Introduction to Machine Learning

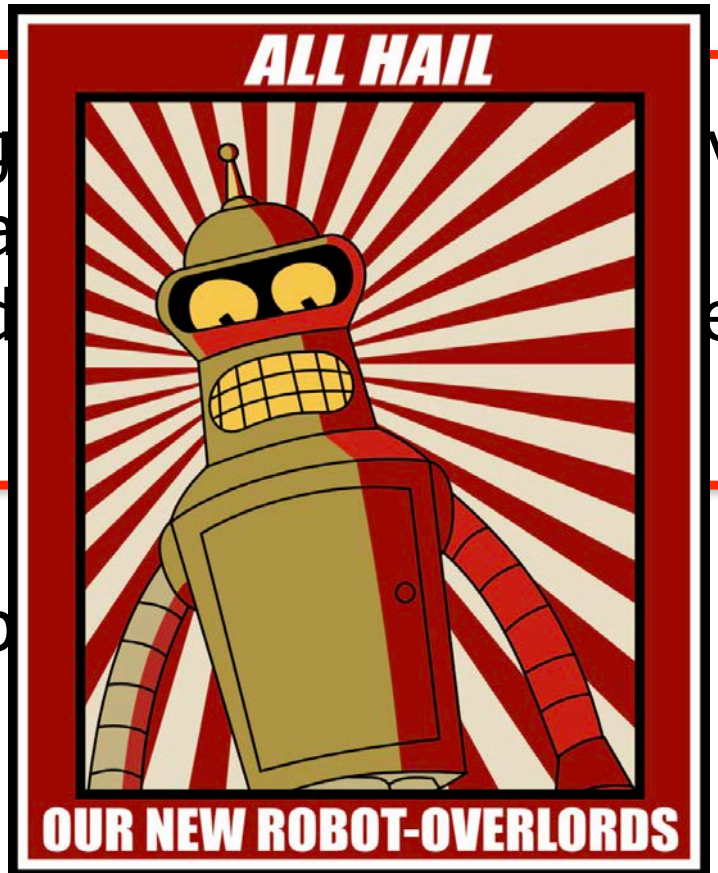
Pascal Fua
IC-CVLab

Human vs Machine Learning

What is Machine Learning?

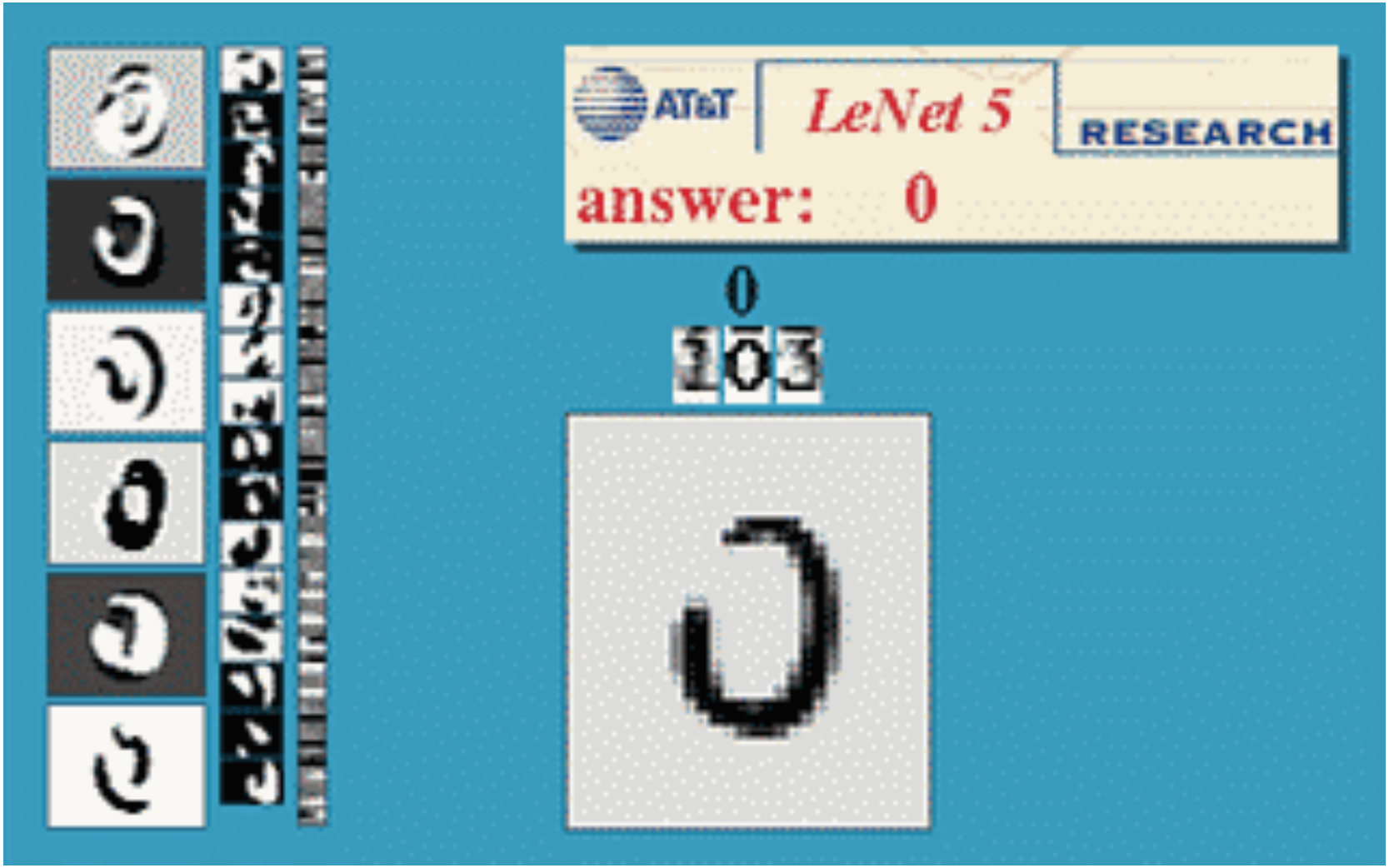
- Machine Learning is the science of getting computers to learn and act like humans do, and improve their learning over time in autonomous fashion, by feeding them data and information in the form of observations and real-world interactions.

