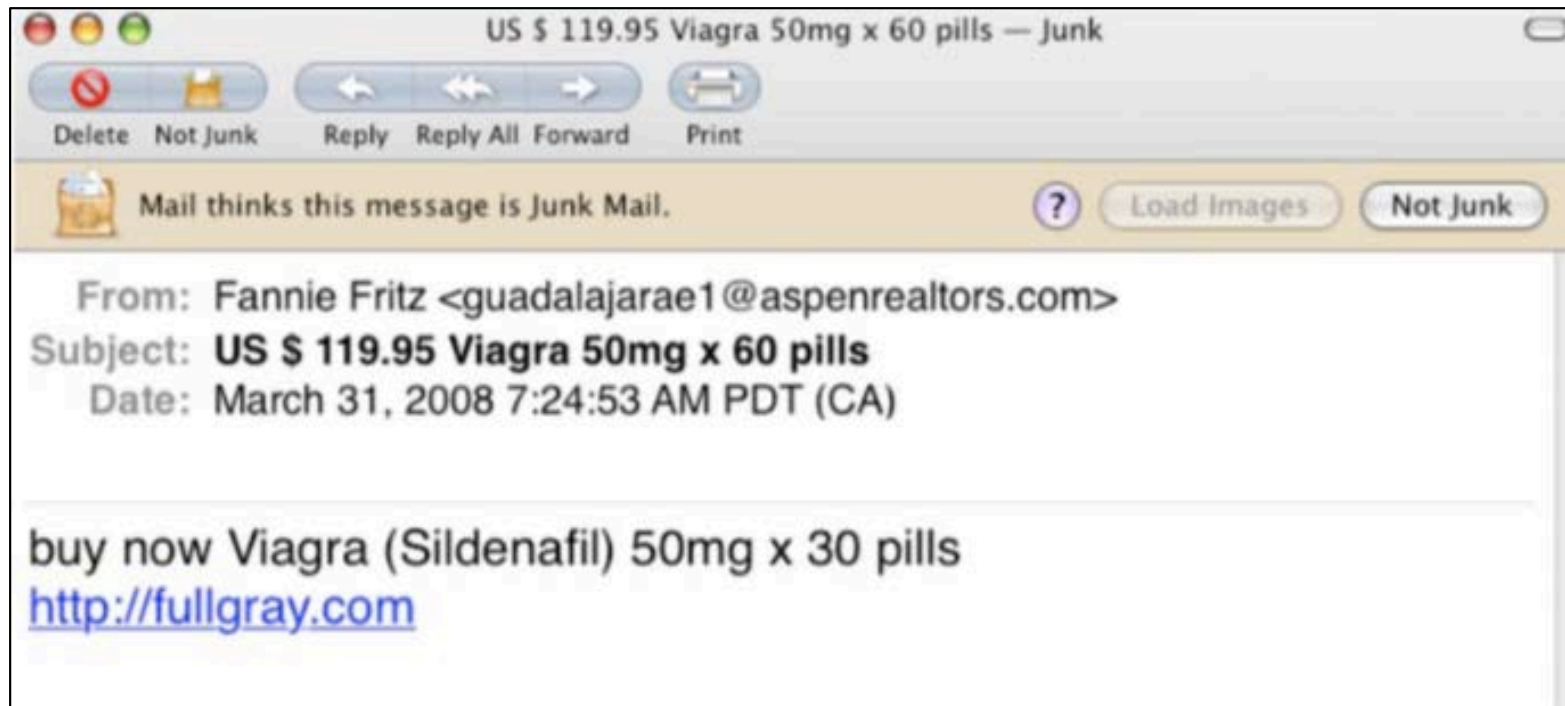
Feature vector:

$$\mathbf{x} = \begin{pmatrix} \#\mathrm{viagra} \\ \#\mathrm{pills} \\ \vdots \end{pmatrix}$$

Labels:

Spam, Not Spam

Model parameters:

$\mathbf{w}$

Predictor: $y(\mathbf{x},\mathbf{w})=\{\mathrm{Spam, Not\ Spam}\}$
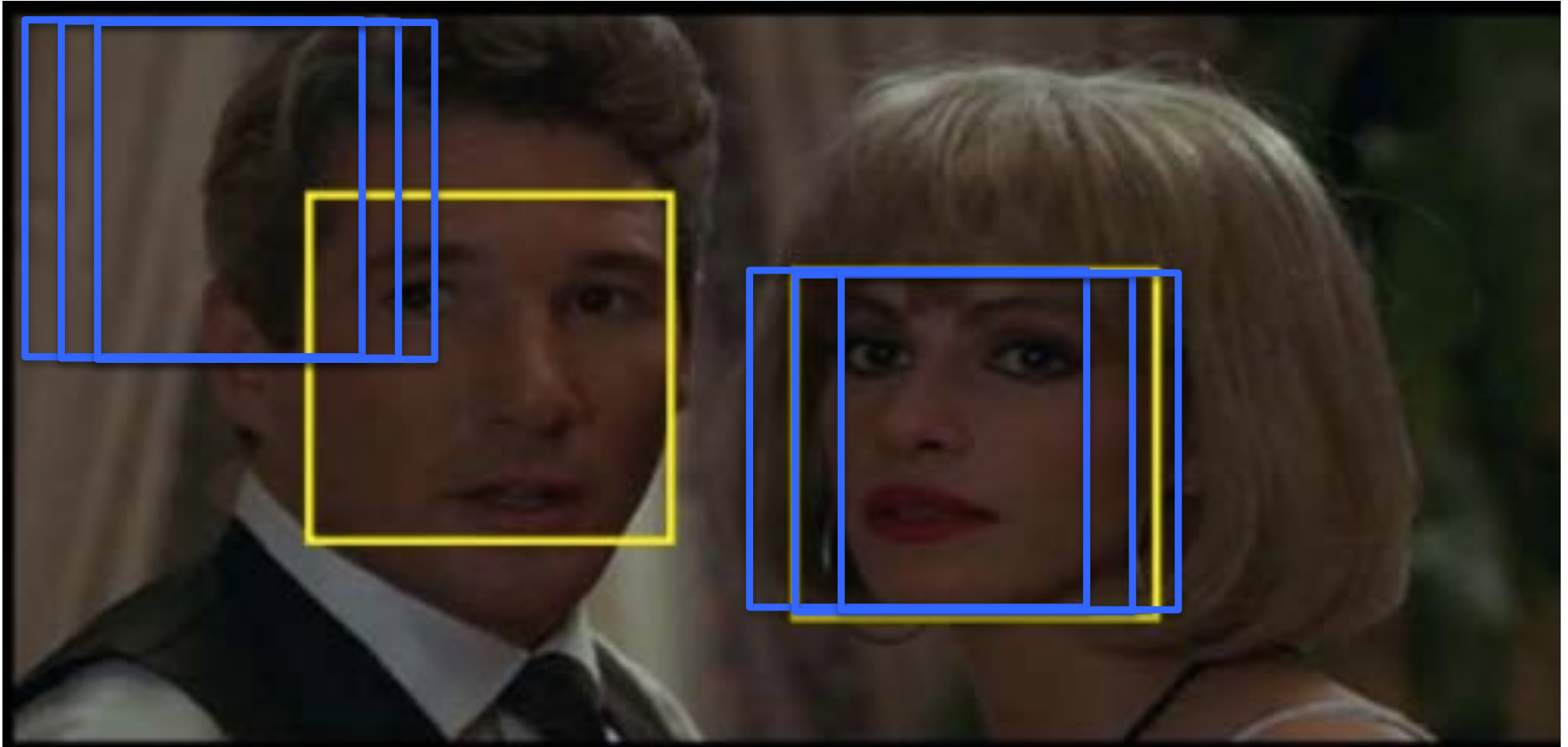
Recommender Systems

