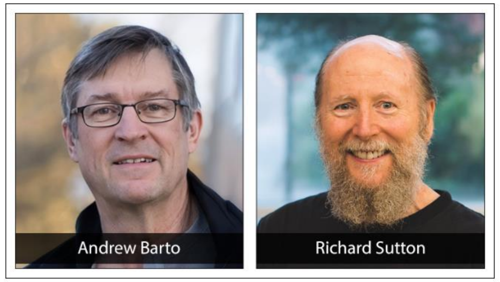
M1. Introduction to Pattern Recognition (PR) and Machine Learning (ML)

Manikandan Narayanan (Mani or MKN for short)

Week 1 (Jul 28-)

PRML Jul-Nov 2025 (Grads section – MTech/MS/PhD/Other students)

Now’s an exciting time to be studying PRML fundamentals!



AWARDS & RECOGNITION

Andrew Barto and Richard Sutton Receive 2024 ACM A.M. Turing Award

Andrew G. Barto and Richard S. Sutton received the 2024 ACM A.M. Turing Award for developing the conceptual and algorithmic foundations of reinforcement learning. In a series of papers beginning in the 1980s, Barto and Sutton introduced the main ideas, constructed the mathematical foundations, and developed important algorithms for reinforcement learning—one of the most important approaches for creating intelligent systems. Barto is Professor Emeritus of Information and Computer Sciences at the University of Massachusetts, Amherst. Sutton is a Professor of Computer Science at the University of Alberta, a Research Scientist at Keen Technologies, and a Fellow at Amii (Alberta Machine Intelligence Institute).

2018 ACM A.M. Turing Award

ACM presented the 2018 A.M. Turing Award to Yoshua Bengio, Geoffrey Hinton, and Yann LeCun at its annual Awards Banquet on June 15 in San Francisco, California.

For conceptual and engineering breakthroughs that have made deep neural networks a critical component of computing.

KUNGL. VETENSKAPS-AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

8 October 2024

The Royal Swedish Academy of Sciences has decided to award the Nobel Prize in Physics 2024 to

John J. Hopfield
Princeton University, NJ, USA

Geoffrey Hinton
University of Toronto, Canada

“for foundational discoveries and inventions that enable machine learning with artificial neural networks”

They trained artificial neural networks using physics

KUNGL. VETENSKAPS-AKADEMIEN
THE ROYAL SWEDISH ACADEMY OF SCIENCES

9 October 2024

The Royal Swedish Academy of Sciences has decided to award the Nobel Prize in Chemistry 2024

with one half to

David Baker
University of Washington, Seattle, WA, USA
Howard Hughes Medical Institute, USA

“for computational protein design”

and the other half jointly to

Demis Hassabis
Google DeepMind, London, UK

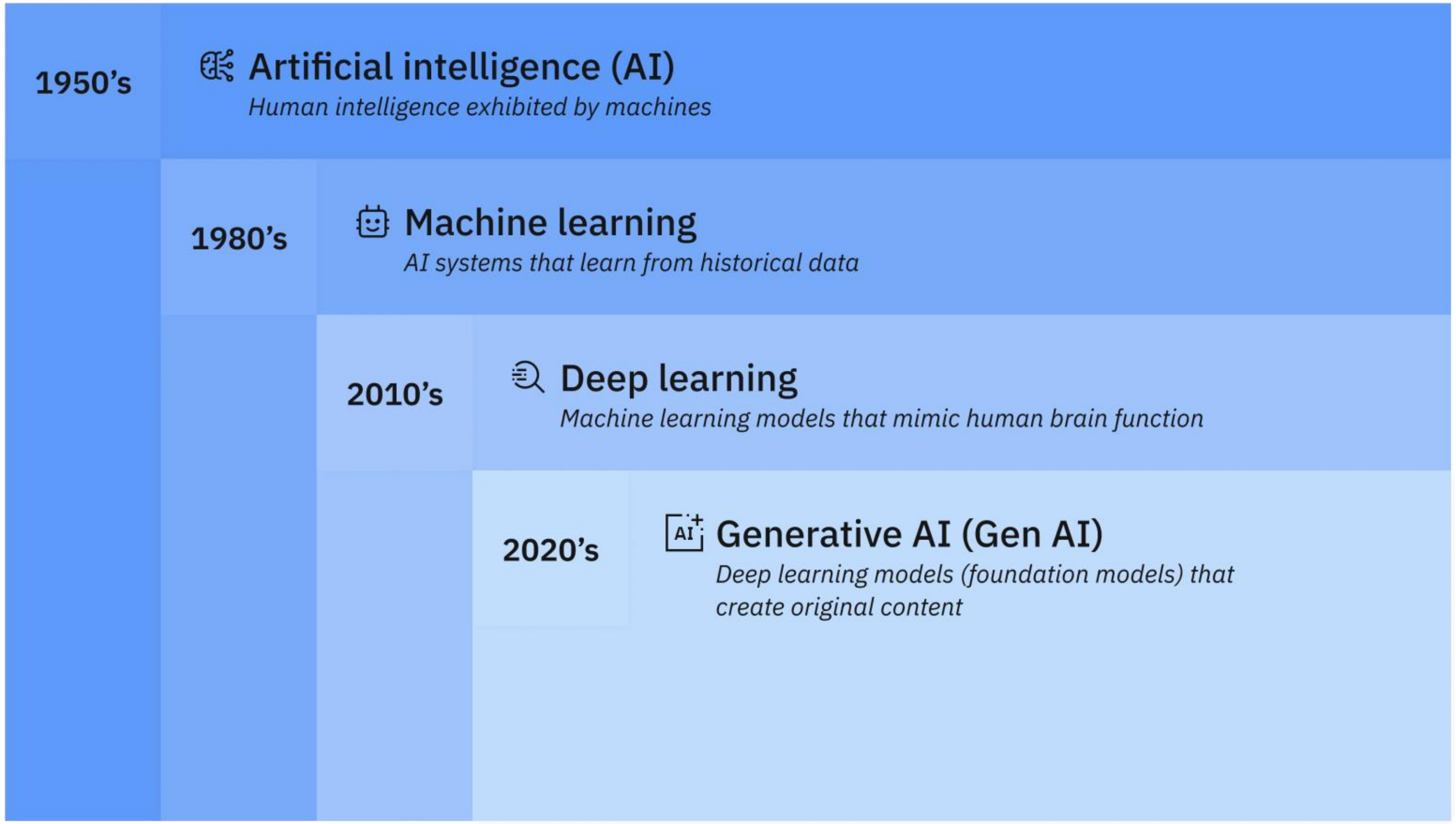
John Jumper
Google DeepMind, London, UK

“for protein structure prediction”

They cracked the code for proteins’ amazing structures

[Sources: https://awards.acm.org/ and https://www.nobelprize.org/]

AI history simplified (excluding winter periods)



[From https://www.ibm.com/think/topics/artificial-intelligence]

Current Spring of AI:

~2020-present

Winter of AI:

Two major "winters" during mid 1970s to ~1980, and late 1980s to early 1990s.