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



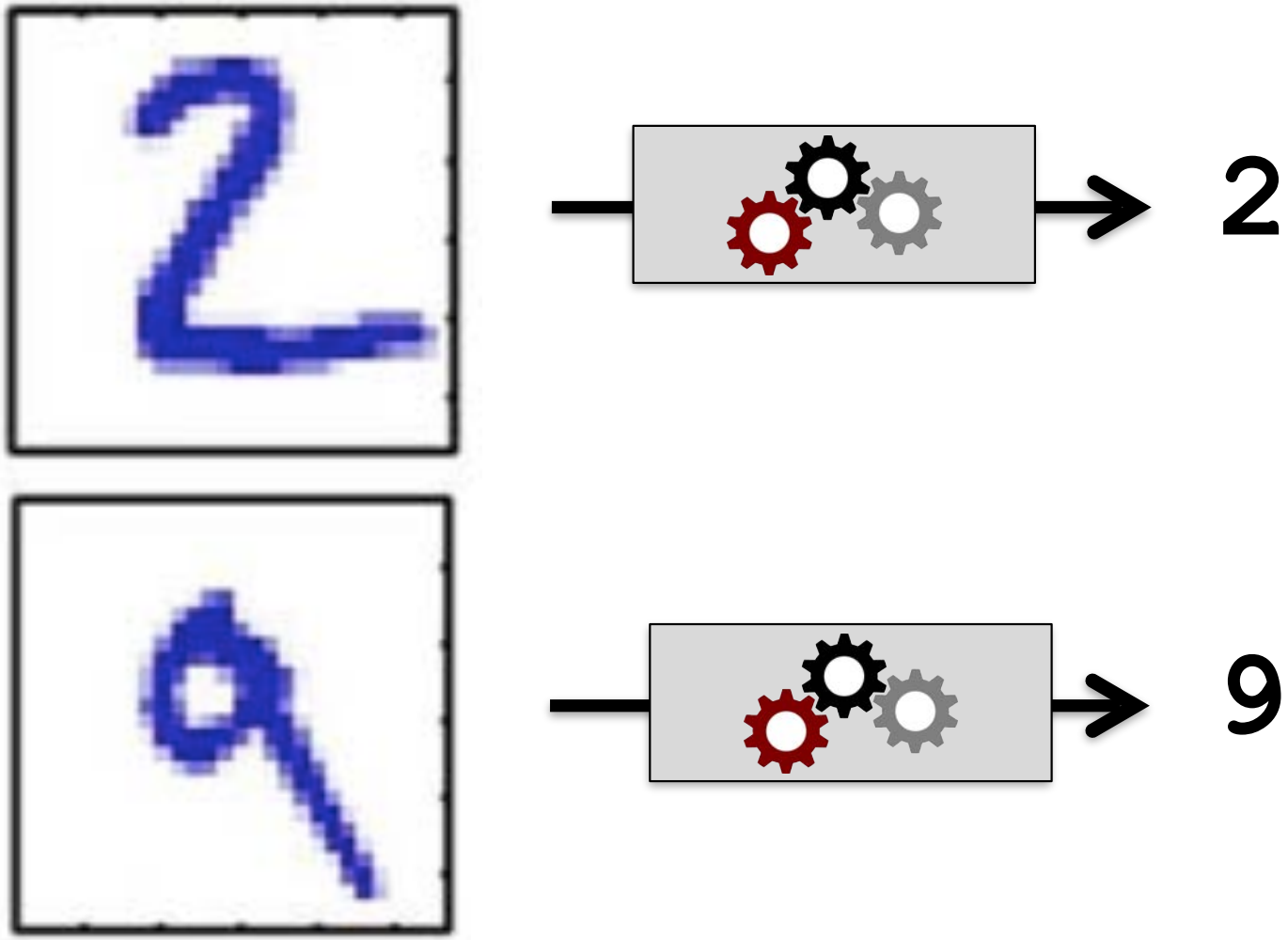
- Machine learning is ap

Recognizing Hand-Written Digits

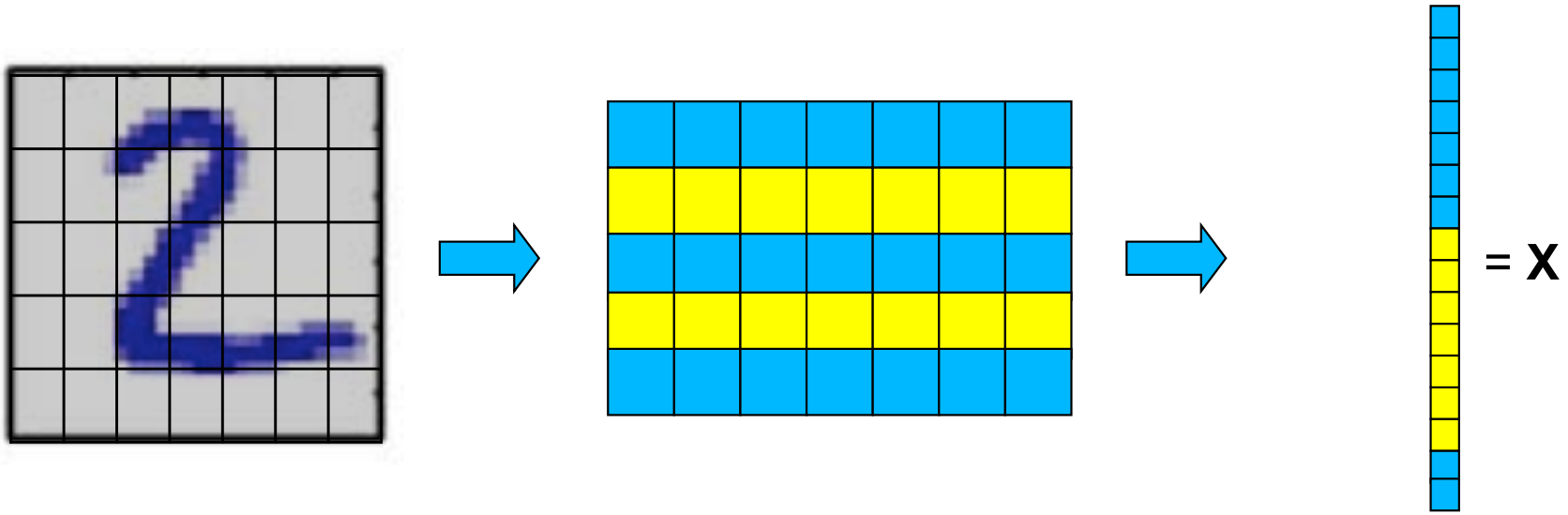


LeNet (1989-1999)

Recognizing Hand-Written Digits



Predictor and Labels



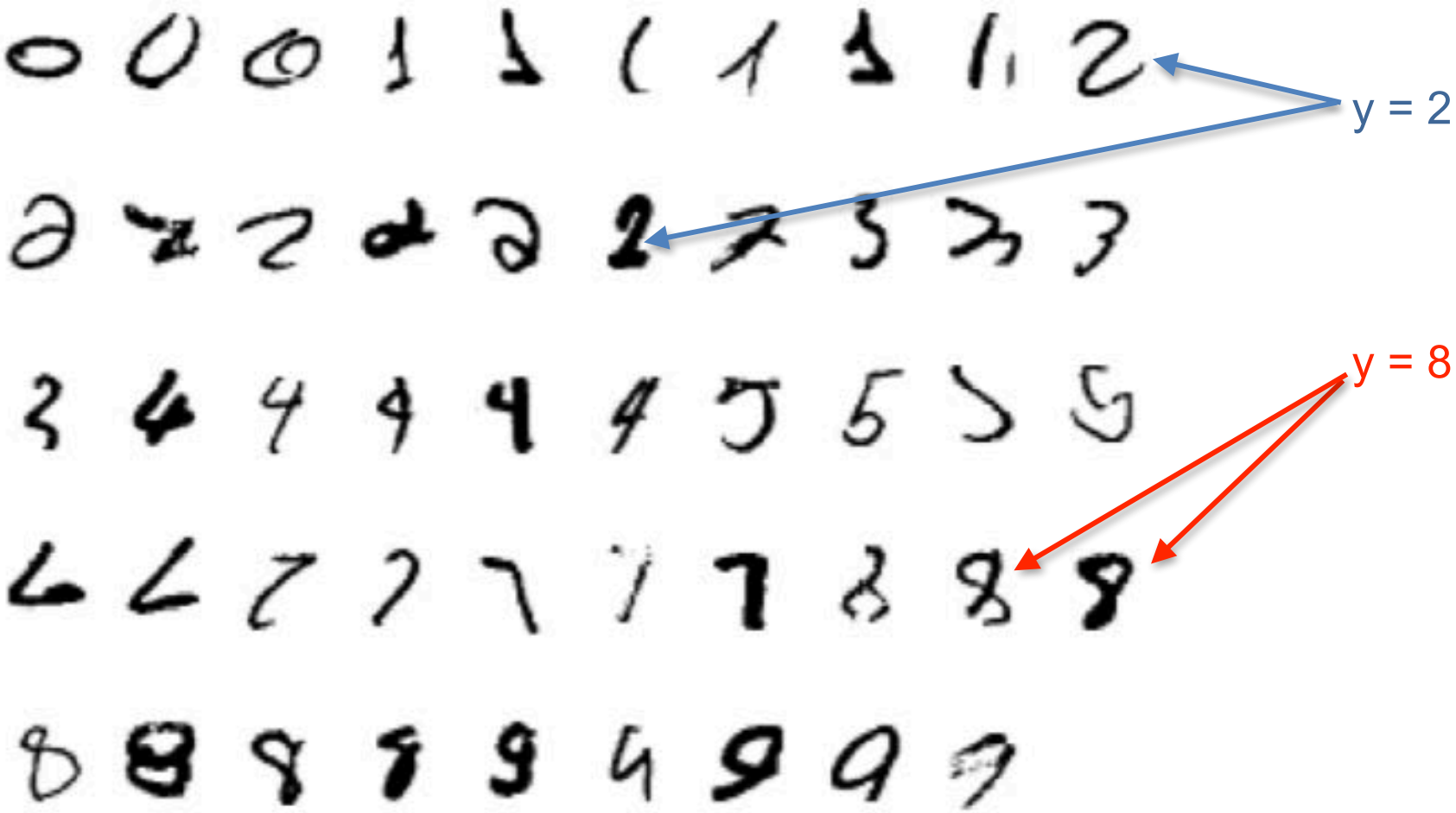
28x28 pixels

X is a 784-D Vector

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

Predictor **Labels**

Labeled Training Set



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

Supervised Classification

Minimize:

Predicted label for sample n

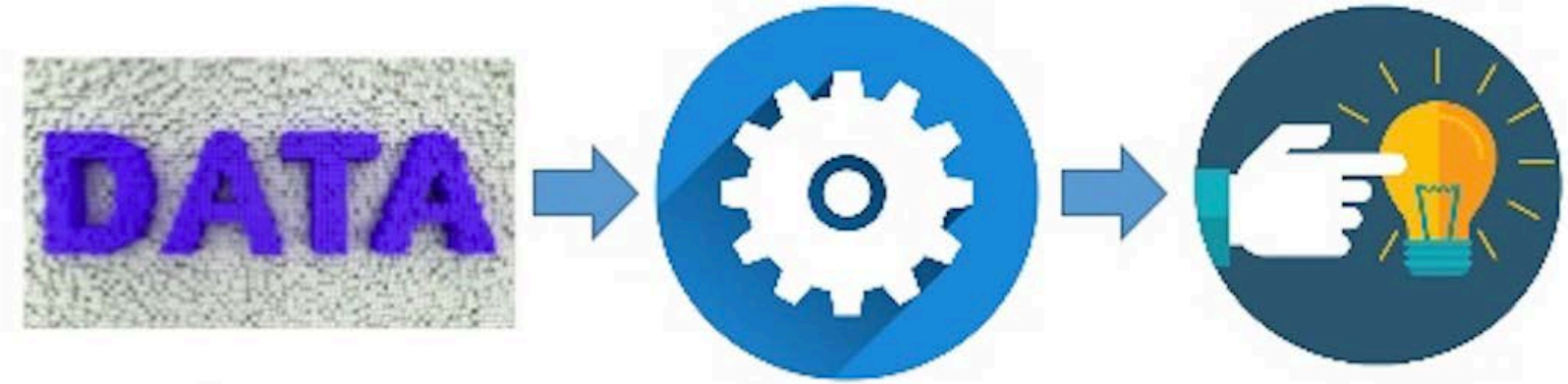
$$E(\mathbf{w}) = \sum_{n=1}^{N} L(y(\mathbf{x}_n; \mathbf{w}), t_n)$$

True label for sample n

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

—> ML is an optimization problem

Generic Scheme



Medical Research

Data: Feature vectors that characterize mothers.

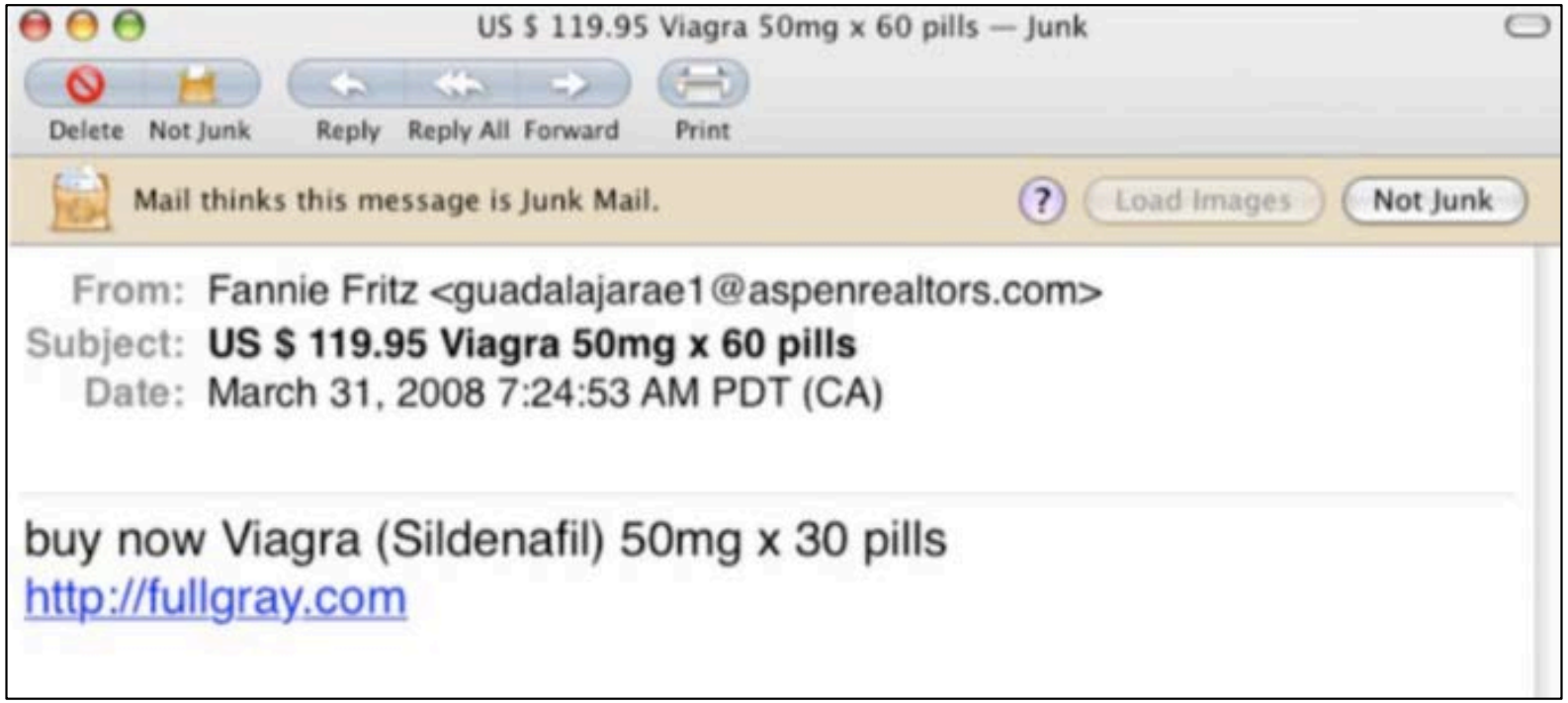
	Age at delivery	Weight prior to pregnancy (pounds)	Smoker	Doctor visits during 1st trimester	Race	Birth Weight (grams)
Patient 1	29	140	Yes	2	Caucasian	2977
Patient 2	32	132	No	4	Caucasian	3080
Patient 3	36	175	No	0	African-Am	3600
*	*	*	*	*	*	*
*	*	*	*	*	*	*
Patient 189	30	95	Yes	2	Asian	3147

Feature vector

Image from Lumen Learning

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

Spam Detection



Feature vector:

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

Labels:

Spam, Not Spam

Model parameters:

w

Predictor: y(**x**,**w**)={Spam, Not Spam}

Recommender Systems

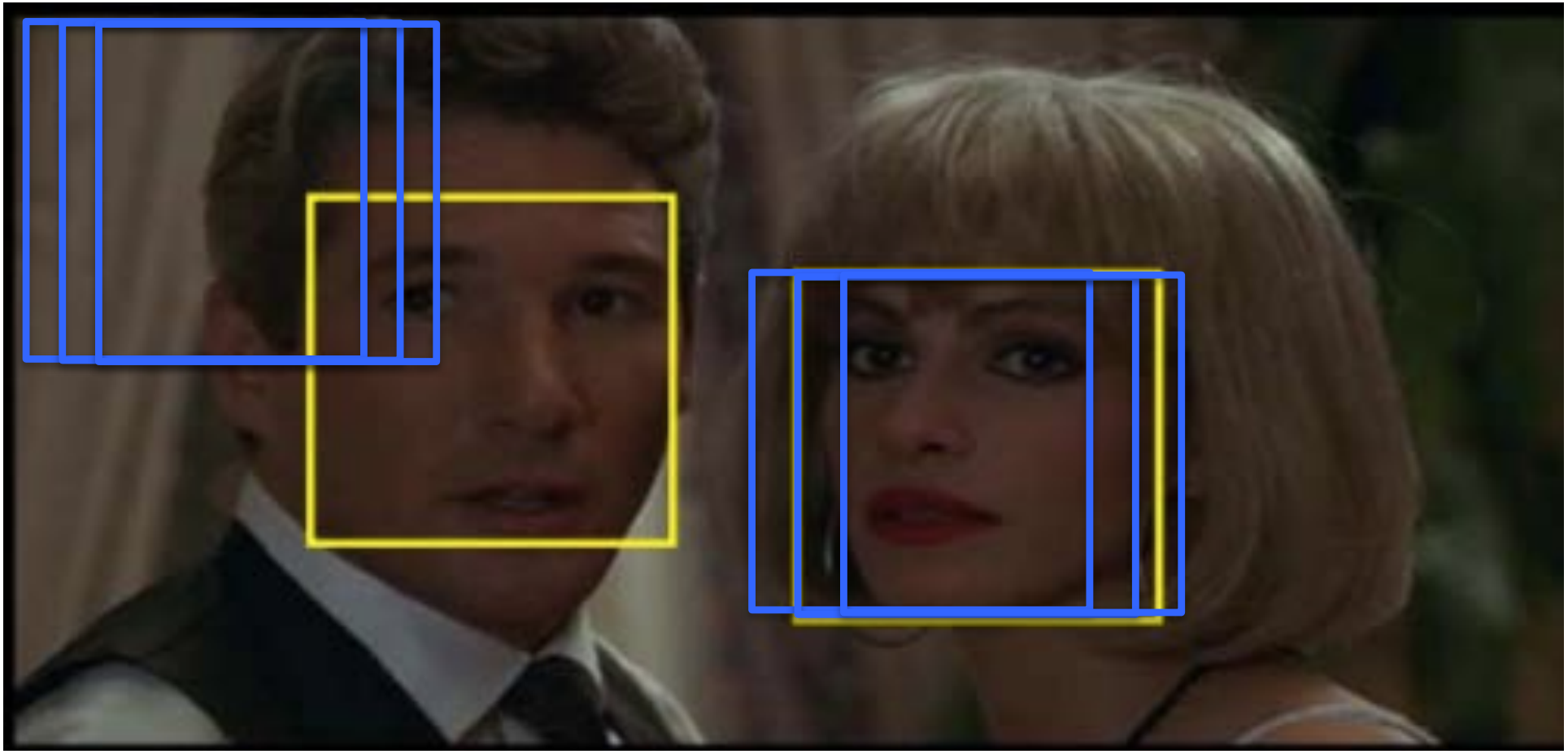
Feature vector:

- What films have you watched?
- Did you like them?

Predictor:

- List of films to propose.

Face Detection



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

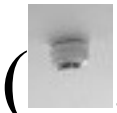
Labeled Training Set

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

Non-faces: Images containing anything else.

Supervised Learning

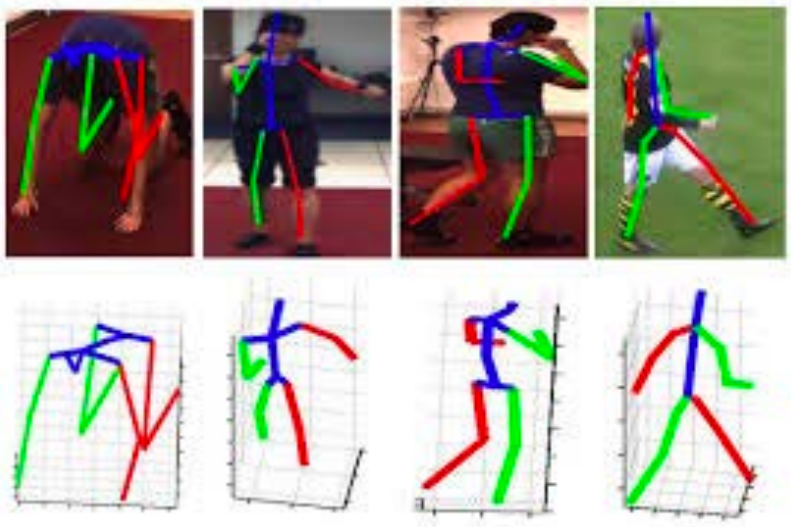
Train using an annotated training set:

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

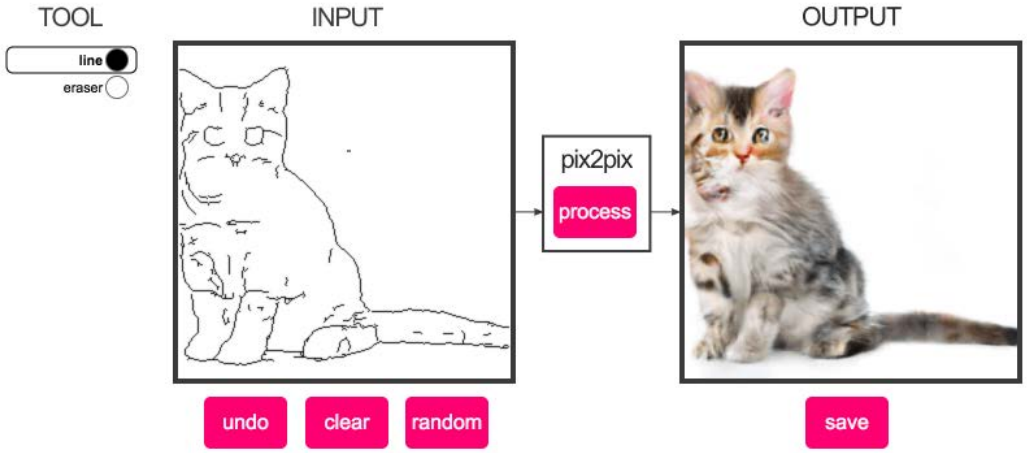
Run on images that do not belong to the test set:

 → Face or not?

