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

B. Ravindran

August 3, 2021

PRML Aug-Nov 2021 (BR section)

Acknowledgment of Sources

- Slides based on content from related
  - Courses:
    - IITM – Profs. Arun/Harish/Manikandan/Chandra's PRML offerings (slides, quizzes, notes, etc.).
    - India – NPTEL PR course by IISc Prof. PS. Sastry (slides, etc.)
  - Books:
    - PRML by Bishop. (content, figures, slides, etc.) – cited as [CMB]
    - Pattern Classification by Duda, Hart and Stork. (content, figures, etc.) – [DHS]
    - Elements of Statistical Learning by Hastie, Tibshirani, Friedman (content, figures, etc.) – [ESL]
    - Mathematics for ML by Deisenroth, Faisal and Ong. (content, figures, etc.) – [DFO]

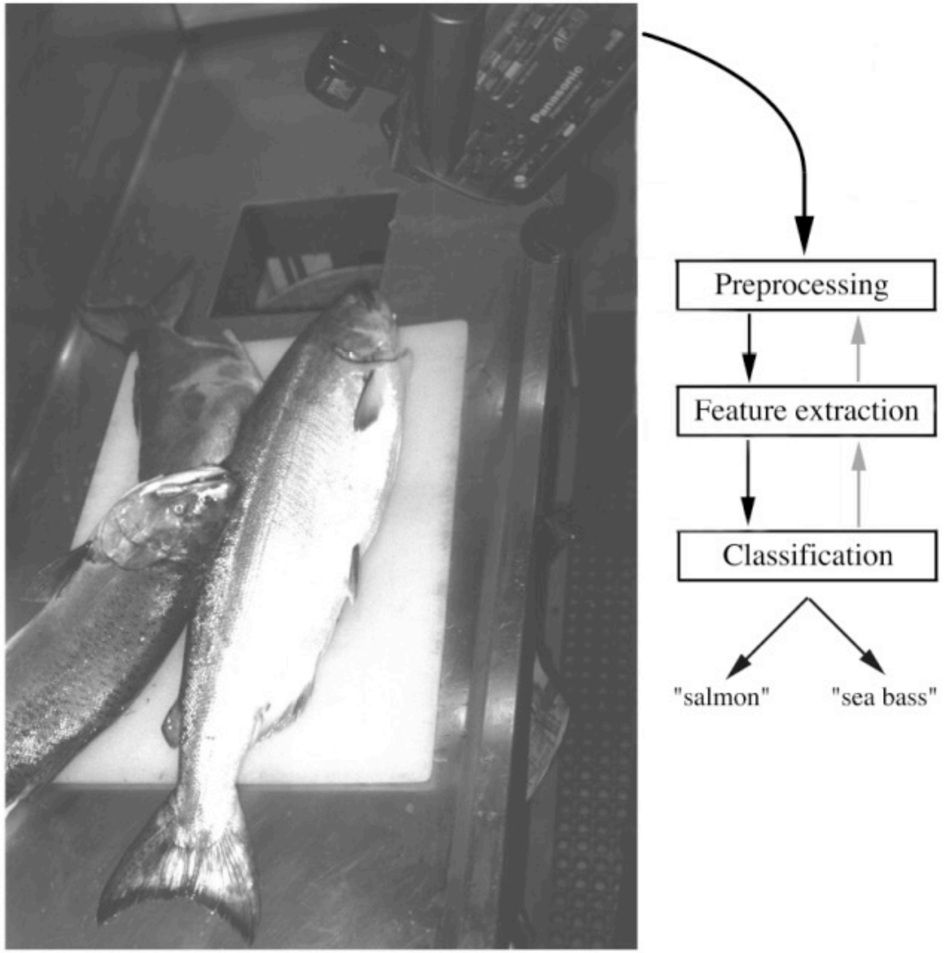
What is a (real-world) PR task?

- 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.

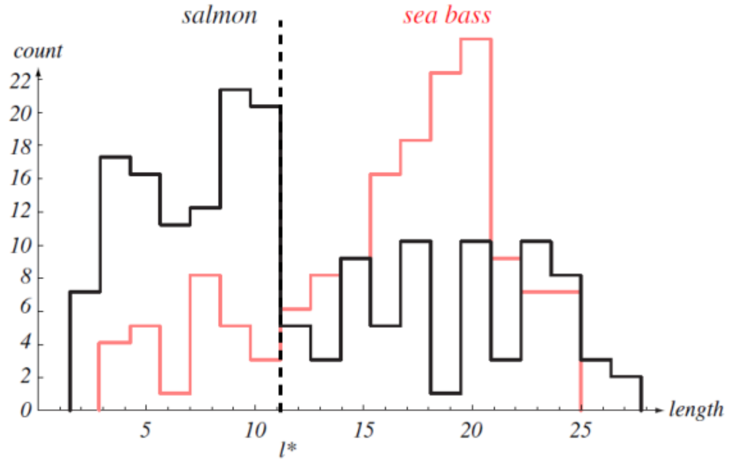
[PSS,AR]

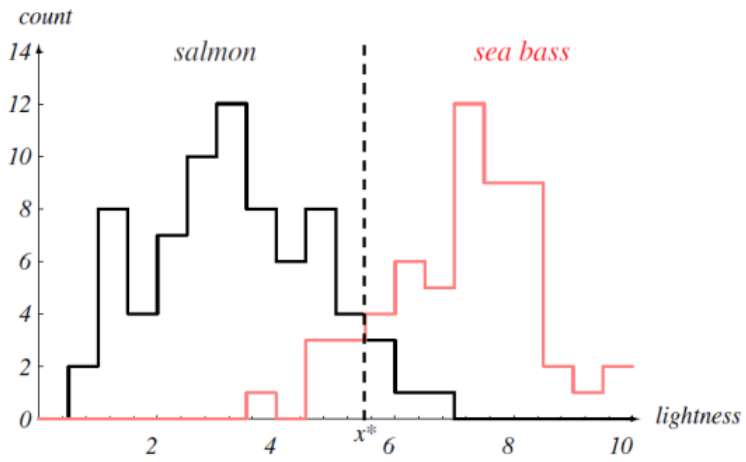
How does a system for a PR task look like?



[DHS]