[https://en.wikipedia.org/wiki/AI_winter]

Focus on foundations in this course, so that you can cut across hype and get to the substance and foundational concepts of (PR)ML.

E.g., The concept of generative models have been around for a very long time (based on simpler ML models, but recently they are based on DL models).

Acknowledgment of Sources

- Slides based on content from related
  - Courses:
    - IITM – Profs. Arun/**Harish**/Chandra/Prashanth's PRML offerings (slides, quizzes, notes, etc.), Prof. Ravi's "Intro to ML" slides – cited (e.g., **[HR]/[HG]**) in the bottom right of a slide.
    - India – NPTEL PR course by IISc Prof. PS. Sastry (slides, etc.) – cited as [PSS] in the bottom right of a slide.
  - Books (and related material):
    - PRML by **Bishop**. (content, figures, slides, etc.) – cited as **[CMB]**
    - Pattern Classification by Duda, Hart and Stork. (content, figures, etc.) – [DHS]
    - Mathematics for ML by Deisenroth, Faisal and Ong. (content, figures, etc.) – [DFO]
    - Foundations of ML by Mohri, Rostamizadeh, and Talwalkar (content, figures, slides by Mohri, etc.). – **[MRT]**

Outline for Module M1

- M1. Introduction to PRML
  - **M1.0 The What?**
    - **What is a PR task/problem?**
    - **What is a ML algorithm?**
  - M1.1 The Why???
    - Your motivation
    - Our motivation (course syllabus & objectives)
    - Do they align? (other course logistics)
  - M1.2 ML Paradigms/Terms/Examples/Next-steps

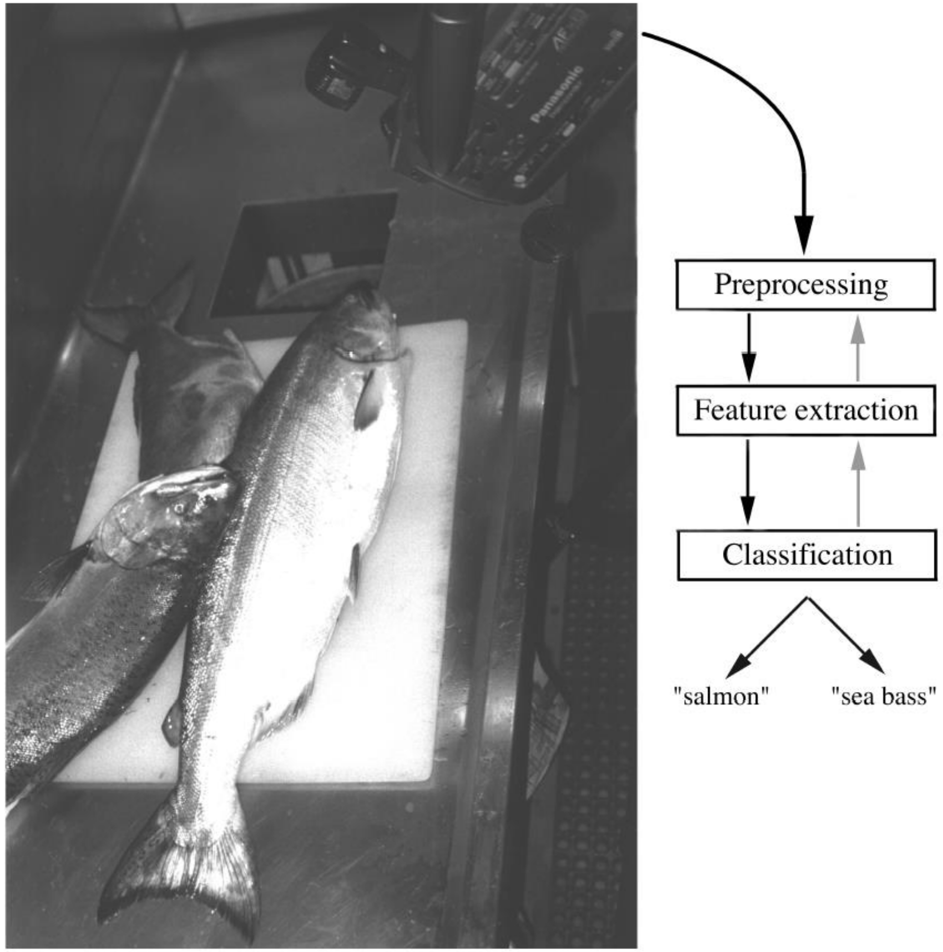
What is a (real-world) PR task?

- Humans routinely categorize sensory inputs (i.e., recognize patterns in sensory inputs).
  - Read facial expressions
  - Recognize speech
  - Read a document
  - Diagnose disease from medical image

- Build a machine (system) to recognize
  - Recognize a person by fingerprints
  - Recognize different types of fruits from an image
  - Predict if a particular region would suffer from a COVID outbreak in the next month.
  - Find relevant movies for you in Netflix based on movies you've watched.

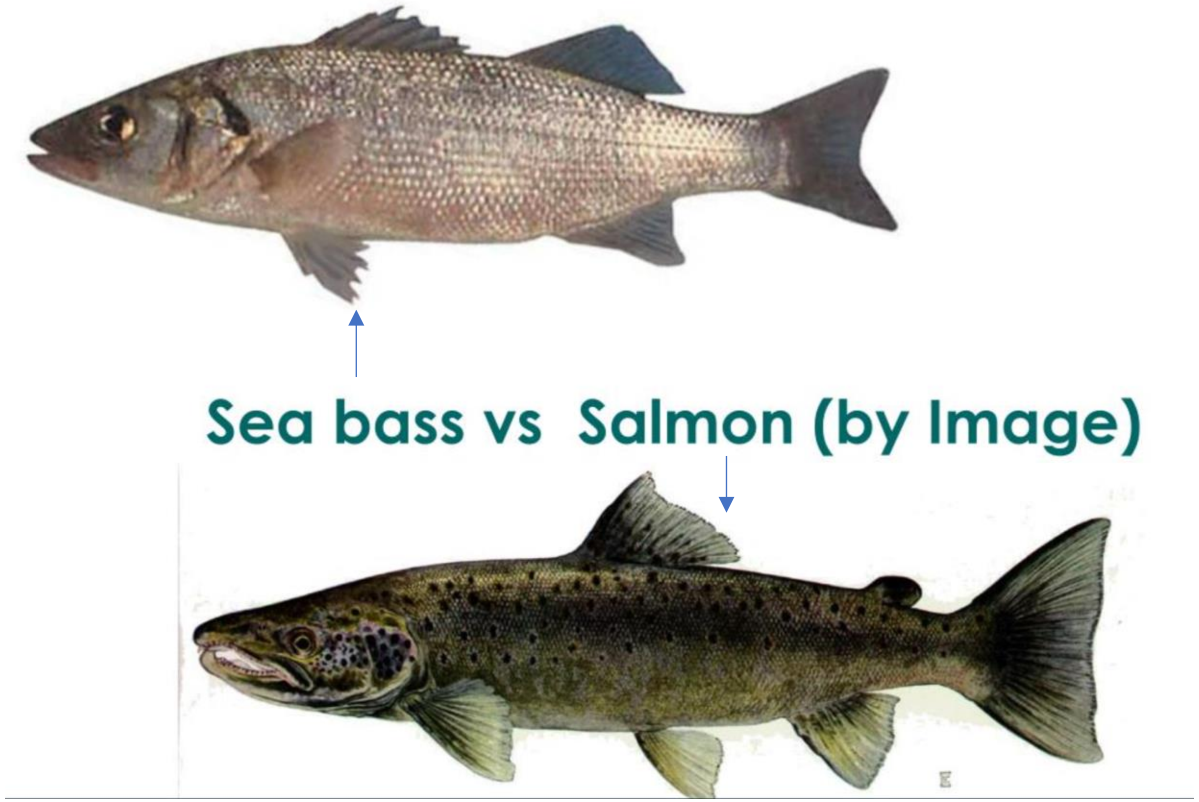
[PSS,AR]