Demos



[Pose Estimation]

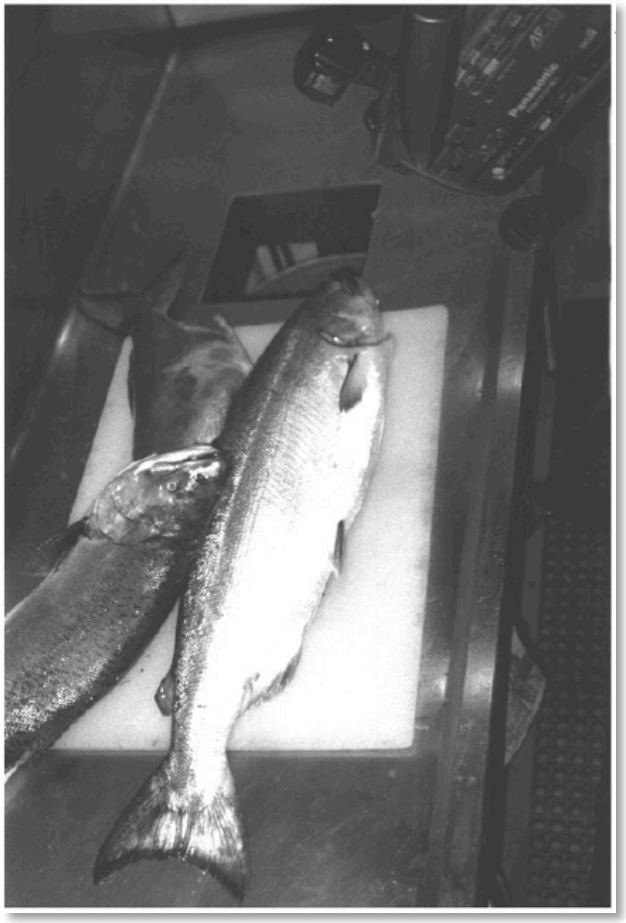


[Image Synthesis]

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

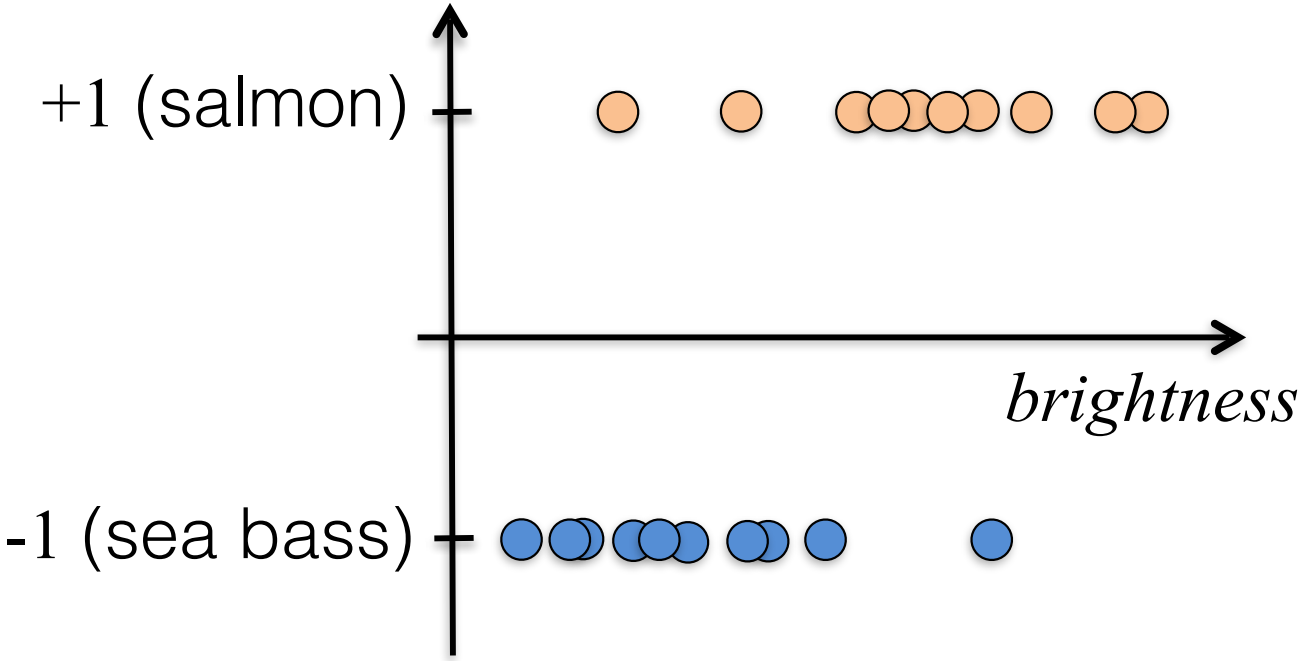


Salmon or sea bass?

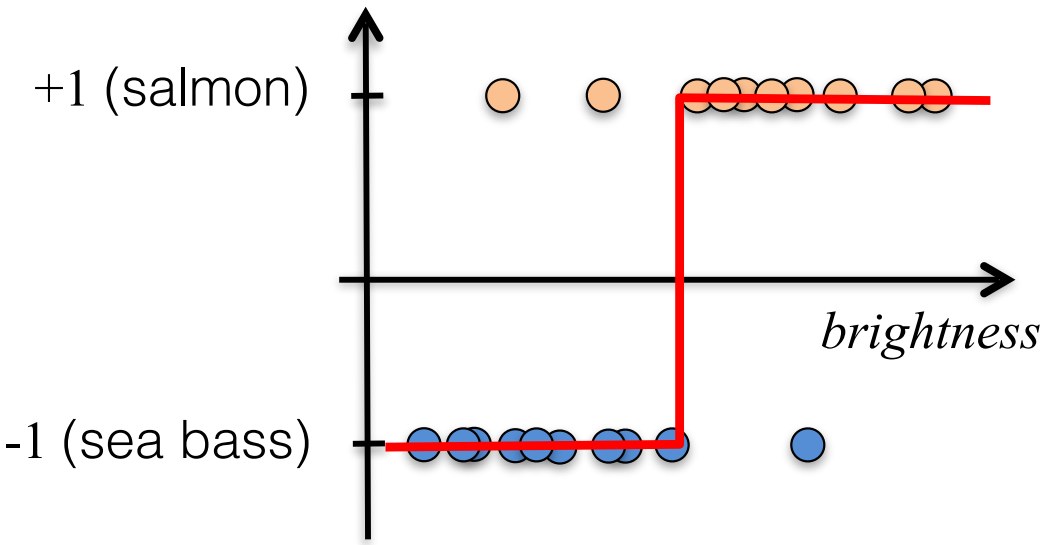
Brightness as a Feature

Some algorithm

brightness



1D Model



Model:

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

Model parameters:

$$\mathbf{w} = \{T\}$$

Error function:

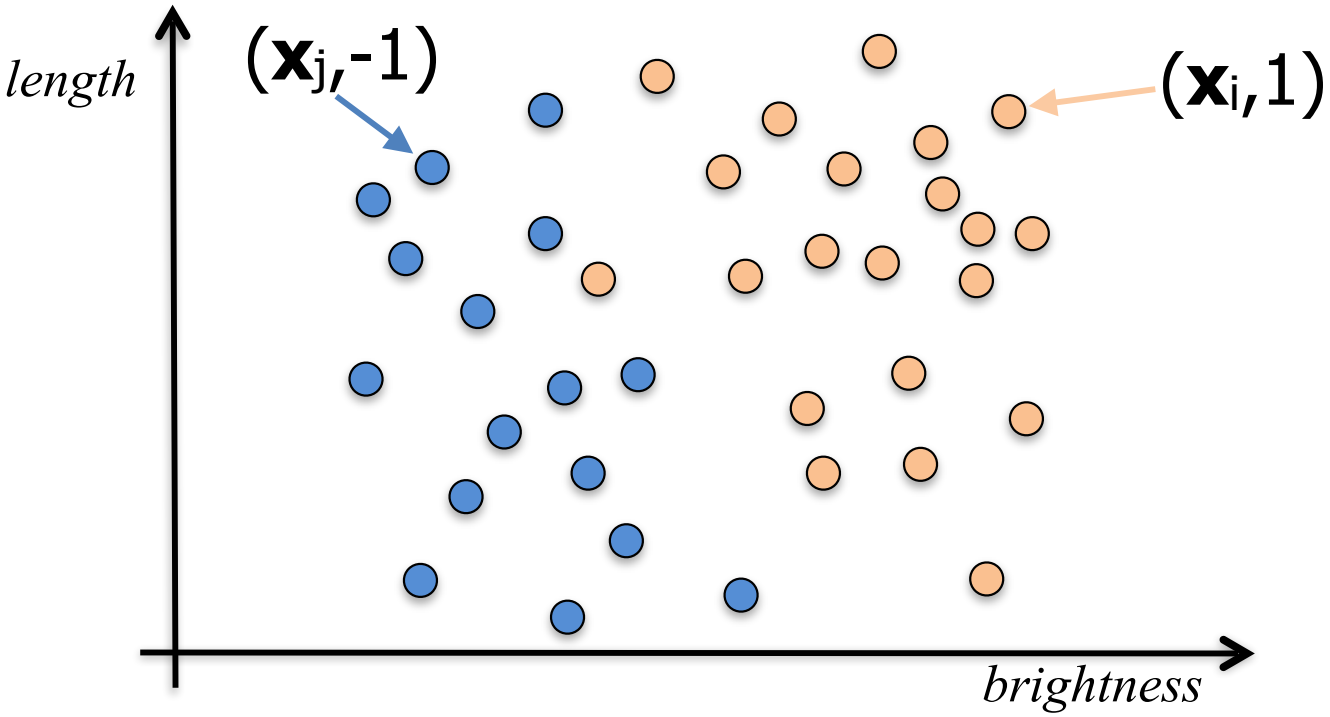
$$E(\mathbf{w}) = \#\text{salmons with brightness } < T + \#\text{seabasses with brightness } \geq T$$