Items

Feature vector:
- What films have you watched?
- Did you like them?

Predictor:
- List of films to propose.

Face Detection

$$y : \mathbf{x} \in \mathbb{R}^{W \times H} \rightarrow v \in \{\text{face}, \text{not face}\}$$

Labeled Training Set

Faces: Near frontal with varying ages, ethnicity, gender, lighting, …

Non-faces: Images containing anything else.

Supervised Learning

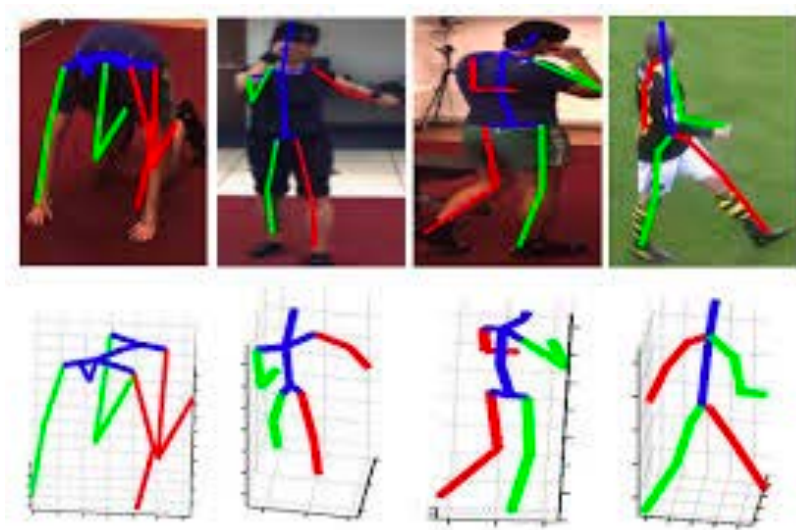Train using an annotated training set:

{(, face), (, face), (, face), … , (, not-face), (, not-face), (, not-face), …..}

Run on images that do not belong to the test set:

 → Face or not?
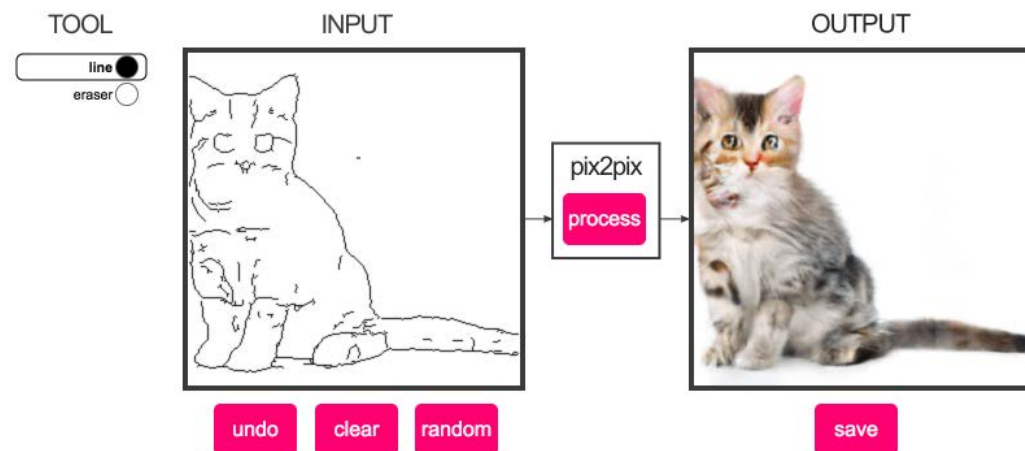
Demos

Pose Estimation

Image Synthesis

Under the IBM **Board** Corporate Governance Guidelines, the **Directors** and **Corporate Governance Committee** and the full Board annually review the financial and other relationships between the independent **director** s and IBM as part of the assessment of director independence. The Directors and Corporate Governance Committee makes recommendations to the Board about the independence of non-management directors, and the Board determines whether those directors are independent. In addition to this annual assessment of director independence, independence is monitored by the Directors and Corporate Governance Committee and the full Board on an ongoing basis.
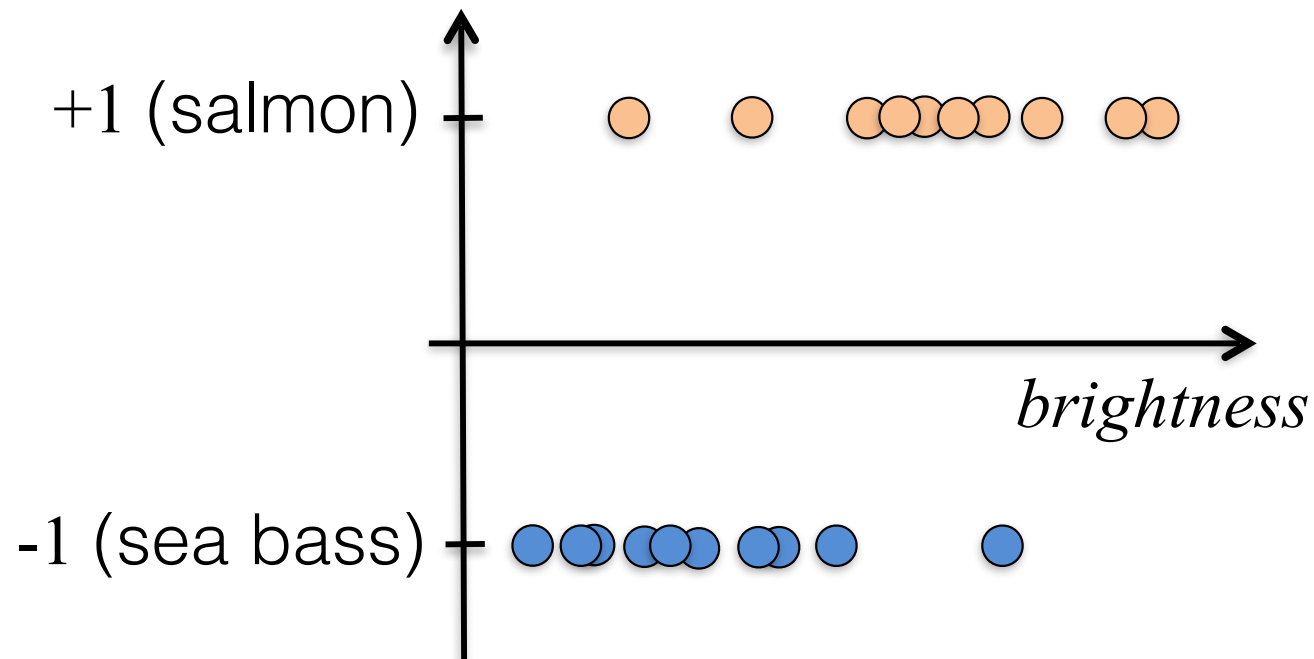
Text Analysis

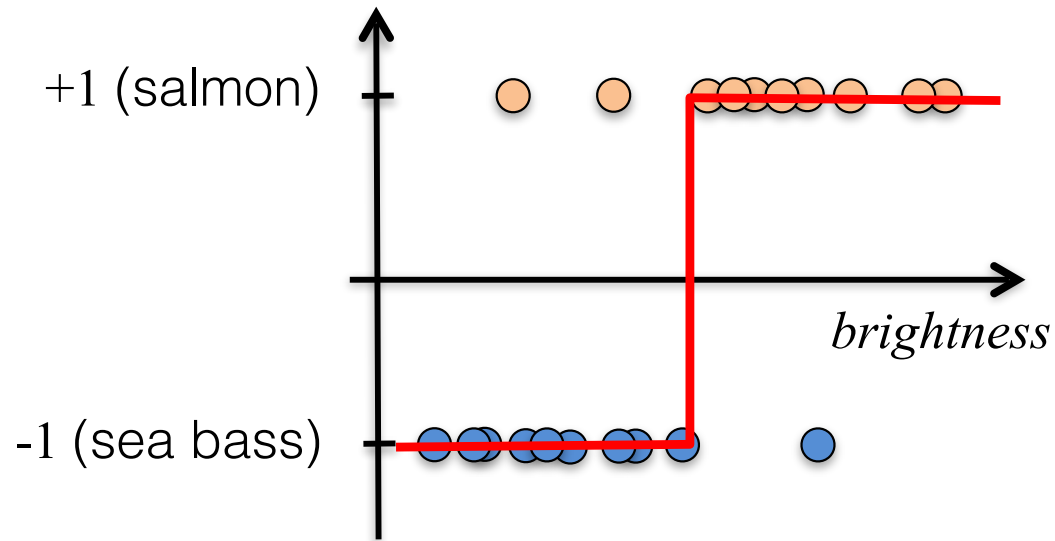# Binary Classification
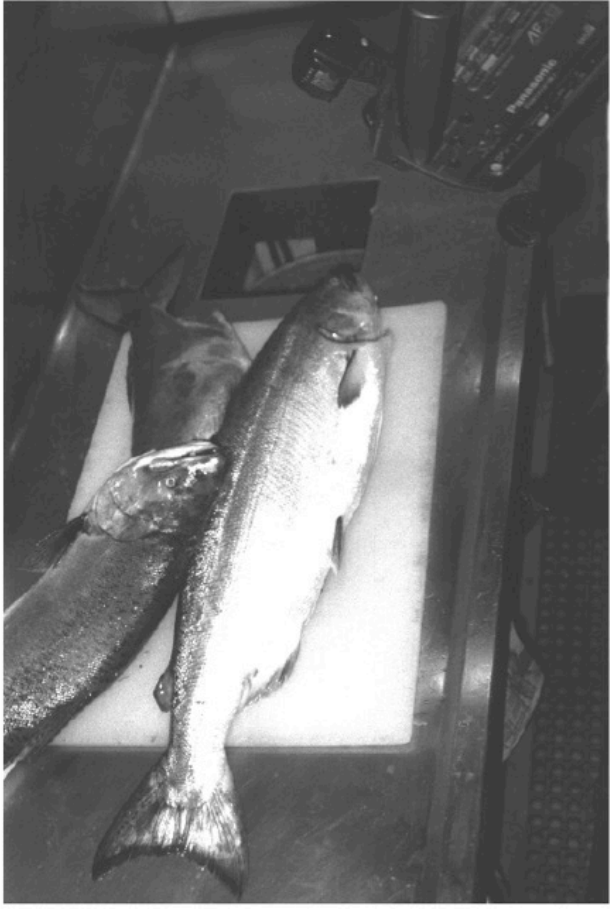
Salmon or sea bass?