How does a system for a PR task look like?



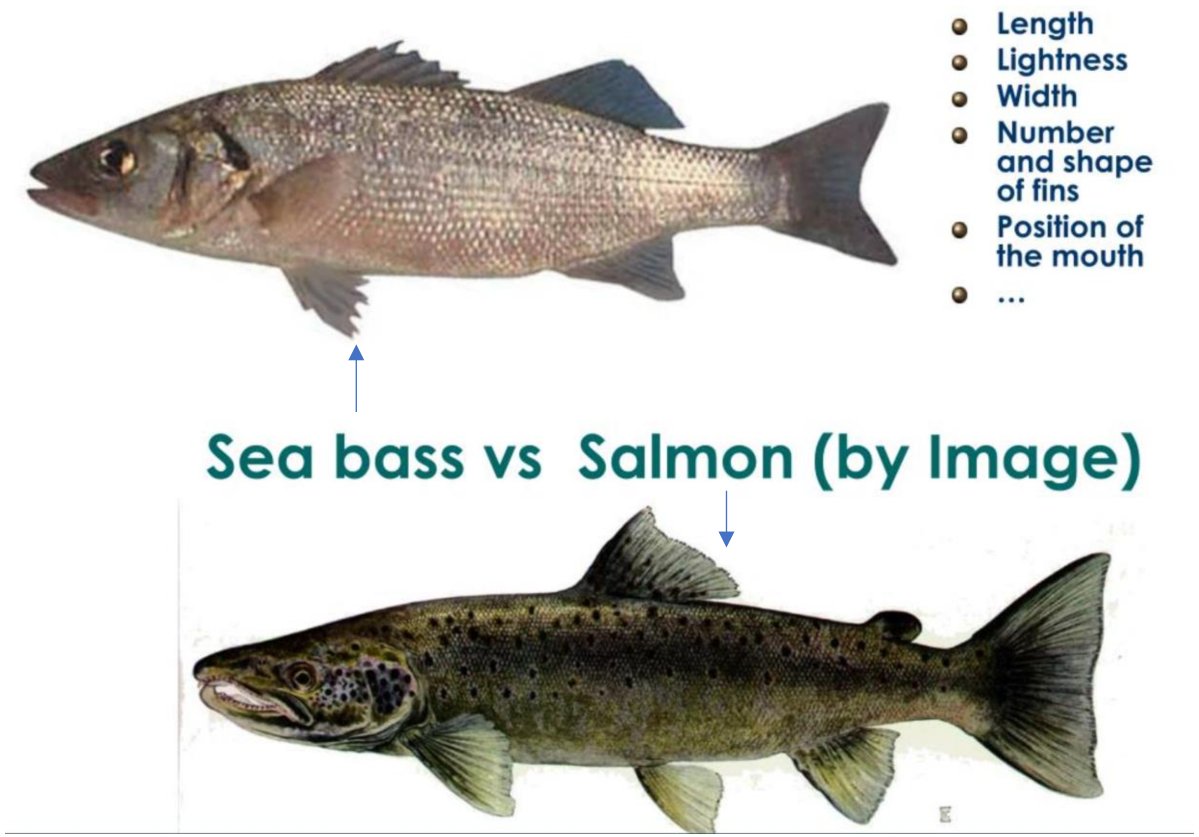
[DHS]

What features would you use?



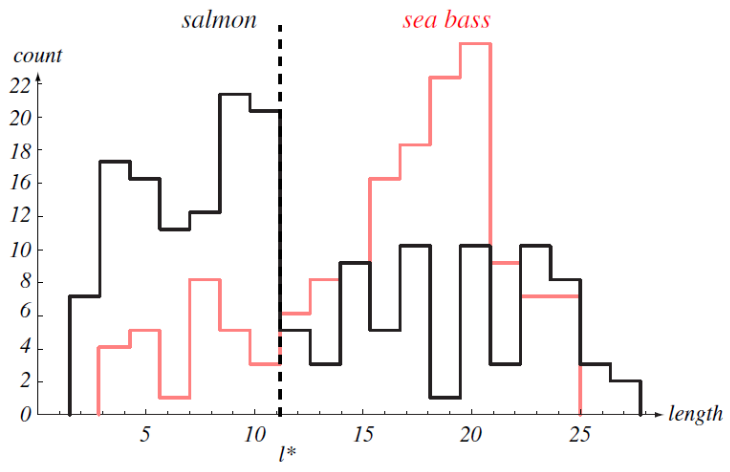
From https://www.slideserve.com/jakeem-barlow/introduction-to-pattern-recognition

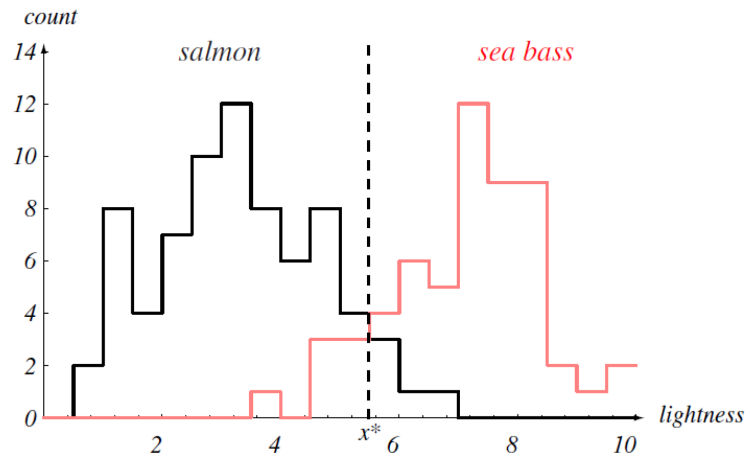
What features would you use?