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

2D Model

Some algorithm →

$$\begin{pmatrix} brightness \\ length \end{pmatrix}$$



● → +1(Salmon)

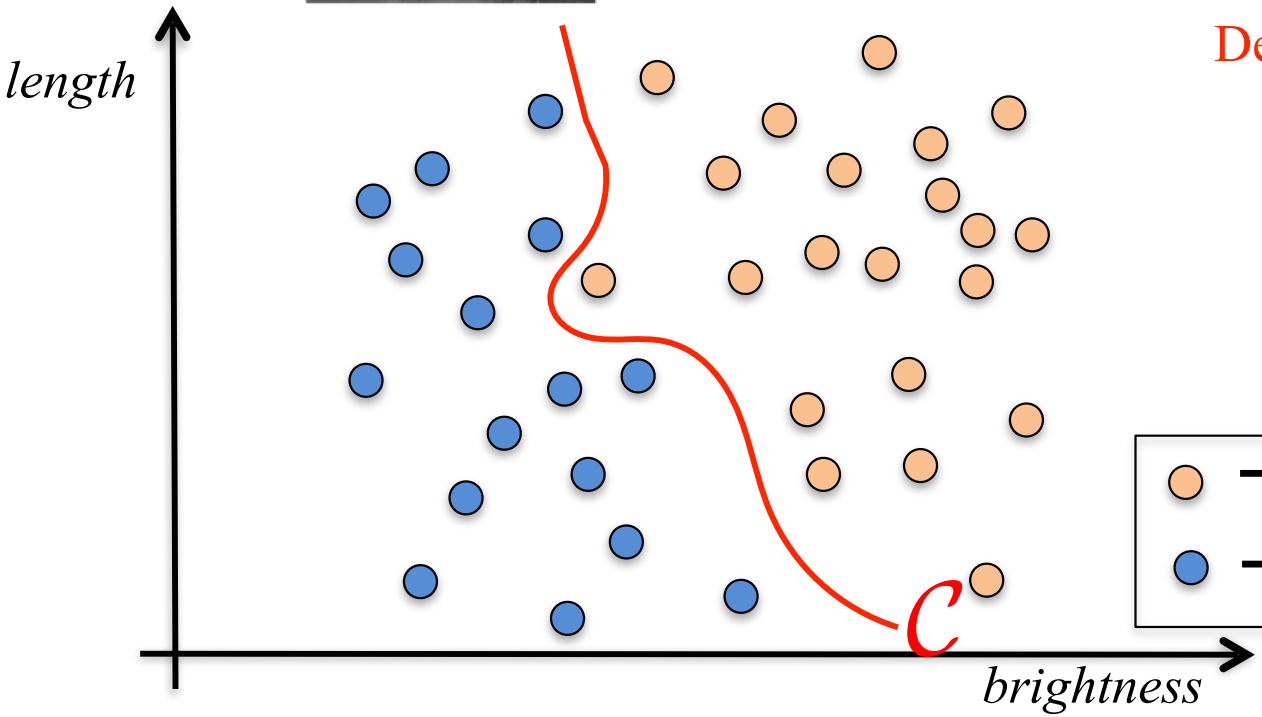
● → -1 (Sea Bass)

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

$$\begin{pmatrix} brightness \\ length \end{pmatrix}$$



Decision boundary:

$$\mathcal{C} = \{\mathbf{x} \in R^2, y(\mathbf{x}, \mathbf{w}) = 0\}$$

$$y(\mathbf{x}, \mathbf{w}) > 0 : Salmon$$

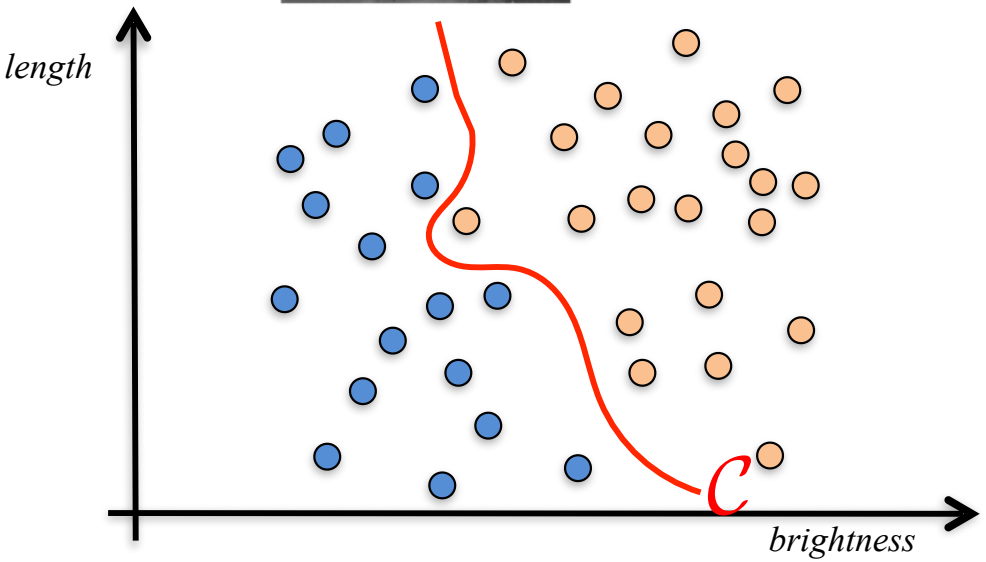
$$y(\mathbf{x}, \mathbf{w}) < 0 : Bass$$

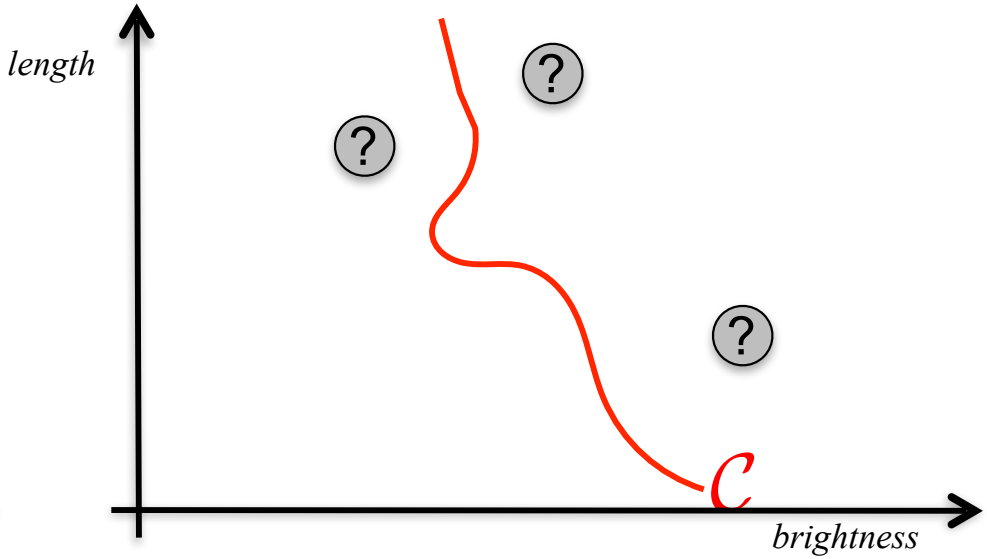
In this class, we will discuss:
- How to define y.
- How to choose **w**.