Brightness as a Feature

1D Model

Model:

$$\begin{cases} -1 & \text{if brightness } < T \\ 1 & \text{otherwise} \end{cases}$$

Model parameters:

$$\mathbf{w} = \{T\}$$
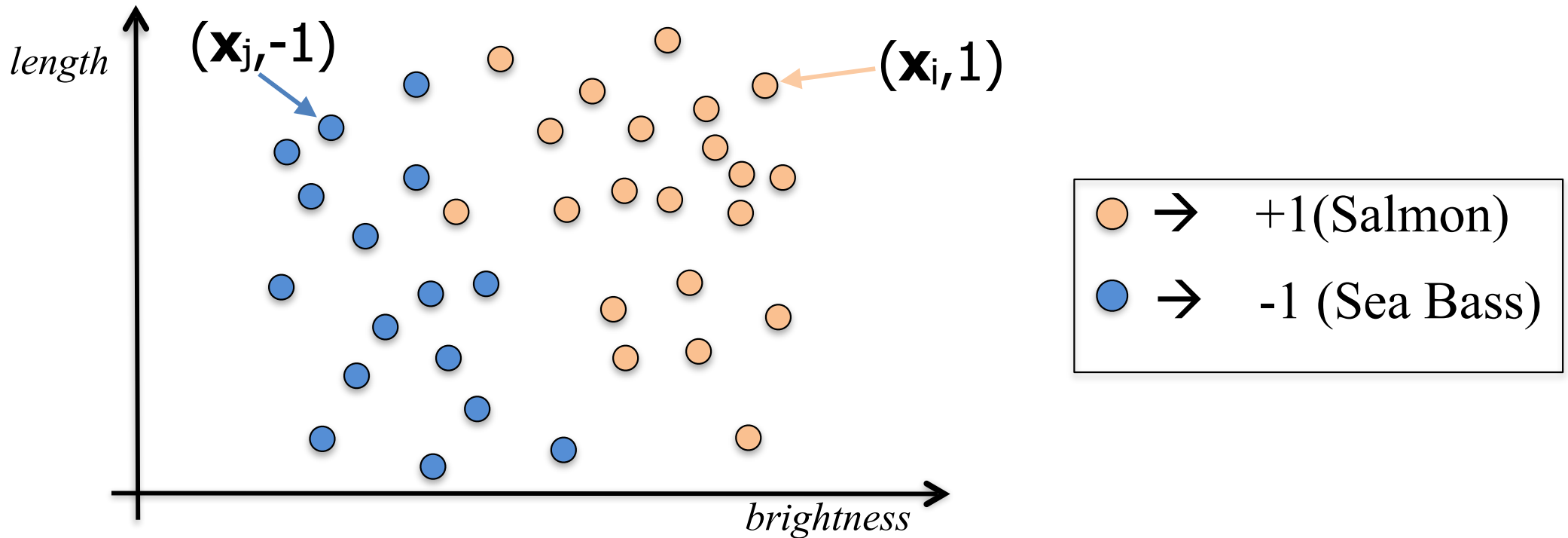
+1 (salmon)

-1 (sea bass)

brightness

Error function:

$$E(\mathbf{w}) = \#\text{salmons with brightness } < T + \#\text{seabasses with brightness } \geq T$$

Learning: Minimizing E(**w**) w.r.t. to **w**

2D Model

$$\text{Some algorithm} \longrightarrow \begin{pmatrix} brightness \\ length \end{pmatrix}$$

$(\mathbf{x}_j, -1)$

$(\mathbf{x}_i, 1)$

length

brightness

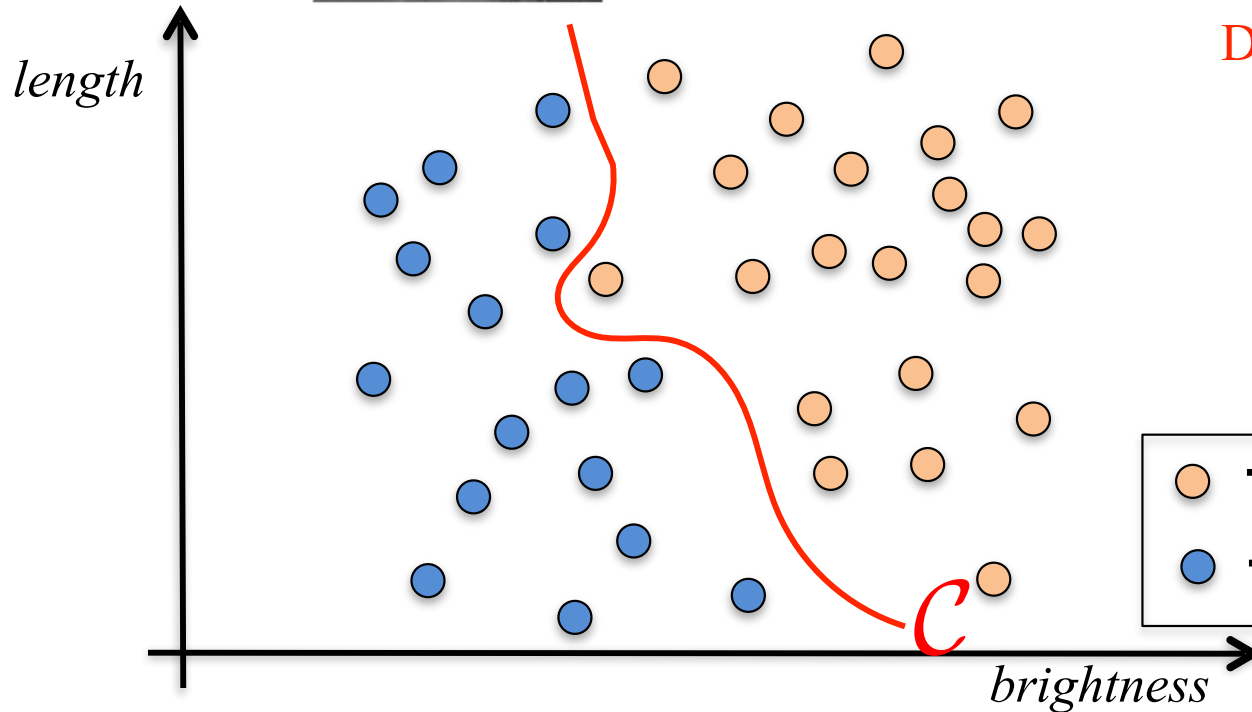- 🟠 → +1(Salmon)
- 🔵 → -1 (Sea Bass)

Each circle denotes a 2-dimensional sample $\mathbf{x}_i$ of dimension 2, which is assigned a label $t_i \in \{-1, 1\}$.