From https://www.slideserve.com/jakeem-barlow/introduction-to-pattern-recognition

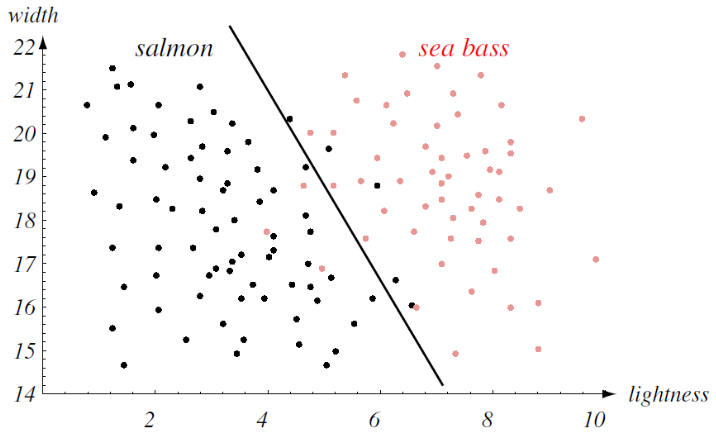
Feature extraction/selection

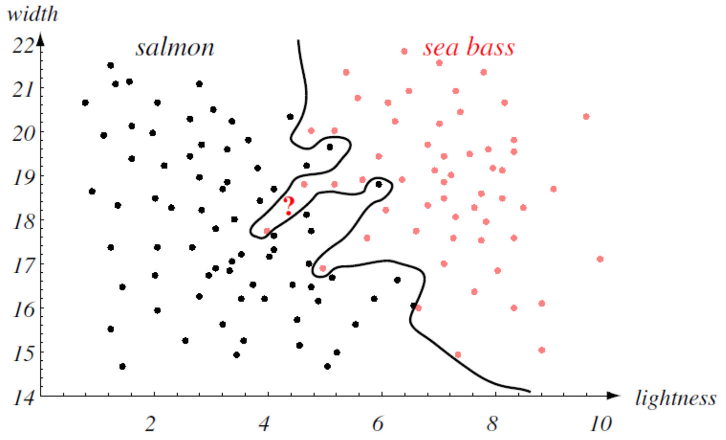


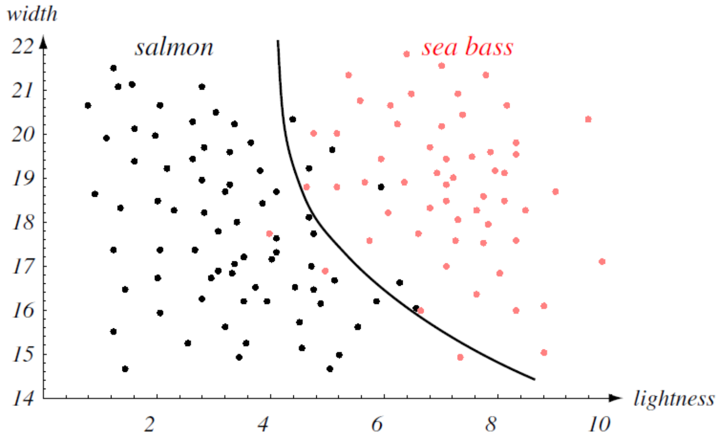


[DHS]

Decision boundary – from simple to complex

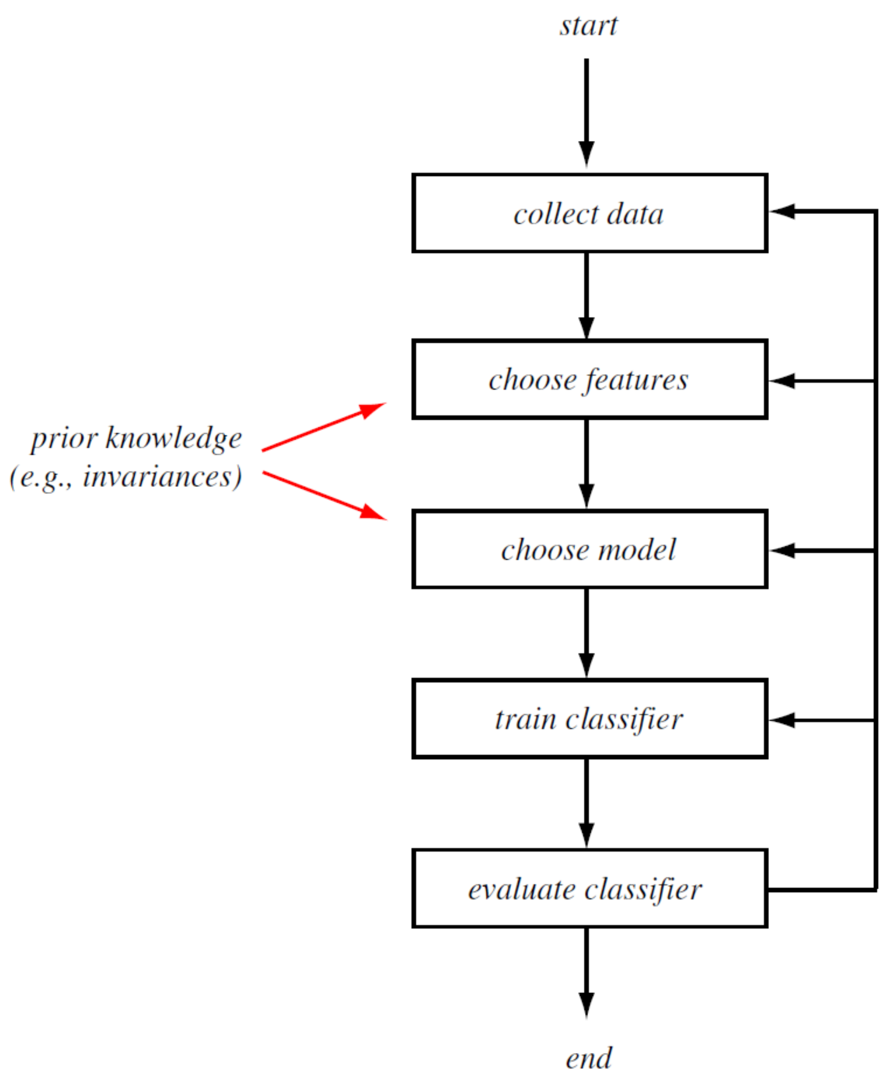






[DHS]

The Design Cycle



[DHS]

What is a PR problem?

- Formulation of the PR task as an abstract computational problem with given input (features) and desired outputs/actions.