Training vs Testing

Some algorithm

$$\begin{pmatrix} brightness \\ length \end{pmatrix}$$





Training set = {●,●}

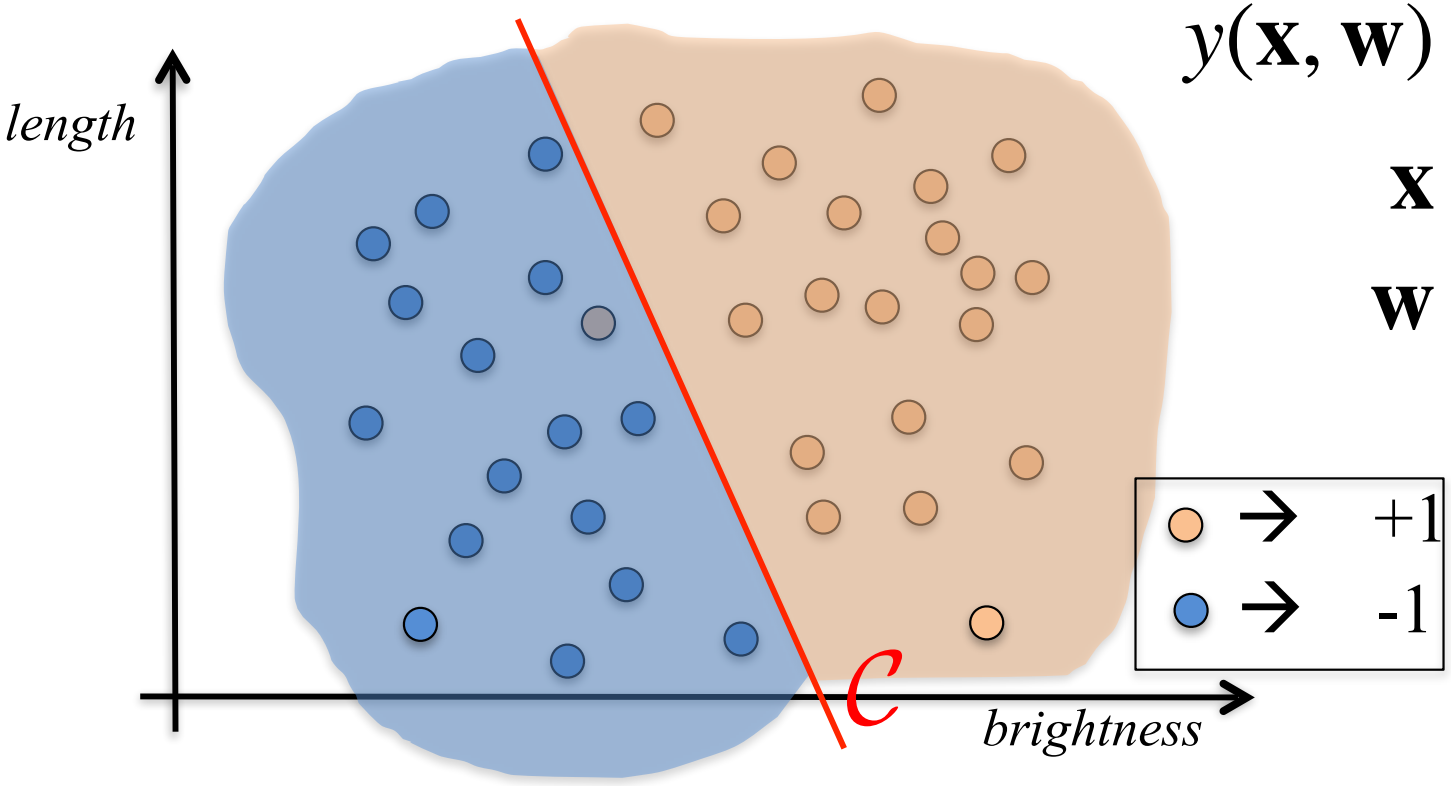
Test set = {?,?,?, ...}

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 →

$$\begin{pmatrix} brightness \\ length \end{pmatrix}$$



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

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

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

—> Important special case.