Decision Boundary

Some algorithm $\longrightarrow$ $\begin{pmatrix} brightness \\ length \end{pmatrix}$

Decision boundary:

$$\mathcal{C} = \{\mathbf{x} \in R^2, y(\mathbf{x}, \mathbf{w}) = 0\}$$
$$y(\mathbf{x}, \mathbf{w}) > 0 : Salmon$$
$$y(\mathbf{x}, \mathbf{w}) < 0 : Bass$$

$\bigcirc \rightarrow +1$

$\bigcirc \rightarrow -1$

$\mathcal{C}$

In this class, we will discuss:
- How to define y.
- How to choose **w**.

Training vs Testing

Some algorithm $\longrightarrow$ $\left(\begin{array}{c} brightness \\ length \end{array} \right)$
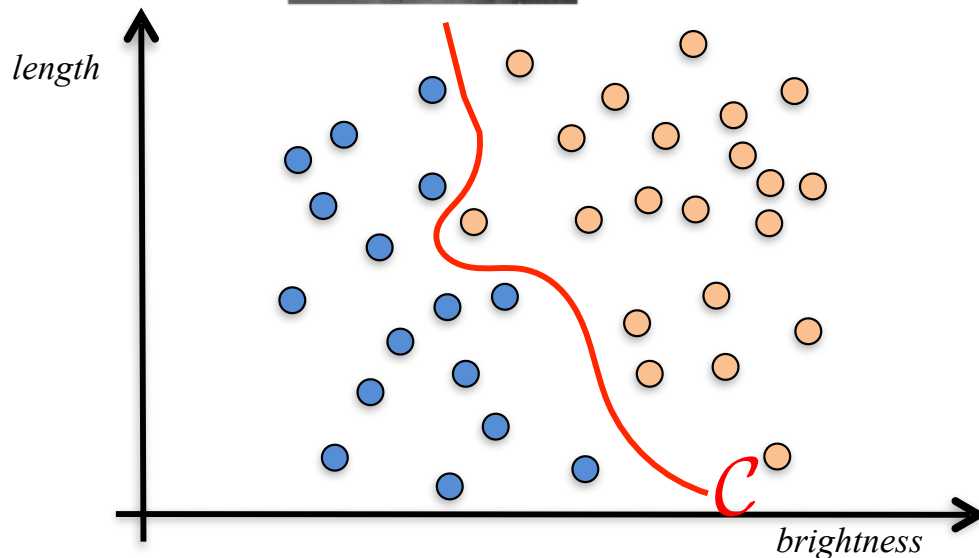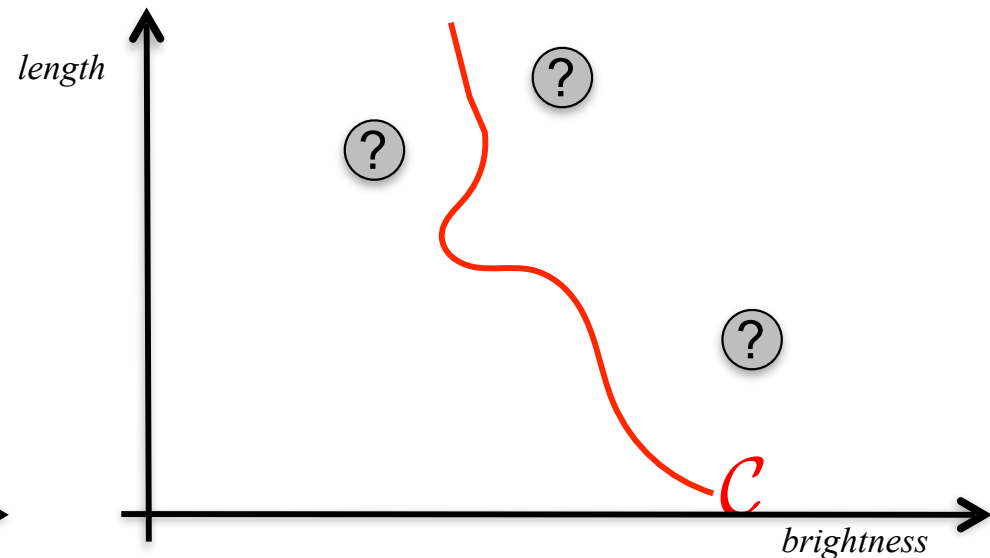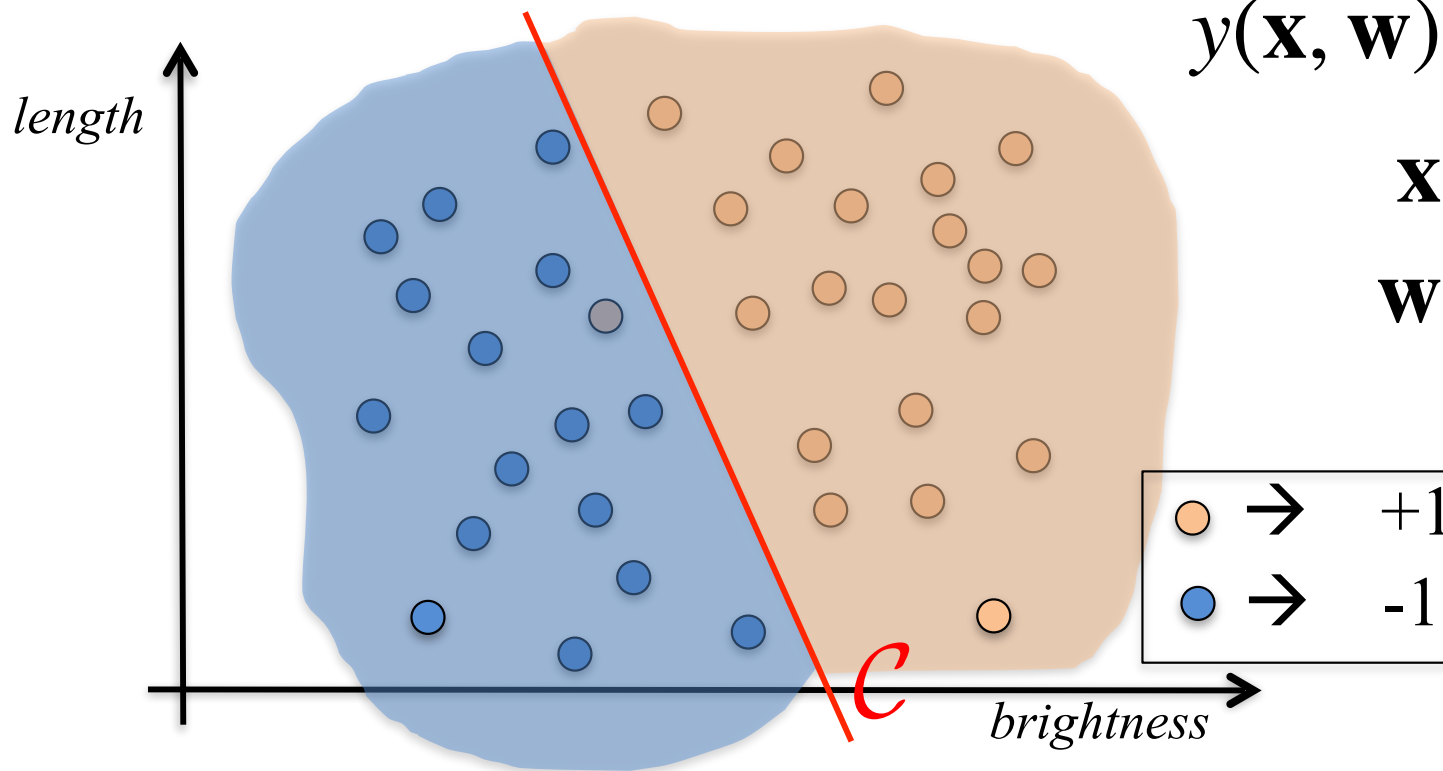
Training set = $\{\bullet, \bullet\}$