- Major classes of PR problems:
  - Clustering
  - Classification
  - Regression
  - …

- How do we solve a PR problem? Typical AI-based approaches include:
  - Logic/Rule/Knowledge-driven approaches
    - (logic/rules typically crafted by humans based primarily on domain knowledge (and to a lesser extent on data); algorithms that do simple/complex reasoning using these rules)
  - Data-driven ML approaches
    - (algorithms focused in this course and defined next)

What is a ML algorithm?

- An algorithm is
  - "… said to learn from experience with respect to some class of tasks, and a performance measure P, if [the learner's] performance at tasks in the class, as measured by P, improves with experience.". [Tom Mitchell 1997]

- In our context of solving a PR task, an ML algorithm recognizes a pattern by:
  - learning it from examples (*training* data provided by a teacher/oracle),
  - so that it can *generalize* to *unseen* (*test*) data.

Clarifying the defn. (PR problem)

- Identify a given image as mango or orange.

[HR]

Clarifying the defn. (ML approach)

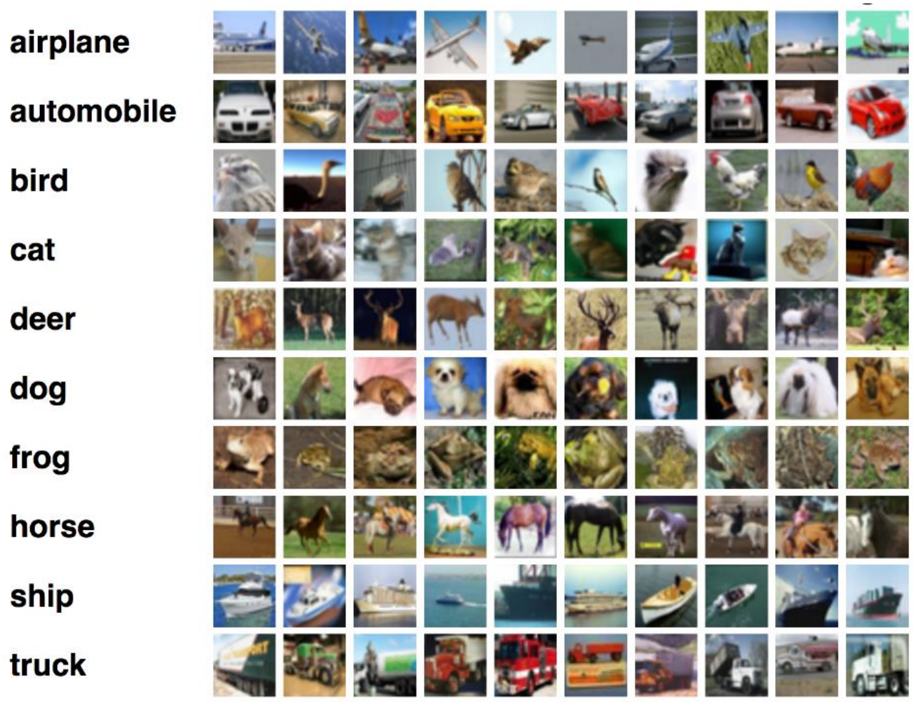
- Identify a given image as mango or orange, given many example images of mango and orange.

M M O M O

[HR]

Clarifying the defn. (another PR problem/ML approach)

Classify an image into one of 10 classes,
given a "training set" of images with classes.



[HR]

Yet another problem!

Check bbcnews article (car vs. van classifier example):
https://www.bbc.co.uk/news/resources/idt-74697280-e684-43c5-a782-29e9d11fecf3

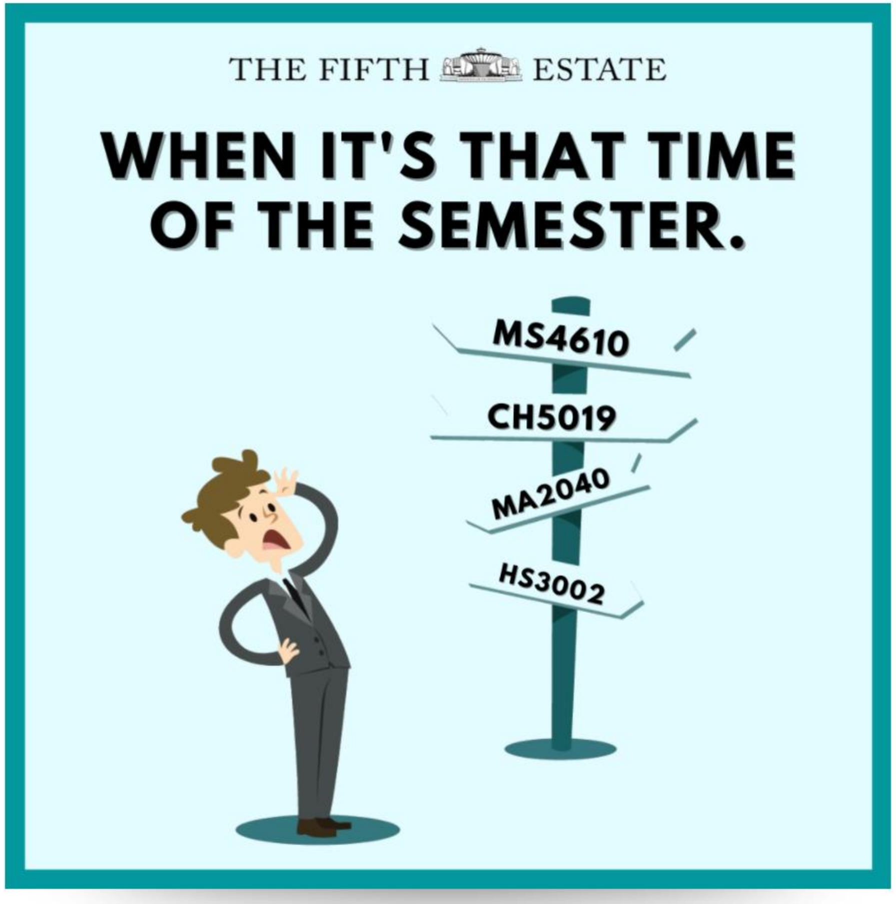
Outline for Module M1

- M1. Introduction to PRML
  - M1.0 The What?
    - What is a PR task/problem?
    - What is a ML algorithm?
  - **M1.1 The Why???**
    - **Your motivation**
    - **Our motivation (course syllabus & objectives)**
    - **Do they align? (other course logistics)**
  - M1.2 ML Paradigms/Terms/Examples/Next-steps

When it’s that time of the Semester

AUGUST 3, 2020 ▪ JAI SANTHOSHI ▪ LEAVE A COMMENT

FEATURE INSTI NEWS



Design: Hardhik Pinjala

Editor: Siddharth D P

A general viewpoint, but what is yours?...

CS Courses:

A course that every Machine Learning enthusiast should look into is the **CS5691, Pattern Recognition and Machine Learning**. Probability is an advised prerequisite for this course and it is quite a hefty course. Nevertheless, it delves into the mathematical fundamentals behind the Machine Learning Algorithm and is a resourceful course when it comes to projects on Data Science and Machine Learning.