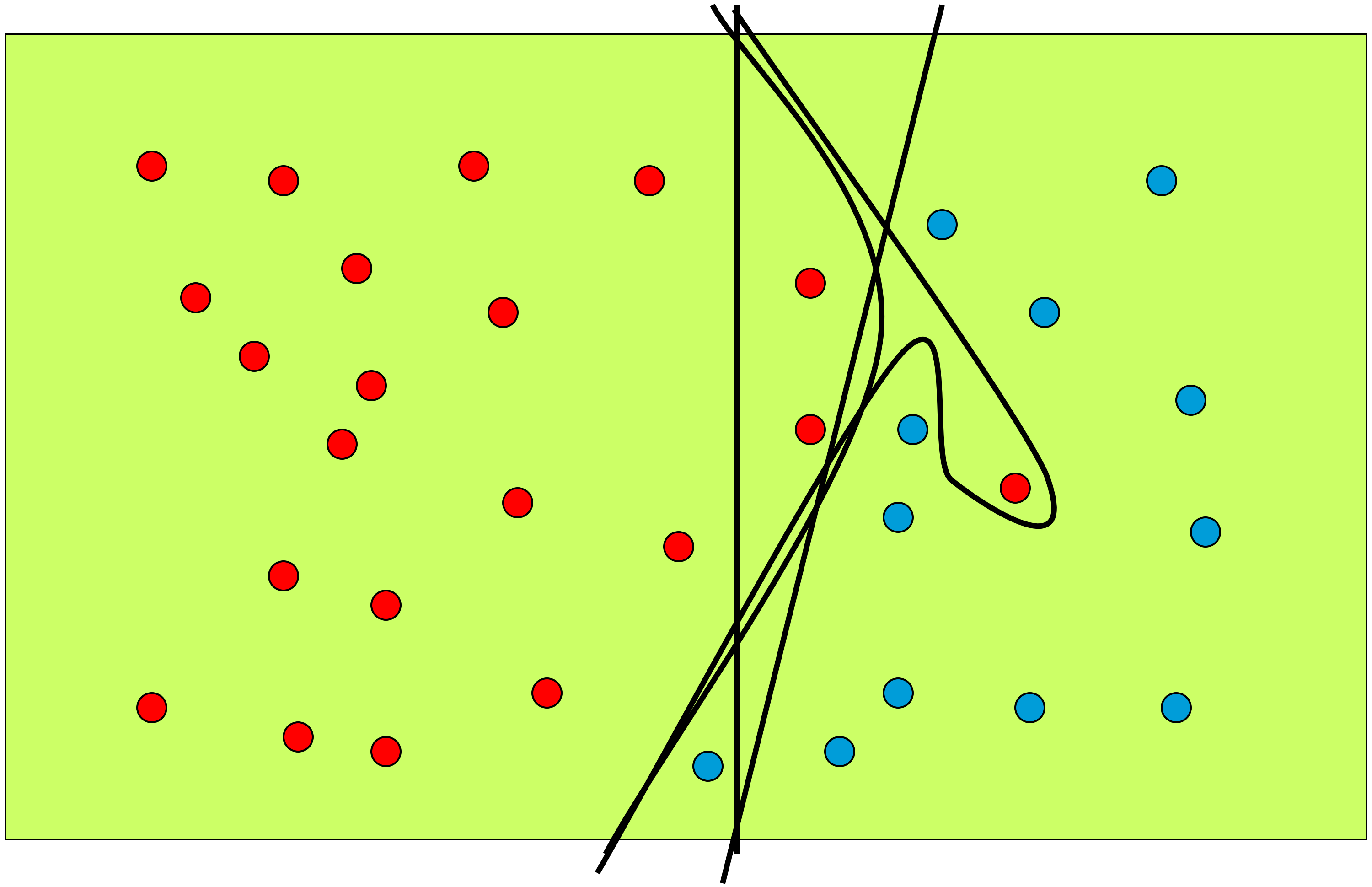
Feature extraction/selection



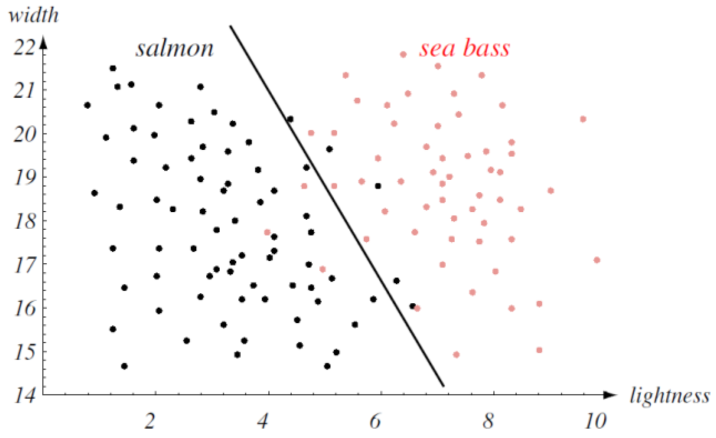


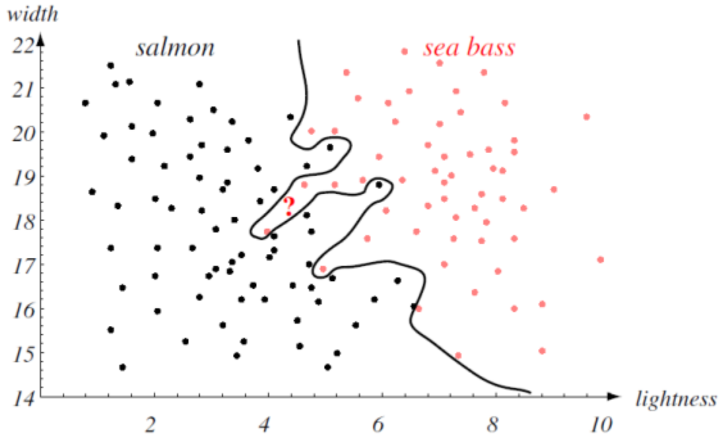
[DHS]

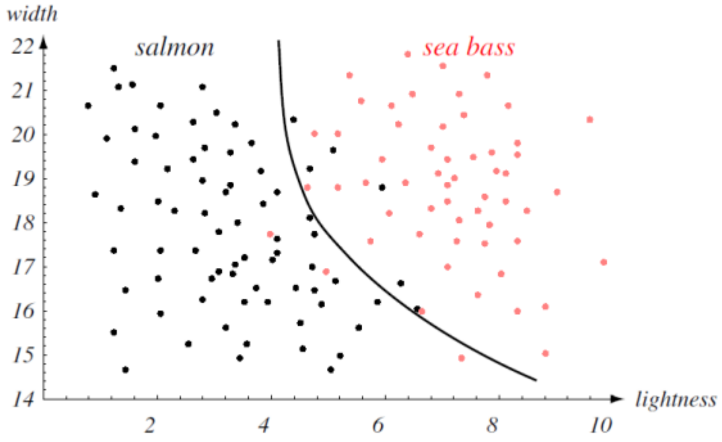
Decision Boundaries



Decision boundary – from simple to complex

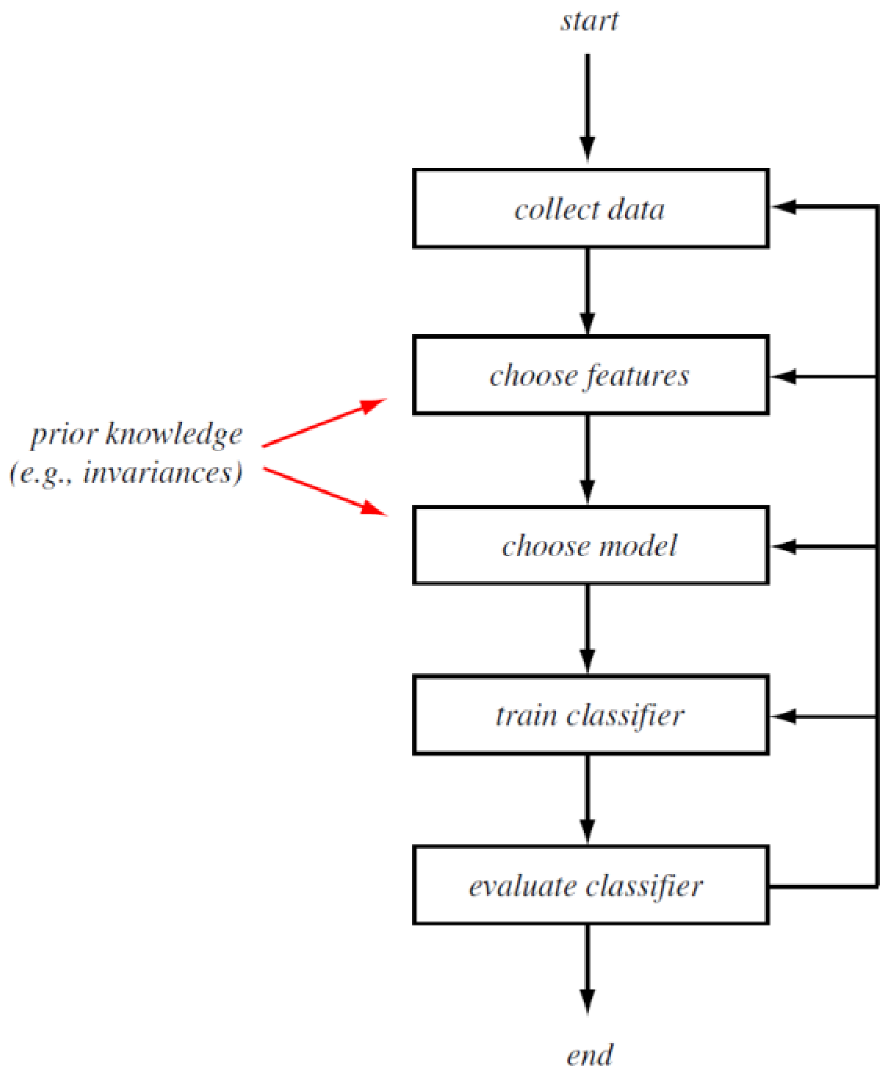






[DHS]

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?
  - Logic/Data-driven approaches

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.

Example PR problem

- Identify a given image as mango or orange.