Test set = $\{?,?,?, ...\}$

1. Use the training set to learn the model.
2. Measure performance on the test set.
   —> Makes sense only if both have the same statistical distribution.

Linear 2D Model

Some algorithm $\longrightarrow$ $\begin{pmatrix} brightness \\ length \end{pmatrix}$

$$y(\mathbf{x}, \mathbf{w}) = w_0 + w_x b + w_y l$$

$$\mathbf{x} = [b, l]$$

$$\mathbf{w} = [w_0, w_x, w_y]$$

—> Important special case.

Course Outline

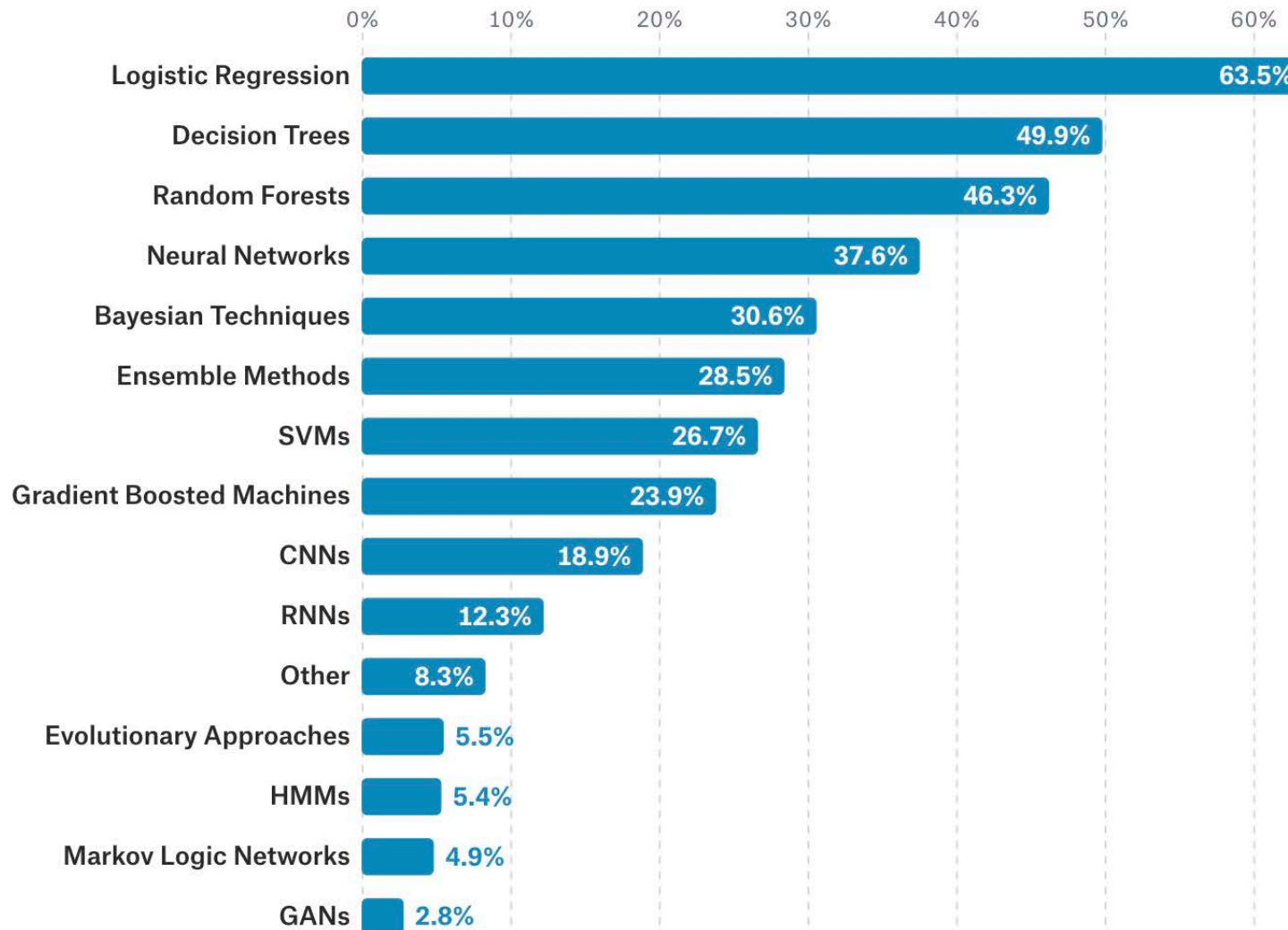**Introduction**
- ML Basics
- K Nearest Neighbors
- K Means