Course Code: CS5691

Course Name: Pattern Recognition and Machine Learning

Credit Count: 15

Perks: Internships and Placements

Get to know you

Are you an

a) Astute listener,
b) Experienced artist, or
c) Fledgling composer

of ML?

(see [DFO] for more on above)

Get to know the course – planned syllabus

Subset of topics below to be covered (not necessarily in the same order):

1. Introduction to PR/ML approach/system (PR task/problem, ML algorithm, ML paradigms/scenarios)
2. Decision theory (optimal (Bayes) classifier, loss functions, optimal regressor)
3. Density estimation (Maximum likelihood, Bayesian estimation, Brief mention of additional topics (Expectation Maximization (EM) for mixture density estimation, and Non-parametric methods))
4. **Linear models for classification and regression** (Naïve Bayes classifier, Logistic regression, and Brief mention of discriminant analysis (hyperplanes); Linear/polynomial/regularized/Bayesian regression)
5. **Non-linear models for classification and regression** (Support Vector Machines and Kernel methods, Neural networks)
6. **Combining models** (Ensemble methods like **Boosting and Bagging, Tree-based** models)
7. Unsupervised learning methods for **clustering and dimensionality reduction** (E.g., hard/soft k-means clustering, Principal Component Analysis (PCA))
8. Pointers to (brief mention of) advanced topics (E.g., Causal ML, Computational learning theory, Structured data algorithms like ones for sequential data (Hidden Markov Models HMM) or graph-structured data; Probabilistic graphical models).

Course objectives, and Expected course outcome

KEY Objectives:

.. Clearly understand the theoretical/statistical principles underpinning different ML algorithms (in particular), and what it means to learn from data (in general).

.. Implement/extend ML algorithms from first principles, tune associated parameters, and evaluate/interpret the results.

Implicit Objectives:

.. Appreciate the different paradigms of ML including supervised and unsupervised learning, and use them to unify my conceptual view of several existing ML algorithms.

.. Identify the right ML algorithm for a given PR problem, by understanding the key differences/similarities and pros/cons of various algorithms.

.. For a real-world task involving data, either solve it using a relevant PR/ML problem/algorithm (or) explain the unsuitability/limitations of a PR/ML approach to the task.

Not focused:

.. Identify the right software package and hardware environment for efficiently running a ML algorithm to solve a PR problem.

.. Very large data contests, which require implementation of a highly efficient ML pipeline.

A big question is:

WHY???

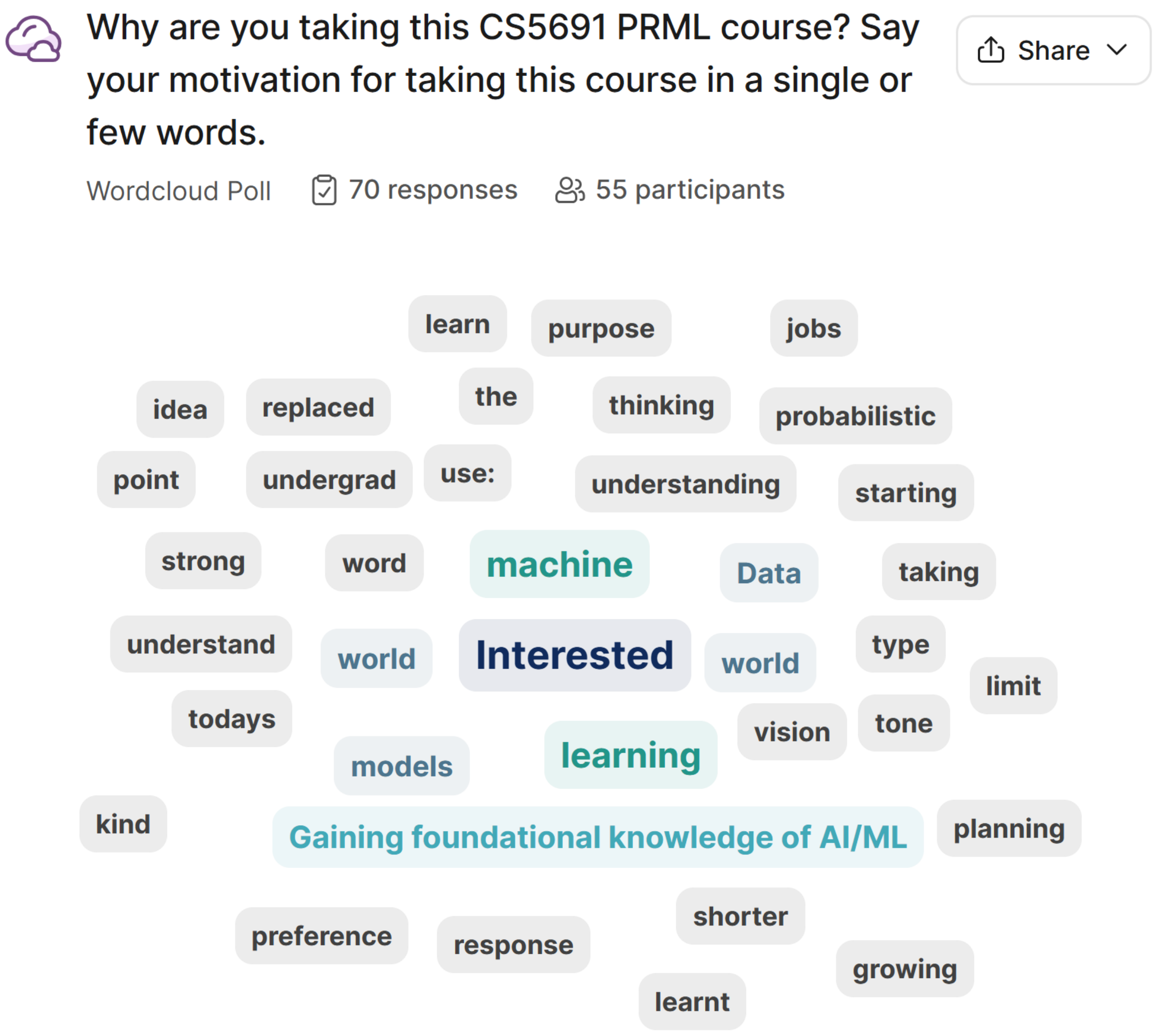
learn so many models/methods, if only one of the methods is all you hear about everywhere?

On a different context, always ask WHY you are taking this course?

<course logistics begin>

Slido poll break (5 mins)

Slido poll results



A big question is:

WHY???

learn so many models/methods, if only one of the methods is all you hear about everywhere?