Course Outline

Introduction

- ML Basics
- K Nearest Neighbors
- K Means

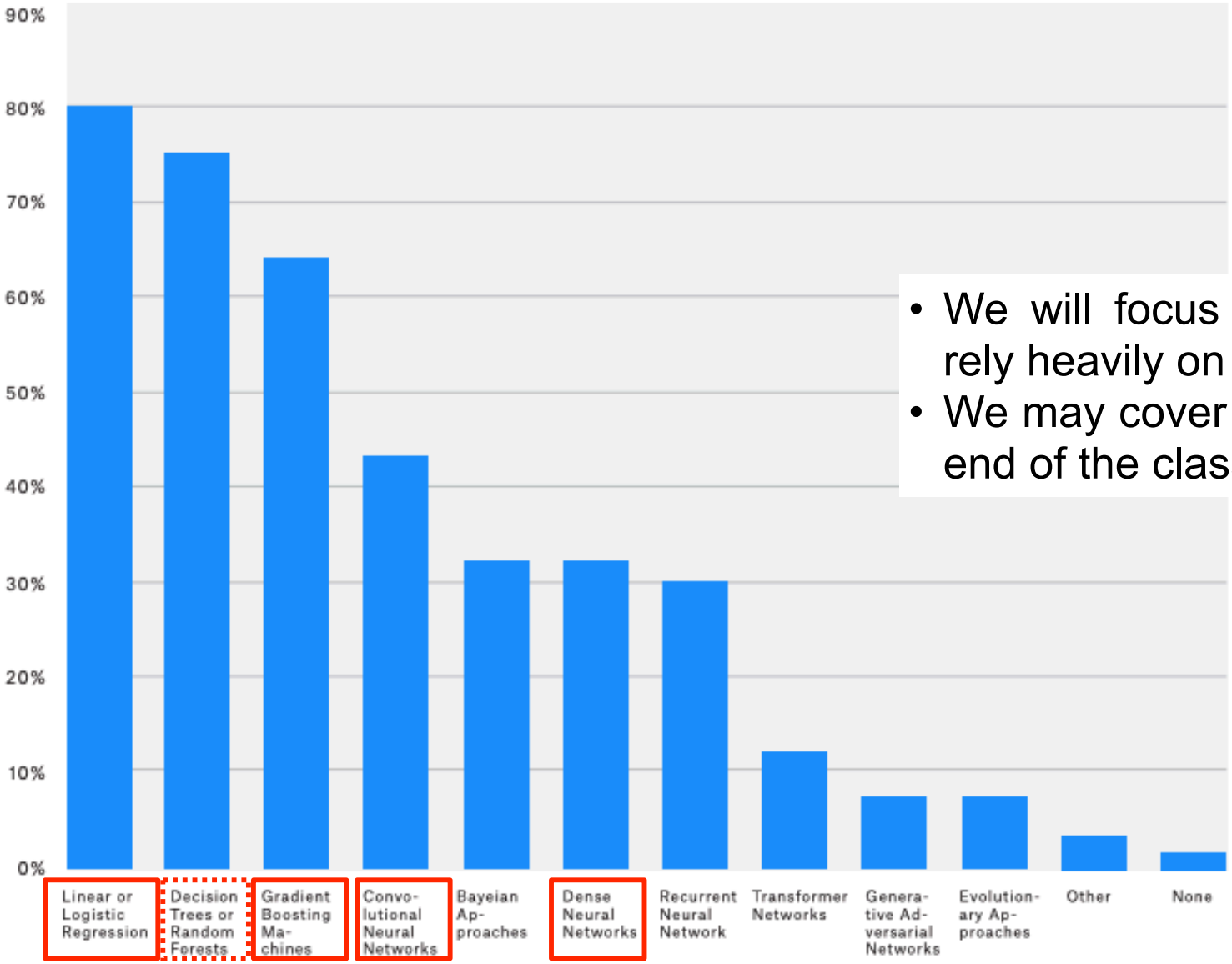
Linear ML

- Linear Regression
- Logistic Regression
- Max Margin Classifiers

Nonlinear ML

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

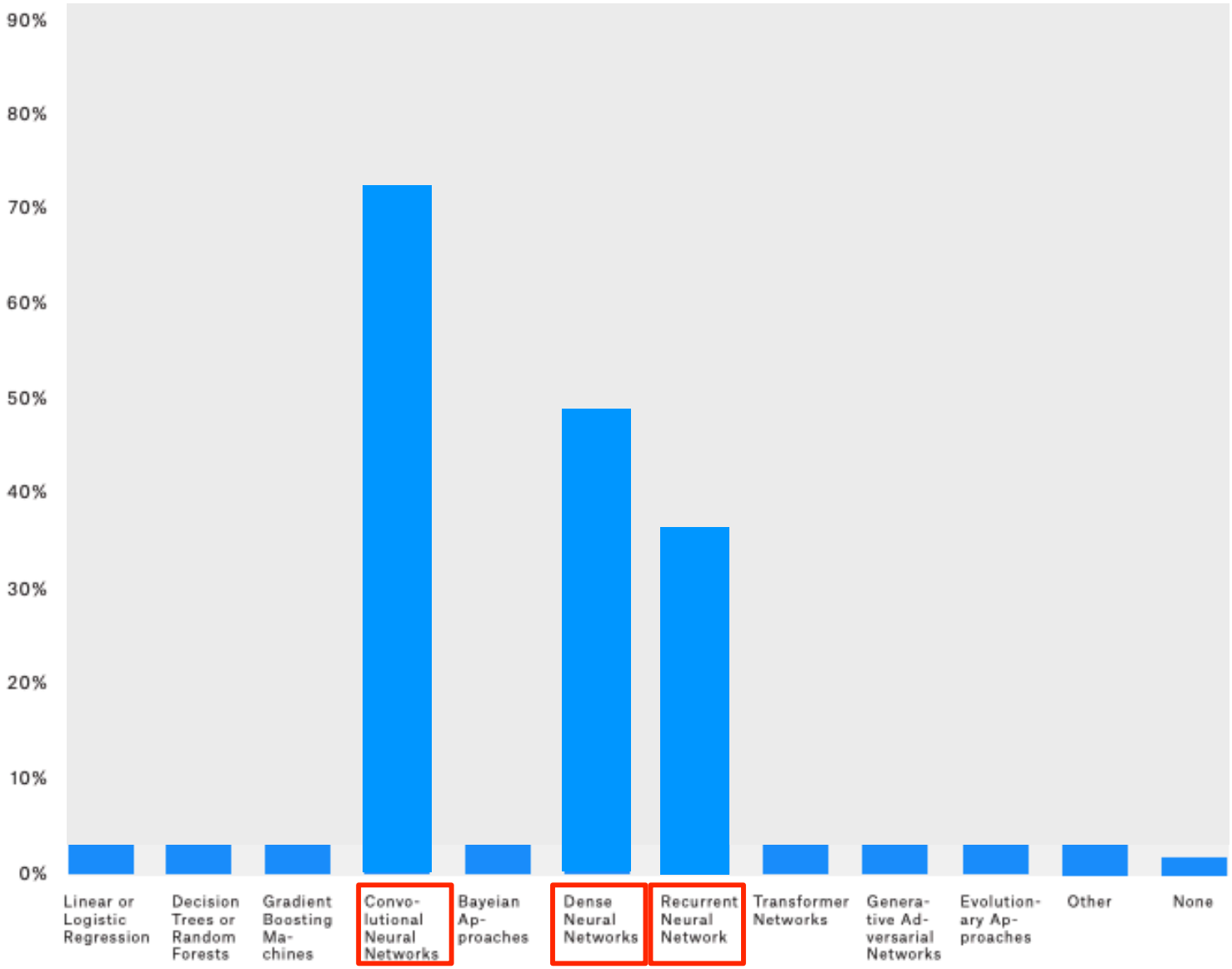
Kaggle Survey (2019)



- We will focus on methods that do not rely heavily on probabilities.
- We may cover some of the others at the end of the class time permitting.

What data science methods do you use at work?

My Erroneous Perception



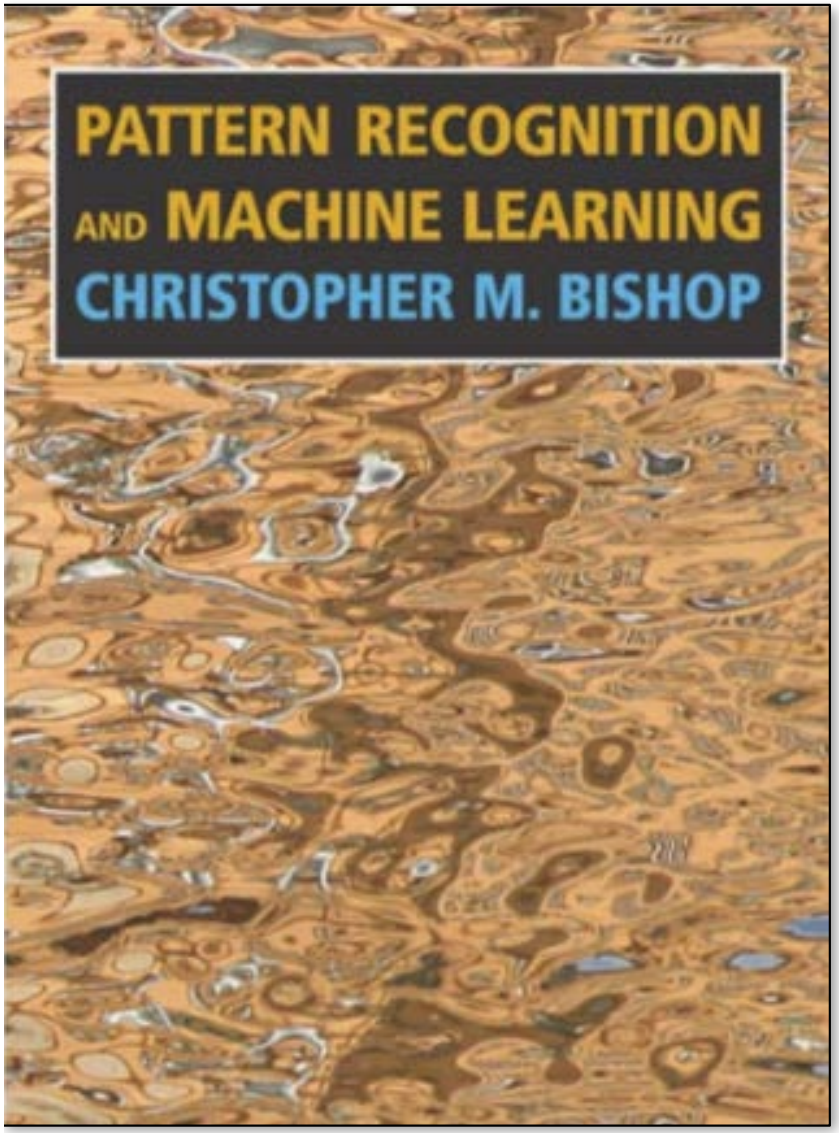
- Will it evolve in that direction?
- Time will tell.

General Organization

- Lectures: Tuesdays 8:15-10 online
- Exercises: Tuesdays 8:15-10 online.
- Written exam with one page of notes.

- 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

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.