[HR]

Example ML approach

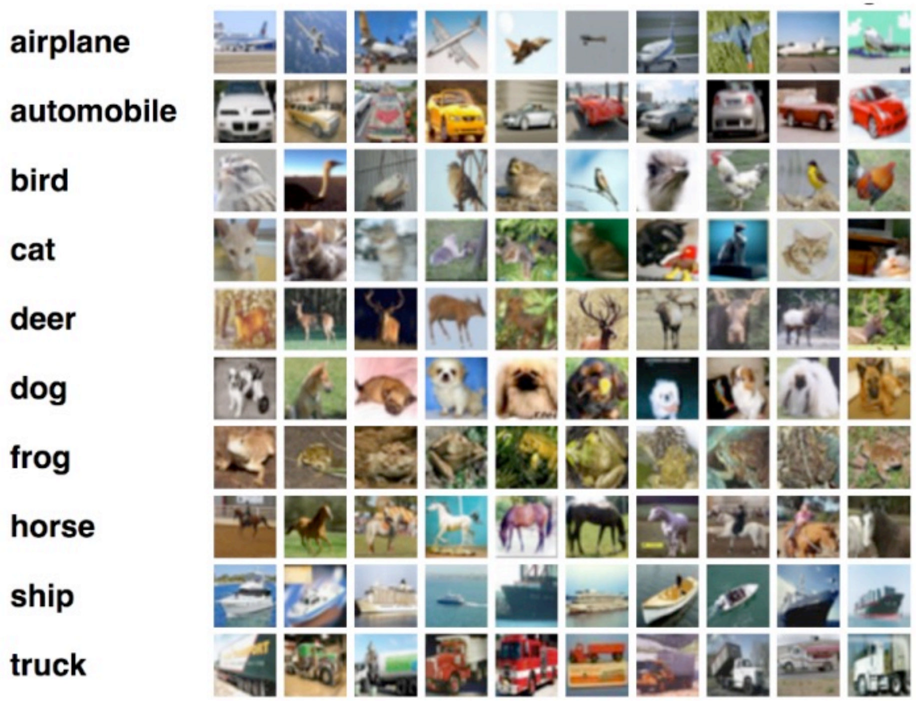
- Identify a given image as mango or orange, given many example images of mango and orange.

M M O M O

[HR]

Example PR/ML problem

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



[HR]

<course logistics begin>

Get to know you - poll

- Level of comfort with prereqs.:

	Moderate	Poor	Very Good
Probability			
Linear algebra			
Calculus			
Optimization			
Programming			

Get to know the TAs

- Rahul Vashisht – Lead TA. Will announce the rest of the TAs shortly.
    - cs18d006@smail.iitm.ac.in

Planned syllabus (tentative)

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

1. Overview of PR/ML problems/algorithms (PR tasks/systems, ML paradigms)
2. Bayesian decision theory (Bayes classifier, loss functions)
3. Density estimation (Maximum likelihood, Bayesian estimation, Expectation Maximization (EM) for mixture density estimation, Non-parametric methods)
4. Linear models for classification and regression (Linear discriminant analysis (hyperplanes), Linear/polynomial regression, 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. Select advanced topics based on time available (E.g., a subset of computational learning theory, algorithms for sequential data (Hidden Markov Models HMM), or graph-structured data; probabilistic graphical models).

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.

Evaluation scheme (tentative)

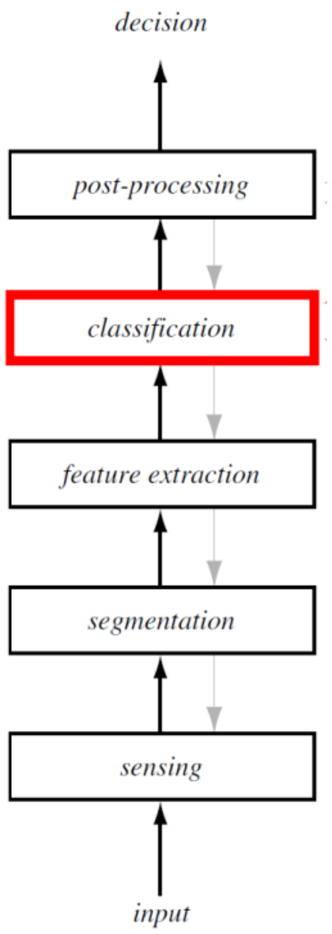
- Aka "Help us help you learn"

- 30% Homework assignments (2 problems/programming-based assignments; done in teams of two students)
- 25% (Real-world) Data contest and writeup (done in teams of two students)
- 5% Tutorial participation – solving problems and interactions during tutorials
- 40% Quiz-I and Quiz-II
  - Proctored, closed-book, online live exam
  - F slot Quiz-II date and F slot EndSem date
- Above amounts to ~6 activities in total – roughly one activity per two weeks.

<course logistics end>

So far: PR system

- PR task and problem
- ML problem
- ML algorithm



[DHS]

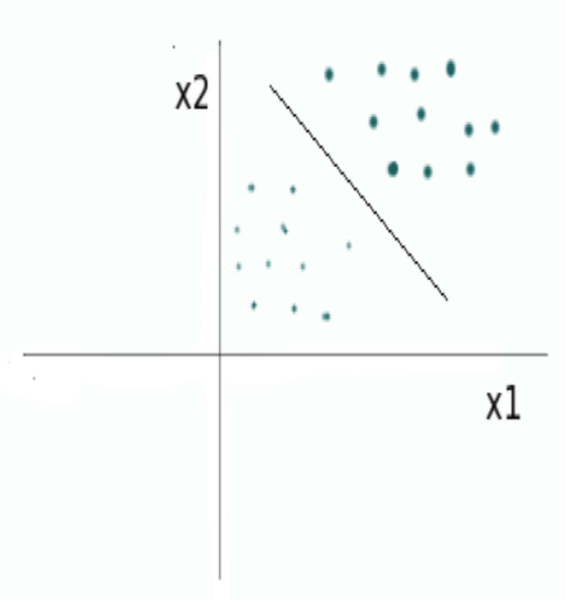
ML Paradigms

- Supervised Learning (informally aka curve-fitting or function learning)
  - 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]

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**

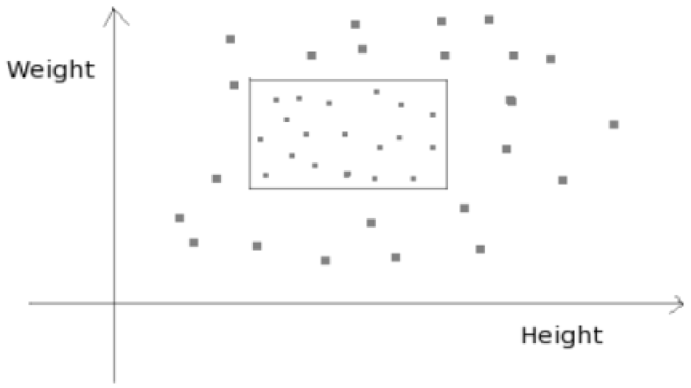


[PSS]

Example (non-linear) classification problem

$f(ax+by) = af(x) + bf(y)$

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

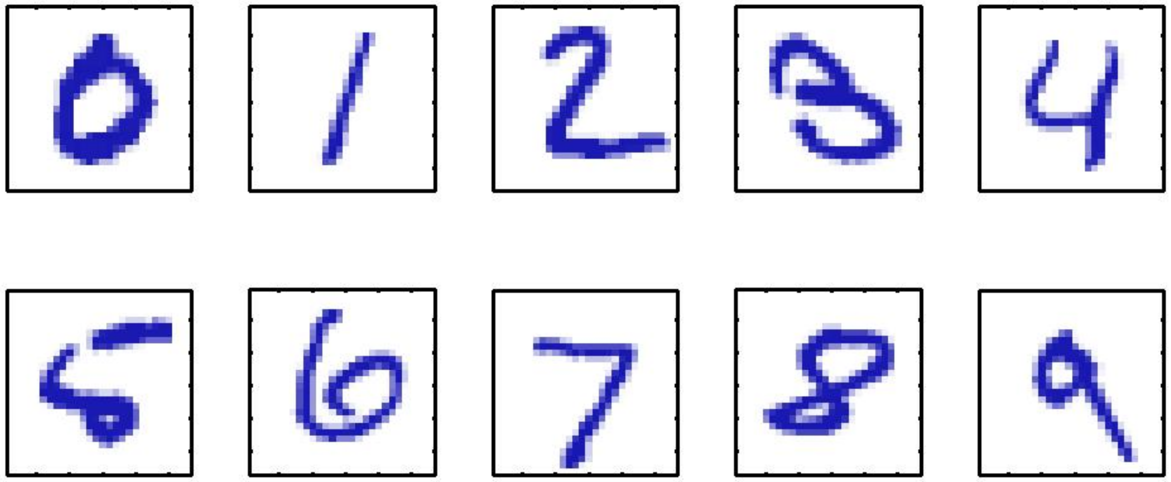


The classifier is ‘nonlinear’ here

[PSS]

Example classification problem

- Handwritten Digit Recognition:
  - 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.