On a different context, always ask WHY you are taking this course? *I hope it is not only campus placement opportunities or hype around ML, but also a general interest in understanding how you can take this field forward with your own creativity and internal ethical compass.*

Get to know you – informal (raise-of-hands) poll

- Level of comfort with prereqs.:

	Moderate	Poor	Very	Action
Probability	..	..	..	(Please self-review axiomatic defns. of probab.)
Linear algebra	..	..	..	(many taking LARP in parallel; will cover LA-heavy ML topics in 2nd half of course)
Calculus	..	..	..	
Programming	..	..	..	(Python tutorial around Assignment 1 if students ask)

Get to know me

- First exposure to ML:
  - Probabilistic Graphical Models (PGM) course by Michael Jordan in UC Berkeley.
  - Offered PGM course here at IITM in different semesters, and PRML course last semester.
  - Bishop’s book was written during a rising wave of interest in PGM.

- Interdisciplinary researcher:
  - Bioinformatics (interface of computer science and biology)
  - Computer science: probabilistic graphical models (causal inference/discovery) and graph algorithms
  - Biology: tissue-tissue communication, disease-disease interactions

Get to know the TAs

- Many of them have taken and/or TAed the course before, and do research in AI/ML; so in a good position to help you. Reach out to them and seek their help!
  - Eight TAs (they introduced themselves in class) in program/alphabetic order:
    Ayushman Bhatt, Debendra Kumar Pal
    Abhishek Ranjan, Rupankar Podder, Sri Saravanan R
    Nency Bansal, Nilesh Subramanian, Rekha Raj
  - ~10-15 students per TA; open hours and tutorial sessions; help with learning topics, clarifying doubts, and preparing for quizzes/exams

- To contact us: use ***Ed Discussion*** (https://edstem.org/us/courses/80982/discussion)
  - Ed Discussion common forums – for any common qns. (on assignments, quizzes, etc.)
  - Ed Discussion "Private" to Staff option – for any specific qns. (visible only to you and TAs/instructor)
  - Your assigned TA or your instructor (email IDs of TAs and instructor in "Course handout" posted in Ed Discussion)

Evaluation scheme (tentative)

- Aka “Help us help you learn”

- Tutorial/Class Participation – 3%
- 3 Assignments - 33% (roughly once every month)
- Quiz I and II - 22% (same Quiz slots as announced in IITM calendar)
- Endsem Exam - 42% (same Endsem slot as in IITM calendar; divided into 3 parts to enable student-friendly testing as explained in “Course handout”)

- **Dates** and other details for all the above course activities are in:
  “Course handout” in **Ed Discussion**.

<course logistics end>

Outline for Module M1

- M1. Introduction to PRML
  - M1.0 The What?
    - What is a PR task/problem?
    - What is a ML algorithm?
  - M1.1 The Why???
    - Your motivation
    - Our motivation (course syllabus & objectives)
    - Do they align? (other course logistics)
  - **M1.2 ML Paradigms/Terms/Examples/Next-steps**

ML Paradigms

- Supervised Learning (informally aka curve-fitting or function learning/approxn.)
  - Learn an input and output map (features to target(s))
    - Classification: categorical output
    - Regression: continuous output
- Unsupervised Learning (informally aka pattern discovery)
  - Discover/uncover patterns in the data (without target/response variable(s))
    - Clustering: cohesive grouping
    - Dimensionality reduction: represent features in low-dimensional space
    - Density estimation: learning model parameters
    - Association: frequent co-occurrence
- Reinforcement Learning
  - Learning control (maximize reward in the long run, via optimal explore-exploit policy)
    - Agent interacting with an environment iteratively via a set of "controlling" actions, and getting information about environment's state/reward to decide next action.
- Other paradigms/categorizations
  - Other paradigms: semi-supervised, online, etc.
  - Other categorizations: linear vs. non-linear models, single vs. ensemble models, methods for independent vs. time-dependent vs. graph-structured data, etc.

[BR,HR]

Review of some terms

- **Example/Point/Datapoint/Sample**: item, instance of the data used.
  - **Features**: attributes associated to an item, often represented as a vector (e.g., length/lightness of a fish image, word counts of a document).
  - **Label**: category (classification) or real value (regression) associated to an item.
- **Parameters:** Parameters that are learnt (estimated) directly from data by the ML algorithm during the training phase.
    - E.g., coefficients in a linear/logistic regression model; weights and biases in a neural network
- **Hyperparameters:** Free parameters not determined by the ML algorithm, but specified as additional inputs to (control) the ML algorithm
    - E.g., learning rate of the gradient descent optimization algo., for ML algos. that use gradient descent to optimize the training phase's objective function
- **Data (consisting of many samples):**
  - training data (typically labeled, for learning parameters).
  - test data (labeled but labels not seen).
  - validation data (labeled, for tuning hyperparameters).

[MRT]

Revisiting ML paradigms/scenarios using these terms

- **Learning stages:**
  - Preprocessing and Feature Selection and Model Selection (based on prior knowledge or model invariances)
  - Training (learning model parameters) and Validation (tuning hyperparameters)
  - Testing (evaluation)
- **Standard (Batch) Scenarios** (with **separate** training and testing stages)
  - **Unsupervised** learning: uses unlabelled data for prediction on unseen points (clustering, dim. redn., density estimation – difficult to evaluate performance due to lack of human-annotated “true” labels, however alternate performance metrics exist)
  - **Supervised** learning: uses labelled data for prediction on unseen points (regression, classification, etc.)
  - **Semi-supervised** learning: uses labeled and unlabeled data for prediction on unseen points.
  - …
- **Other Scenarios** (with rounds of **intermixed** training and testing stages)
  - **Online learning** – in one round, get one training example, predict its label, receive its true label, and incur loss; repeat multiple rounds and minimize cumulative loss (or regret).
  - **Reinforcement learning** – same as above, but interact with or affect environment and receive instant reward for each action; maximize cumulative reward; explore vs. exploit dilemma
  - **Active learning** – actively choose what new training examples to get (adaptively or interactively from oracle)
  - …

[MRT]

A note on real-world applications

- Complex paradigms/scenarios come up in practical applications:
  - E.g. scenarios: all combinations of (batch vs. online) x (active vs. passive) x (supervised vs. unsupervised) learning
  - E.g. tasks: online (passive) regression, (minibatch) active classification, online clustering, etc.
- In this course, we focus on simple scenarios (mostly standard batch scenarios),
  - so that we can delve deeper into the fundamentals of ML algorithms for these scenarios;
  - and you can later try out combinations based on this understanding.

Examples of PR Tasks/Problems in different real-world application domains

- Text: document classification, spam detection.
- Language: NLP tasks (e.g., morphological analysis, POS tagging, context-free parsing, dependency parsing), language modelling (e.g., LLMs).
- Speech: recognition, synthesis, verification.
- Image (computer vision): annotation, face recognition, OCR, handwriting recognition, image understanding, etc.
- Bioinformatics / Computational Biology: protein structure prediction, protein-DNA binding strength prediction, disease/medical diagnosis, function prediction, etc.
- Games (e.g., chess, backgammon, go).
- Unassisted control of vehicles (robots, car).
- Fraud detection, network intrusion, etc.