Linear ML
- Linear Regression
- Logistic Regression
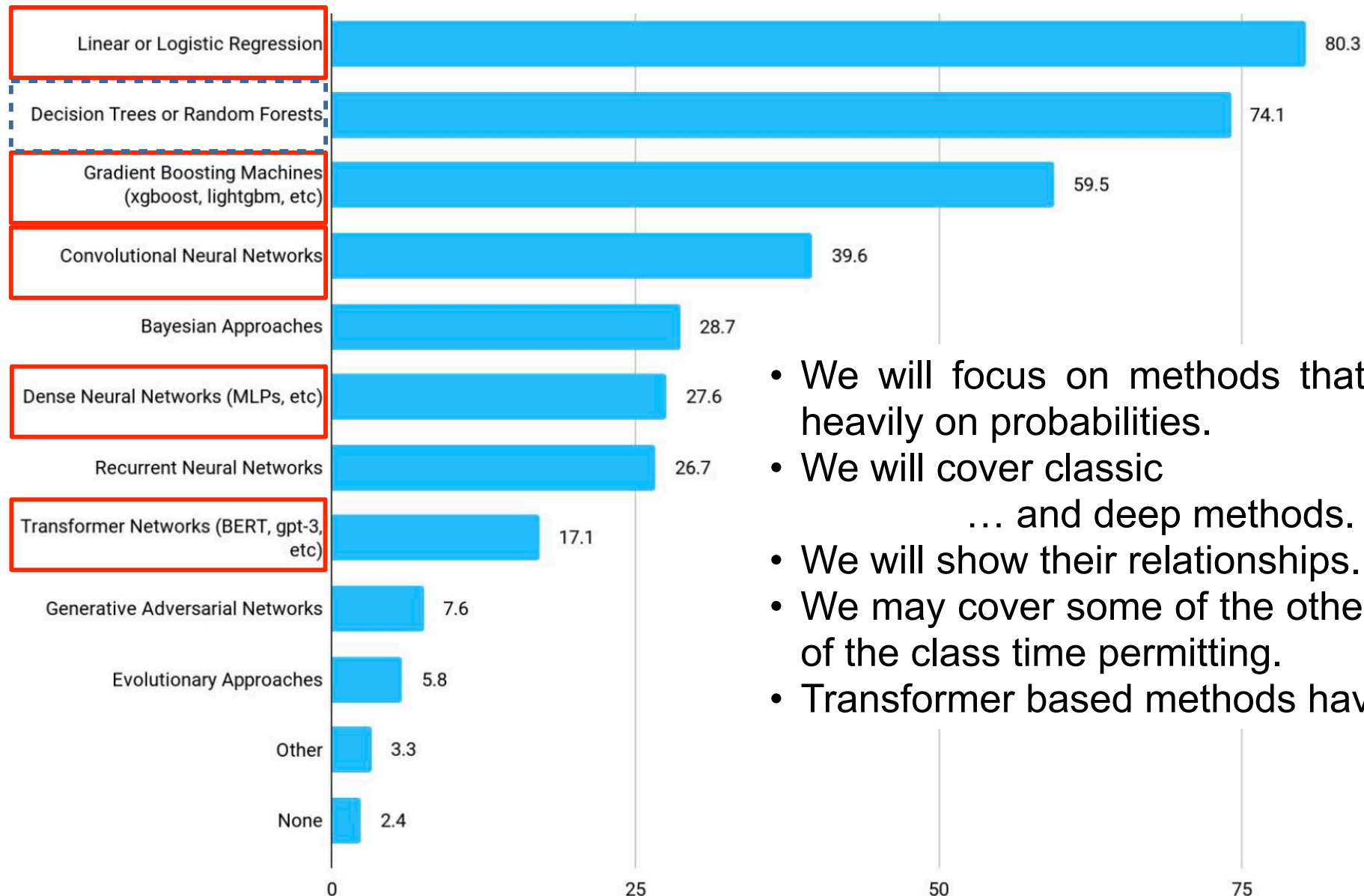- Max Margin Classifiers

Nonlinear ML
- AdaBoost
- Polynomial Support Vector Machines
- Kernel Methods
- Artificial Neural Networks

Kaggle Survey (2017)


What data science methods do you use at work?

Kaggle Survey (2021)

- We will focus on methods that do not rely heavily on probabilities.
- We will cover classic
  … and deep methods.
- We will show their relationships.
- We may cover some of the others at the end of the class time permitting.
- Transformer based methods have taken off.

What data science methods do you use at work?