[CMB]

Classification Applications

- Credit Card fraud detection
  - Valid transaction or not
- Sentiment Analysis
  - Opinion mining; buzz analysis; etc.
- Churn prediction
  - Potential churner or not
- Medical diagnoses
  - Risk analysis

Regression (vs. classification) problem

- Closely related problem. Output is continuous-valued rather than discrete as in classifiers.
- Here training set examples could be $\{(X_i, y_i),\ \ i = 1, \ldots, \ell\}$, $X_i \in \mathcal{X}$, $y_i \in \Re$.
- The prediction variable, y, is continuous; rather than taking finitely many values. (There can be noise in examples).
- Similar learning techniques needed to infer the underlying functional relationship between X and y. (Regression function of y on X).

$X \quad Y \in \{-1, +1\}$

$\{1, 2, \ldots, K\}$

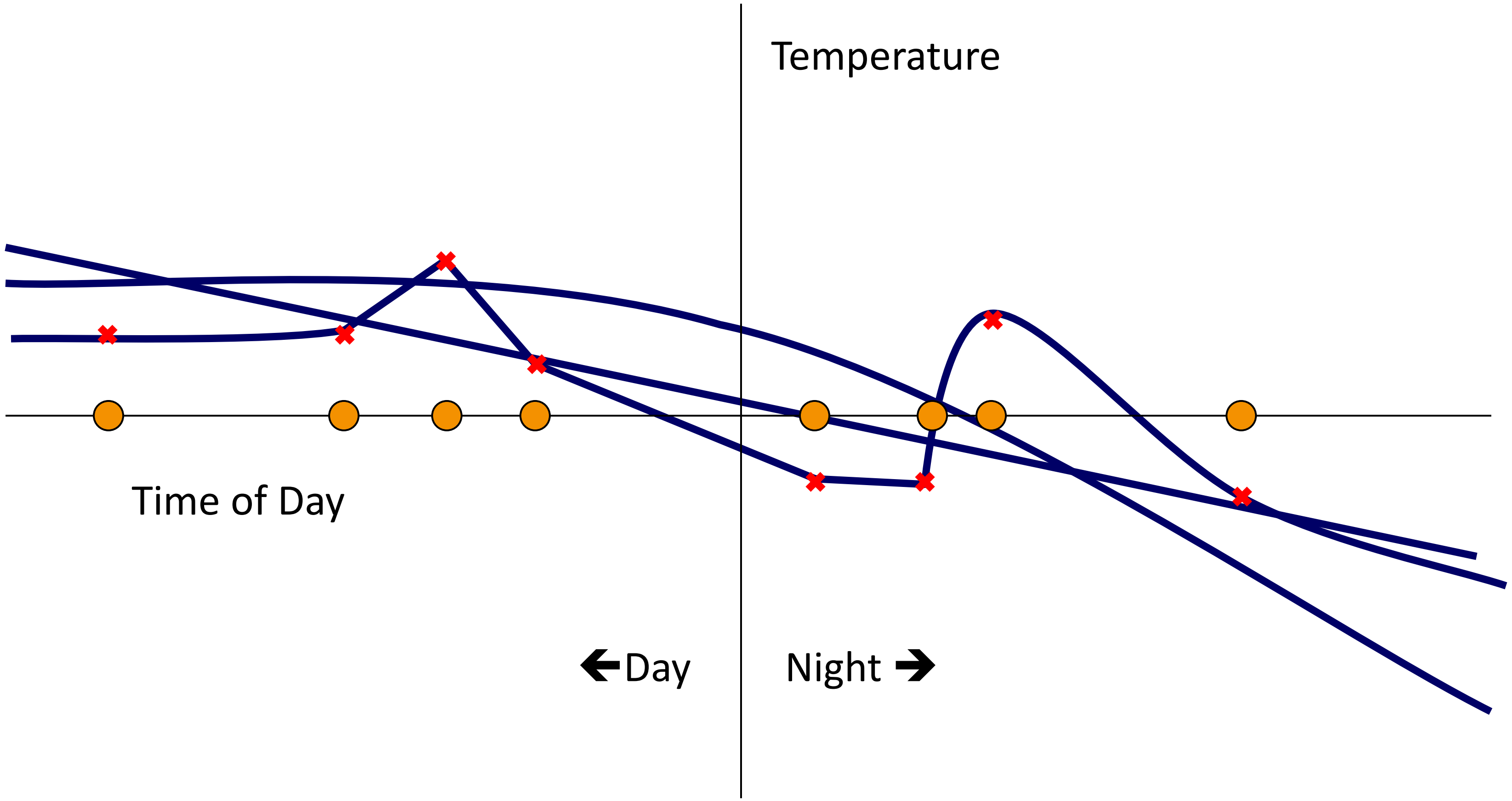
$X \quad Y \in [0, 1]$

$Y \in \mathbb{R}$

$Y \in \cancel{\mathbb{R}}$

[PSS]

Prediction or Regression



Example regression problems

- Time series prediction: Given a series $x_1, x_2, \cdots$, find a function to predict x_n.
- Based on past values: Find a 'best' function $\hat{x}_n = h(x_{n-1}, x_{n-2}, \cdots, x_{n-p})$
  - Predict stock prices, exchange rates etc.
  - Linear prediction model used in speech analysis
- More general predictors can use other variables also.
  - Predict rainfall based on measurements and (possibly) previous years' data.
  - In general, System Identification. (An application: smart sensors)

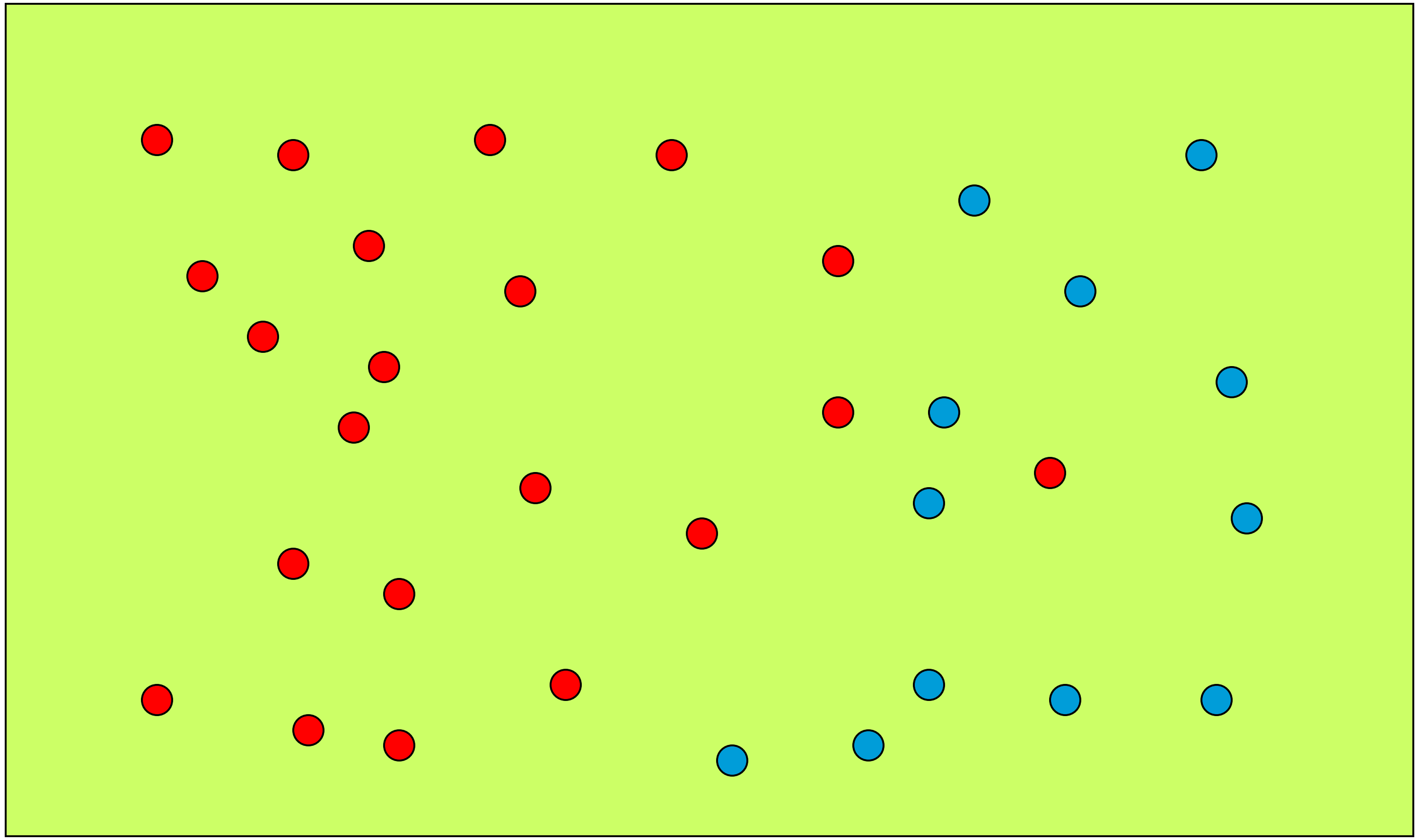
[PSS]