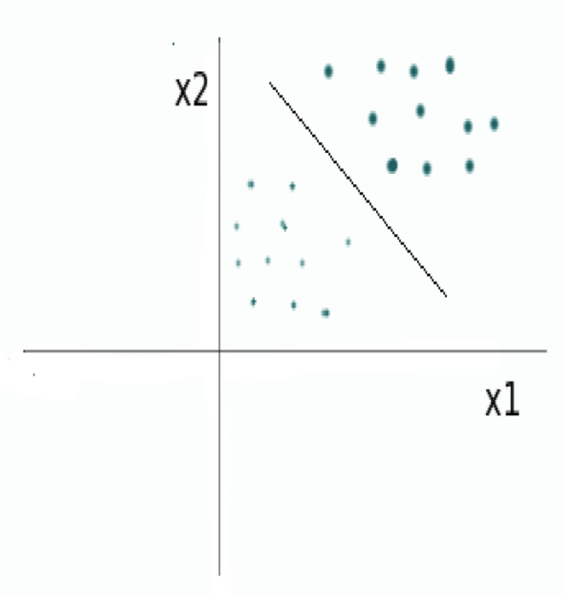
[MRT]

Examples (contd.): Example regression problem from bioinformatics

- Input: a protein sequence and a DNA sequence
  - ($\mathrm{x} \in \mathbb{R}^p$, with the number of features p depending on the length of the two sequences, and the alphabet size of these sequences)
- Output: strength/affinity of binding of the protein to the DNA
  - ($y \in \mathbb{R}$, determined using wet-lab experiments)
- Problem: Find a function that approximates the output of different input instances in the training data; i.e., find a function f s.t.
  - $\hat{y} = f(\mathrm{x})$ **"best" approximates** the true value y of x (denoted as $\hat{y} \simeq y$), across all examples $\{(\mathrm{x}_1, y_1), (\mathrm{x}_2, y_2), \dots, (\mathrm{x}_n, y_n)\}$ in the training data.

Example (linear) classification problem

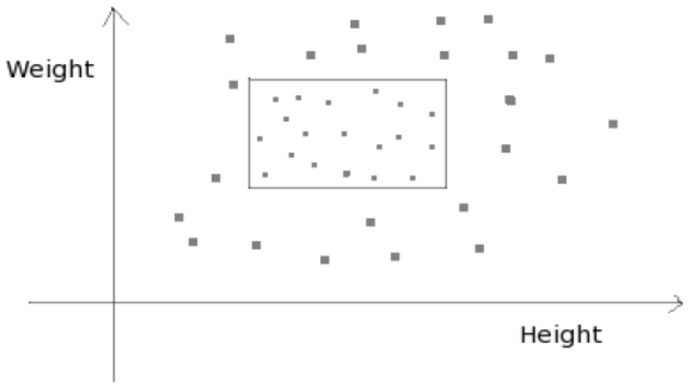
- : Problem: ‘Spot the Right Candidate’
- : Features:
  - x_1: Marks based on academic record
  - x_2: Marks in the interview
- A Classifier: $ax_1 + bx_2 > c \Rightarrow$ ‘Good’
  We have chosen a specific form for the classifier.
- Design of classifier: What values to use for a, b, c?
- **Information available: ‘experience’ – history of past candidates**



[PSS]

Example (non-linear) classification problem

- Problem: recognize persons of ‘medium build’
- Features: Height and Weight

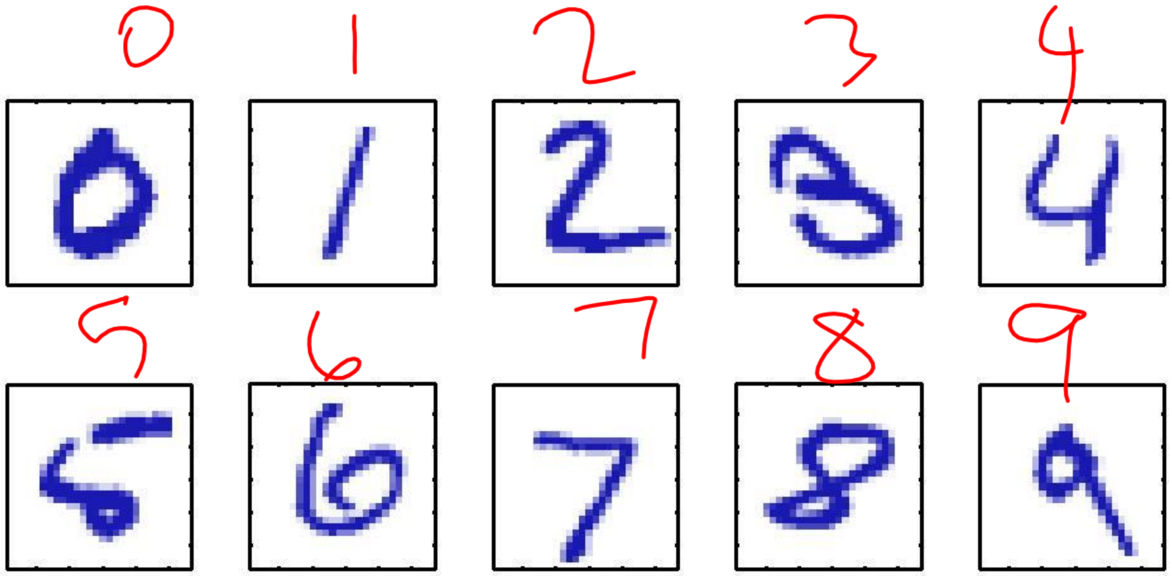


The classifier is ‘nonlinear’ here

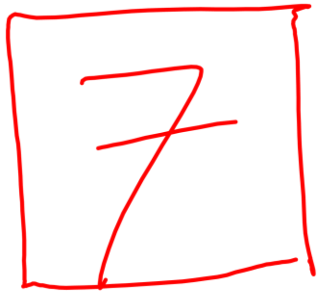
[PSS]

Example classification problem

- Handwritten Digit Recognition:
  - PR Problem: Take a digit's 28 x 28 pixel image, represented by a vector of 784 real numbers, as input feature vector **x**, and output the identify of the digit 0,…,9.
  - Non-trivial problem due to the wide variability of handwriting.
  - ML Algorithm: Learns this variability from training examples, and uses it to predict the digit of unseen images accurately.



Tr.



?

→ 7

Pred

Te

[CMB]

So far: PR system, and what next?

- PR task and problem
- ML algorithm (“computational method that uses experience to improve performance” [MRT])
- ML scenarios/paradigms

- What next?
  - Probability Review
  - Decision Theory (incl. optimal (Bayes) classifier and optimal regressor)

[DHS]

Thank you…

…for your interest in the course and your attention!

All the best to find your optimal learning path!

Backup slides

Jul 2023 offering results