Kaggle Survey (2022)
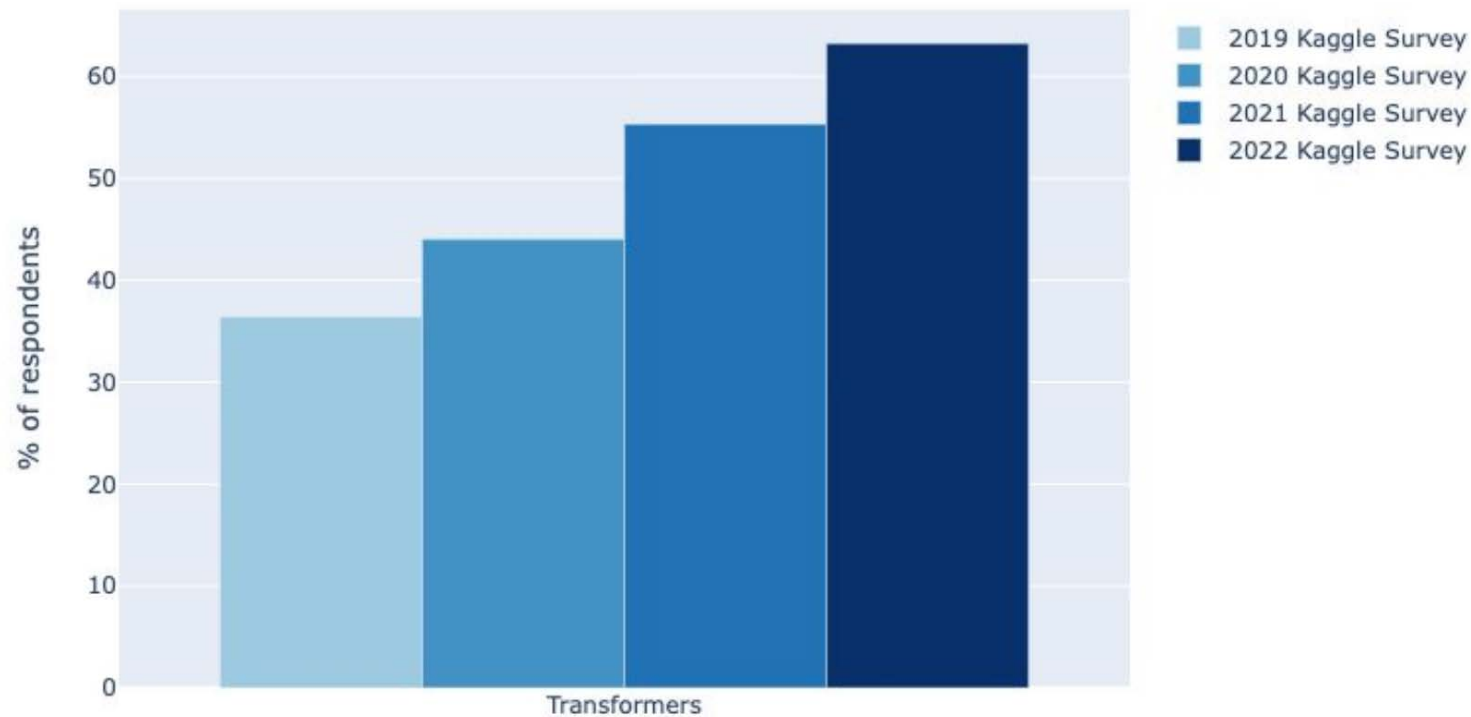
- Transformer architectures are becoming more popular for deep learning models?
- Will they dominate?
- ➡ Time will tell

General Organization

- Lectures: Tuesdays   8:15-10

- Exercises: Tuesdays 17:15-19 (Pen and paper)

- Projects: Fridays 15.15-17.   (Coding)

- Two milestones will be graded (10% of grade, each).

- Written exam (80% of grade).

- Slides: https://moodle.epfl.ch/course/view.php?id=16159

- Main references:

    - C.M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006

    - M. Welling, A First Encounter with Machine Learning, 2011

Recommended Book
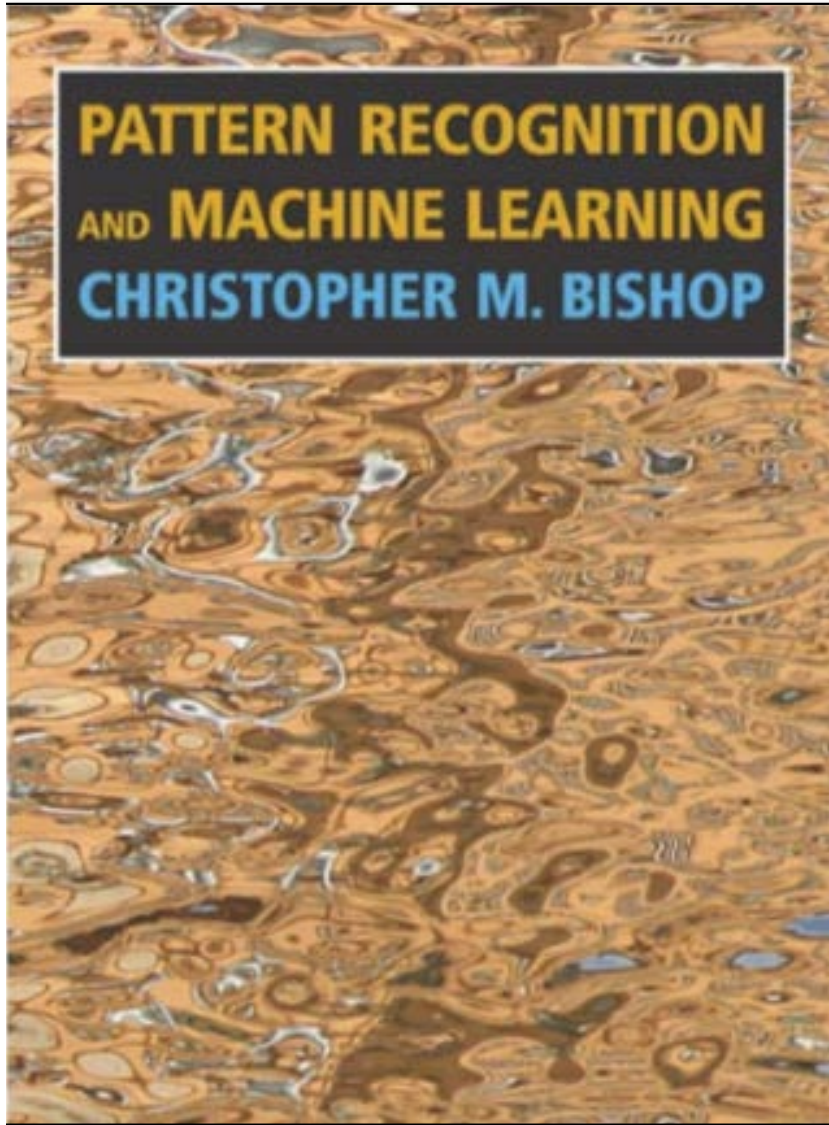
Pattern Recognition and Machine Learning.

Christopher Bishop, Springer, 2006.

- Available for free on the web.
- We will use the same notations.

https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf

Slide Codes

Training vs Testing

Training vs Testing

Reminder

Training vs Testing

Optional Bishop, xxx

Normal slide: It is part of the course and I may ask exam questions about it.

Reminder slide: We have already covered this earlier in the class. Go back to the appropriate lecture if you do not remember.

Optional slide: This is additional material for people interested in more details. I will not ask direct exam questions on this.

Reference to book or paper for even more details.

Moodle Page

Shape from Stereo

Recovering Depth from Multiple Images

📄 Shape from Stereo

📄 Nerfs

📄 Computer Vision: Algorithms and Applications. Chapter 12.

Optional