ML Paradigms

- Supervised Learning (informally aka curve-fitting or function learning)
  - 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.

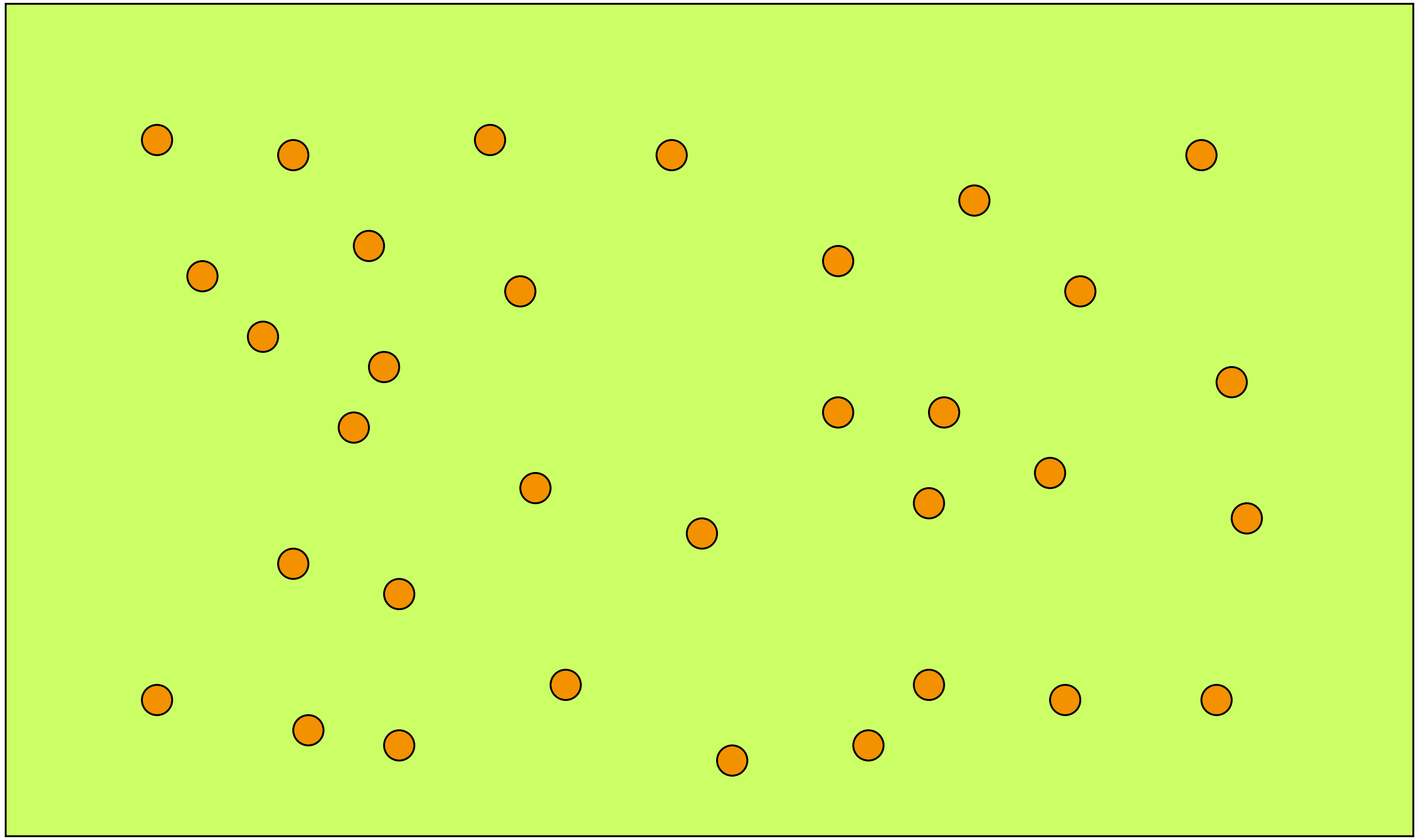
[BR,HR]

Labeled Training Data

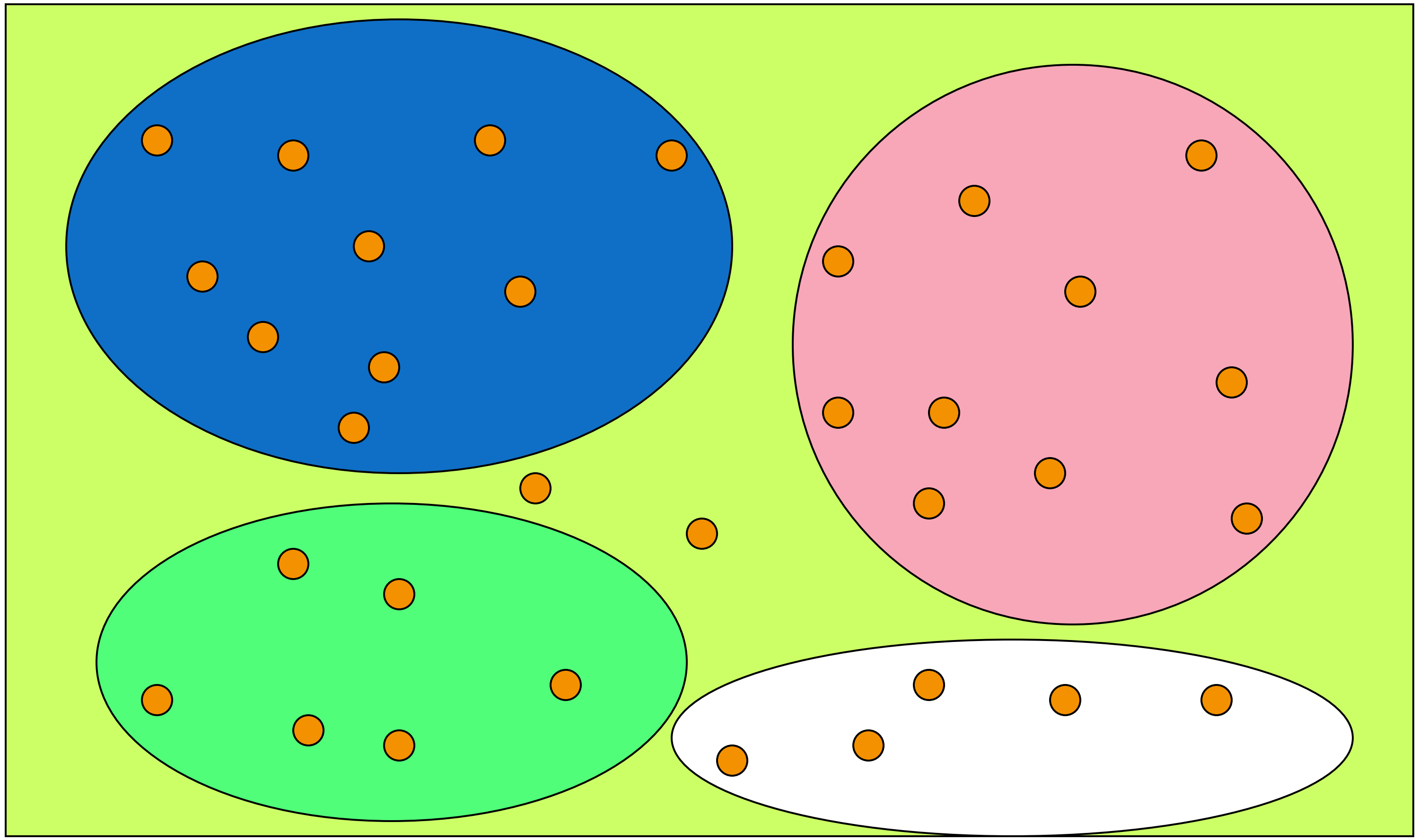


Unlabelled Training Data

Clustering



Possible Clusters



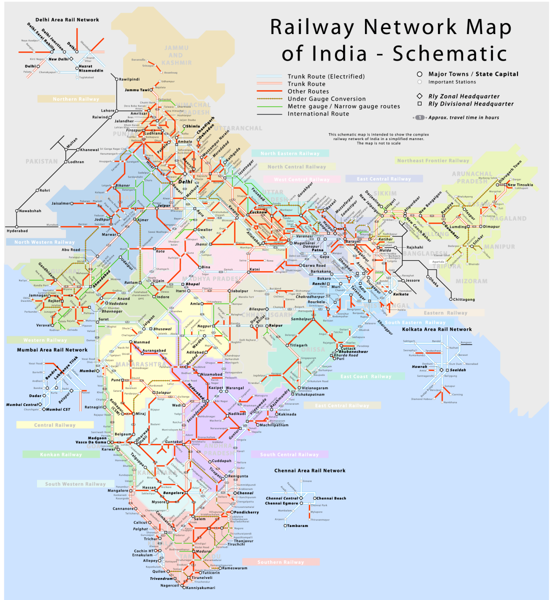
Clustering Applications

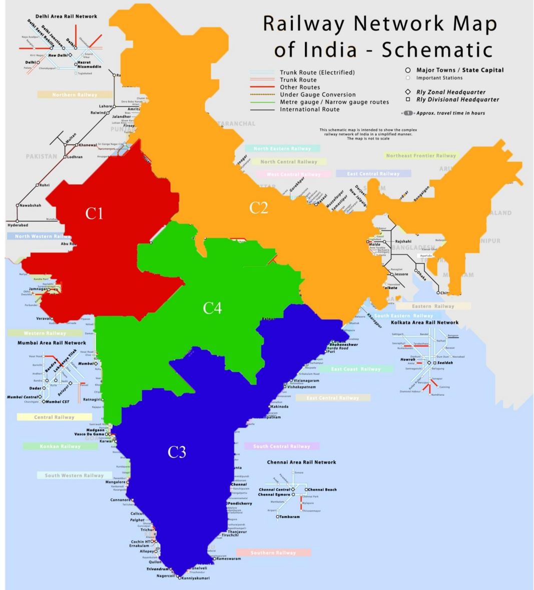
- Customer Data
  - Discover classes of customers
- Image pixels
  - Discover regions
- Biomarkers
  - Genes or proteins vs. disease
- Transportation Networks
  - Zones

Image Courtesy: http://cs.brown.edu/~pff/segment/

Clustering Applications

- Customer Data
  - Discover classes of customers
- Image pixels
  - Discover regions
- Biomarkers
  - Genes or proteins vs. disease
- Transportation Networks
  - Zones





ML Paradigms

- Supervised Learning (informally aka curve-fitting or function learning)
  - 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]

Association Rule Mining

- Mining frequent patterns and rules
- Association rules: conditional dependencies
- Two stages
  - Find frequent patterns
  - Derive associations (A => B) from frequent patterns
- Find patterns in
  - Sequences (time series data, fault analysis)
  - Transactions (market basket data)
  - Graphs (social network analysis)

Mining Transactions

- Transaction is a collection of items bought together
  - A (sub)set of items is called an itemset
- Find frequent itemsets
- Itemset A=> itemset B, if both A and A U B are frequent itemsets.

Transaction data bases

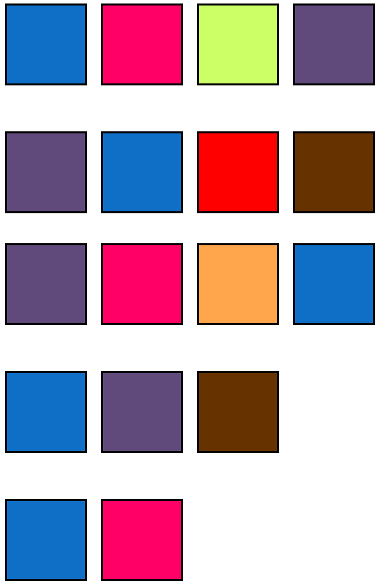
T1:

T2:

T3:

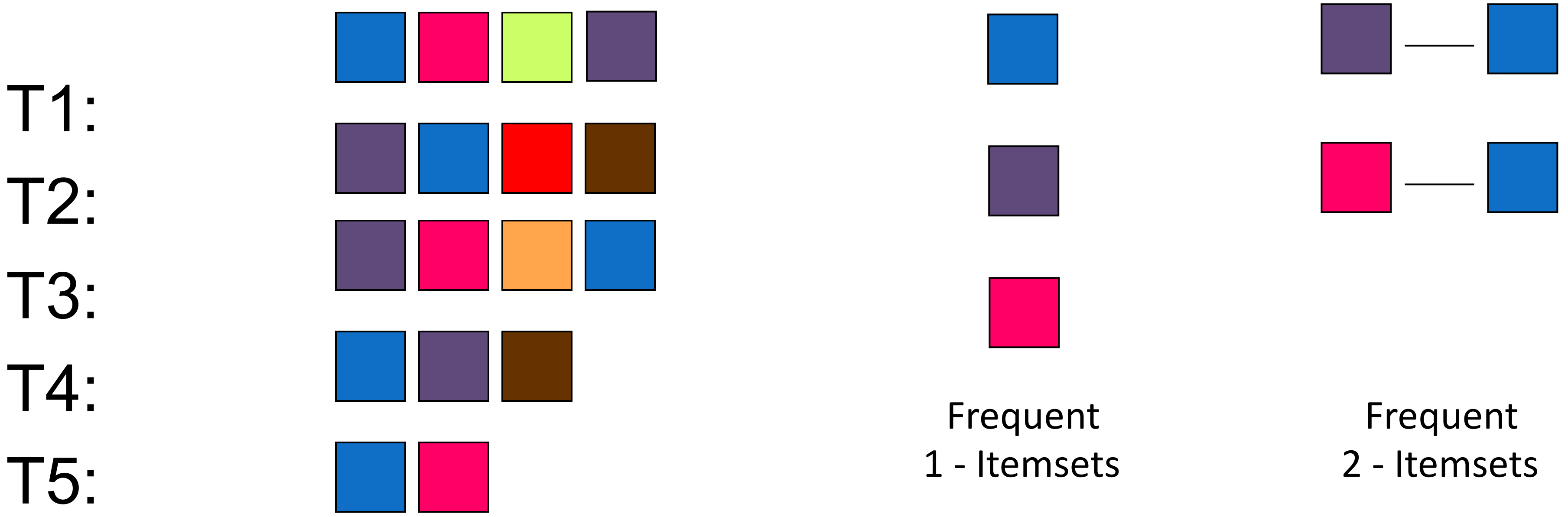
T4:

T5:



Frequency Threshold: 3

Frequent Pattern Mining



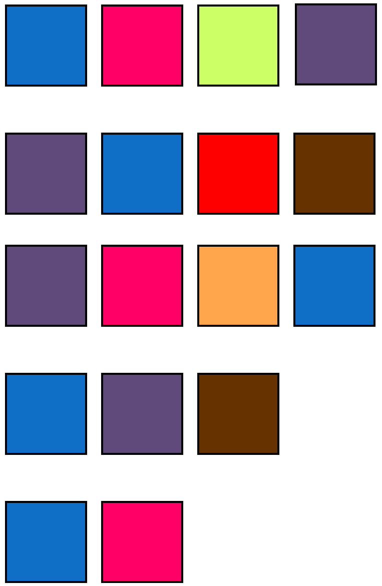
Frequency Threshold: 3

Mining Transactions

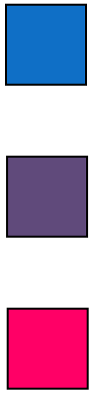
- Transaction is a collection of items bought together
  - A (sub)set of items is called an itemset
- Find frequent itemsets
- Itemset A => itemset B, if both A and A U B are frequent itemsets.
- Support of a rule is the percentage of itemsets containing A U B
- Confidence of a rule is the percentage of itemsets containing A that also contain A U B
- We look for rules with both high support and confidence
  - Can be determined from the frequent itemsets; hence more effort focused on that

Frequent Pattern Mining

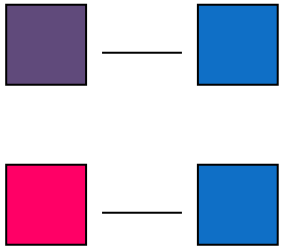
T1:
T2:
T3:
T4:
T5:



Frequency Threshold: 3



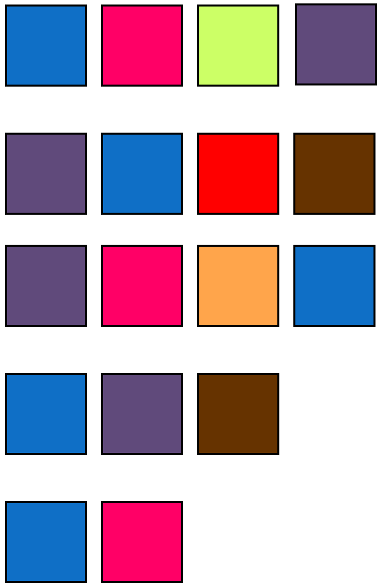
Frequent
1 - Itemsets

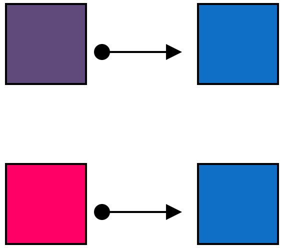


Frequent
2 - Itemsets

Association Rules

T1:
T2:
T3:
T4:
T5:





Support: 4/5; Conf: 1

Support: 3/5; Conf: 1

Association Rules

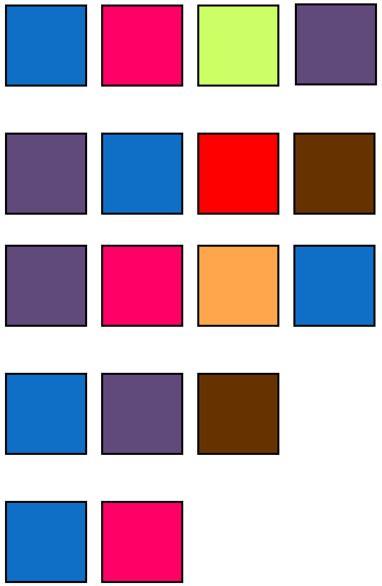
T1:

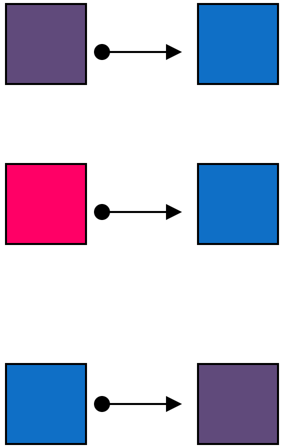
T2:

T3:

T4:

T5:





Support: 4/5; Conf: 1

Support: 3/5; Conf: 1

Support: 4/5; Conf: 4/5

Association Mining Applications

- Market Basket analysis
- Topic identification
  - co-occurrence of words
- Plagiarism Detection!
- Biomarkers
  - Genes or proteins vs. disease
- Time series analysis!
  - Trigger Events

ML Paradigms

- Supervised Learning (informally aka curve-fitting or function learning)
  - 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]

Learning to Control

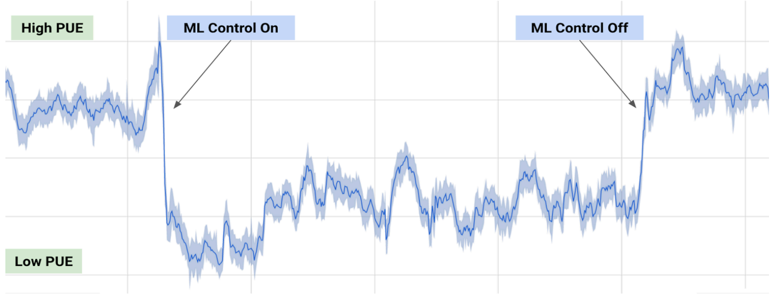
- Familiar models of machine learning
  - Learning from data.
- How did you learn to cycle?
  - Not from Data!
  - Trial and error!
  - Falling down hurts!
  - Learn from Evaluation!

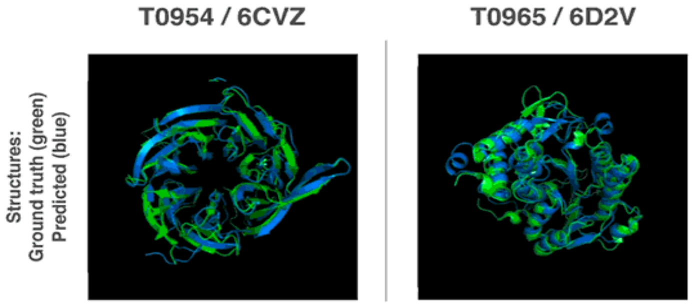
Learning to Control

- Familiar models of machine learning
  - Learning from data.

- How did you learn to cycle?
  - Not from Data!
  - Trial and error!
  - Falling down hurts!
  - Learn from Evaluation!
- Walking, Talking, etc.

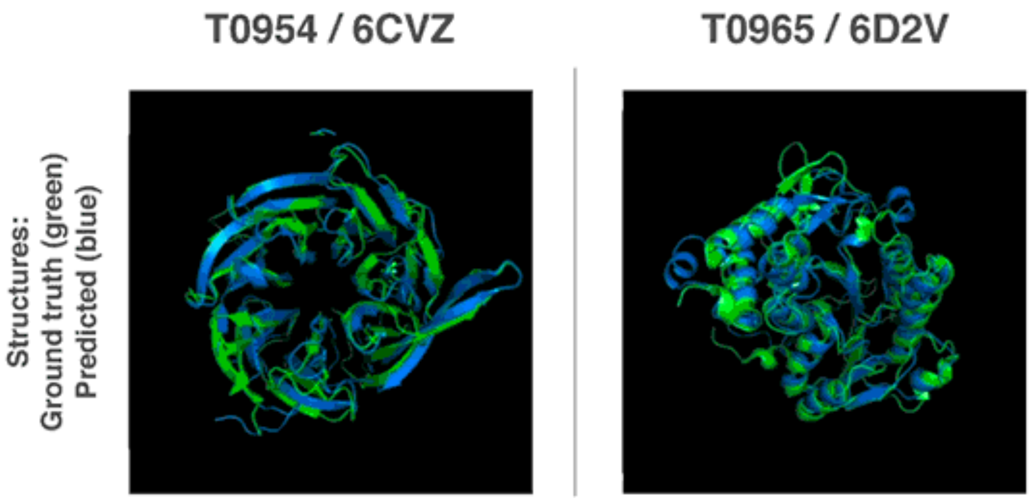
Reinforcement Learning Works!

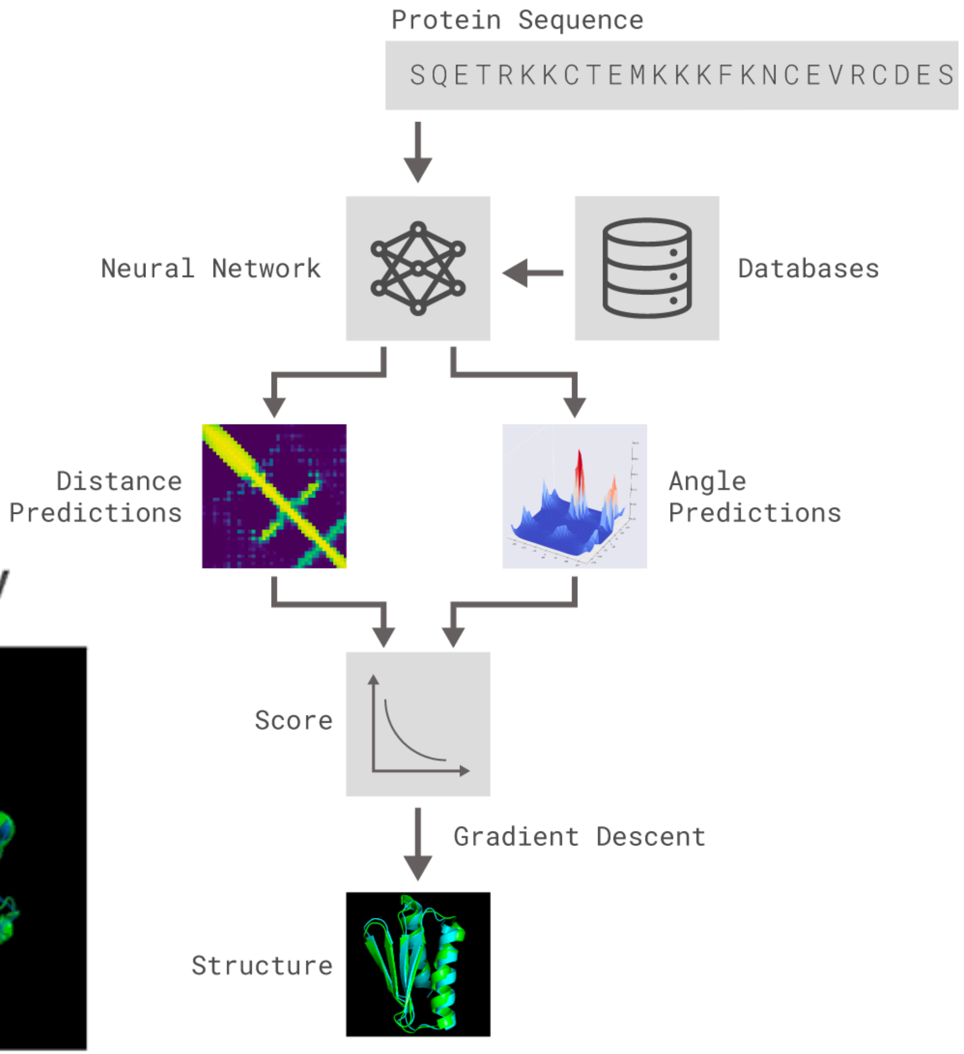




AlphaFold

- Protein Folding: One of the hardest problems in biology
- An AI agent achieved 25% improvement over best human effort

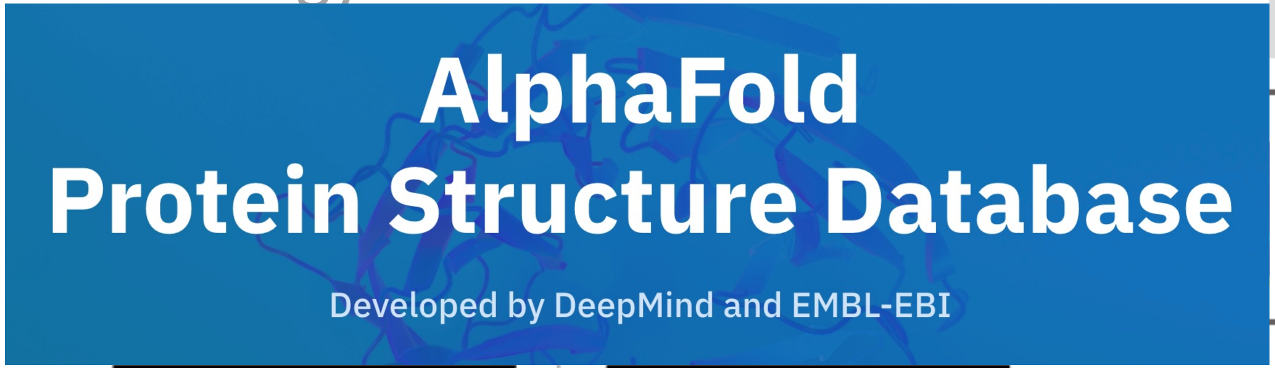


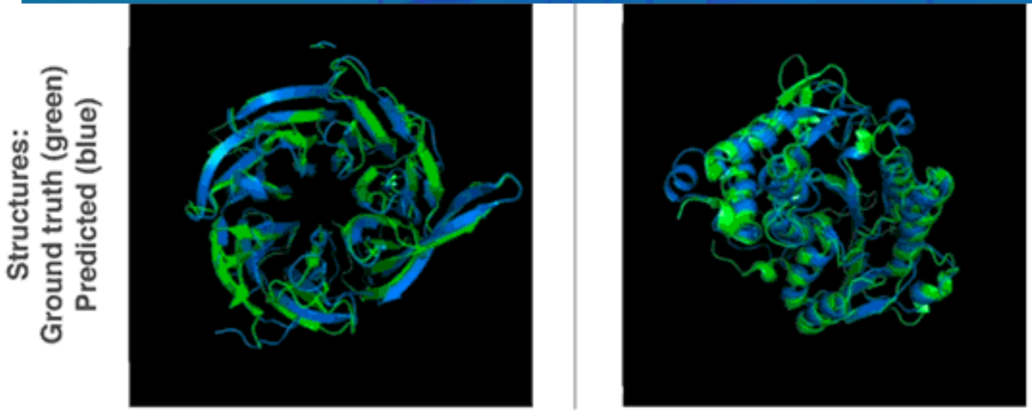


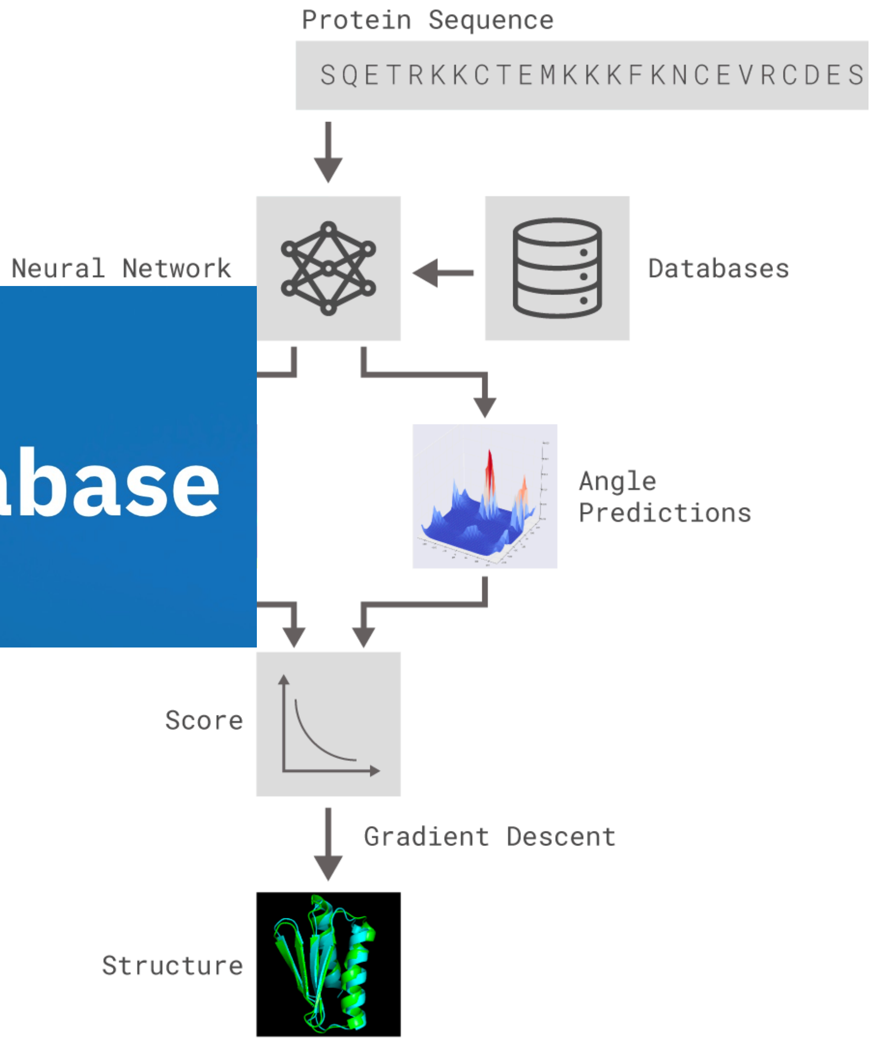
AlphaFold 2.0

- Protein Folding: One of the hardest problems in biology

SOLVED?







Get to know the TAs

- Rahul Vashisht – Lead TA. Will announce the rest of the TAs shortly.
  - cs18d006@smail.iitm.ac.in