Artificial Intelligence Foundations and Applications

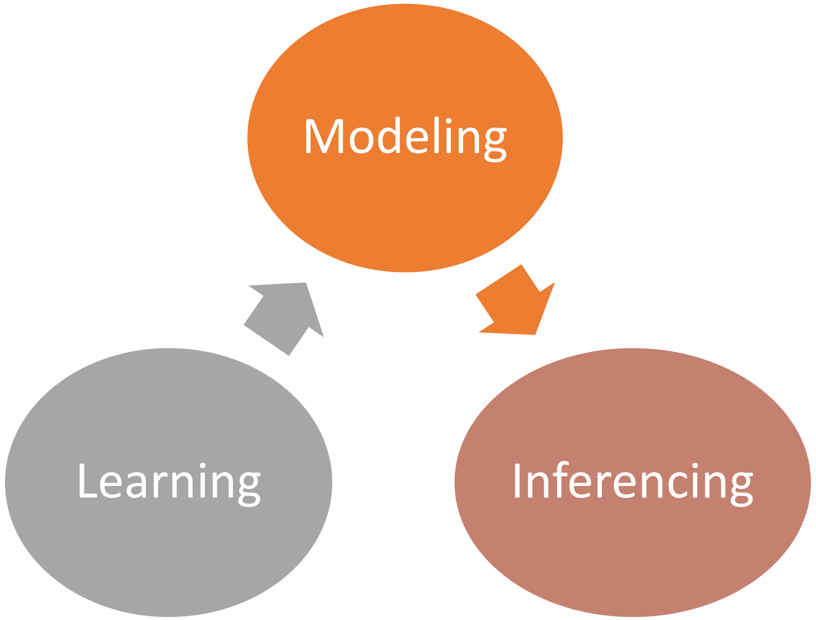
Machine Learning – Part 1
Introduction

Sudeshna Sarkar

Centre of Excellence in Artificial Intelligence, IIT Kharagpur

Machine Learning : Definition

Learning is the ability to evolve behaviours based on data (experience).



- Learn from data such as build a model from data
- Use the model for prediction, decision making or solving some tasks

Components of a learning problem

- **Task:** The behaviour or task being improved.
  - For example: classification, acting in an environment
- **Data:** The experiences that are being used to improve performance in the task.
- **Measure of improvement :**
  - For example: increasing accuracy in prediction, improved speed and efficiency

Models for Prediction

Input x → Predictor f → Output y

A. Features

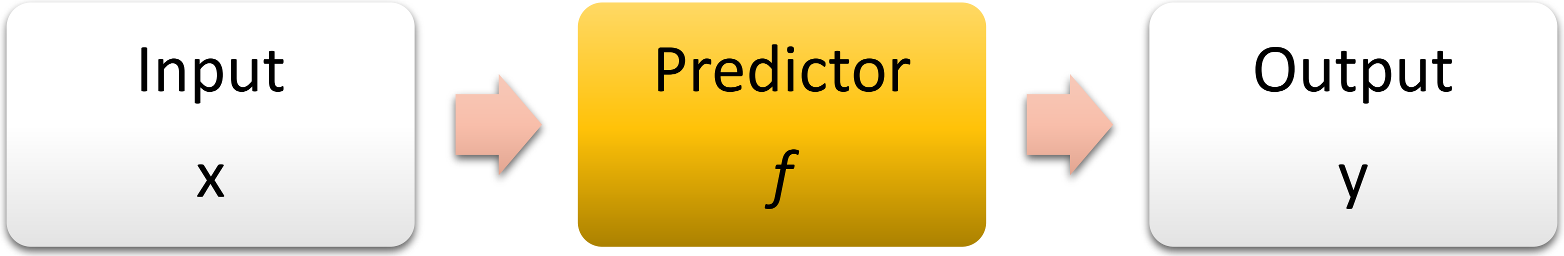
- Feature vector of *n* features

$$\bar{x} = (x_1, x_2, \dots, x_n)$$

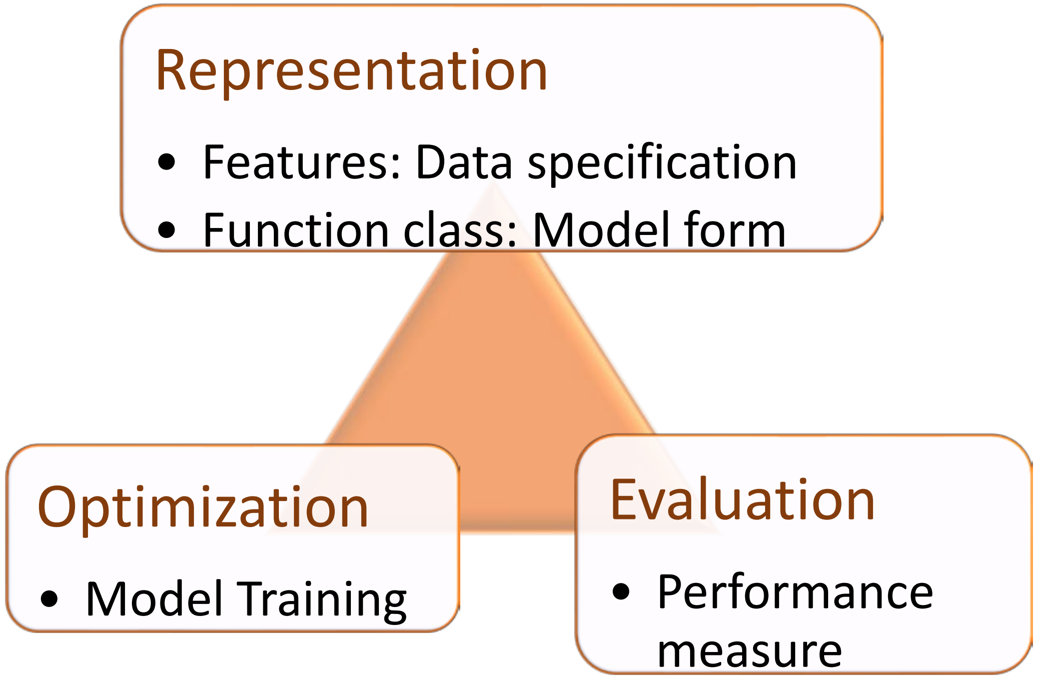
B. Convert input to a vector of basis functions

$$\left(\phi_0(\bar{x}), \phi_1(\bar{x}), \dots, \phi_p(\bar{x})\right)$$

Design a Learner

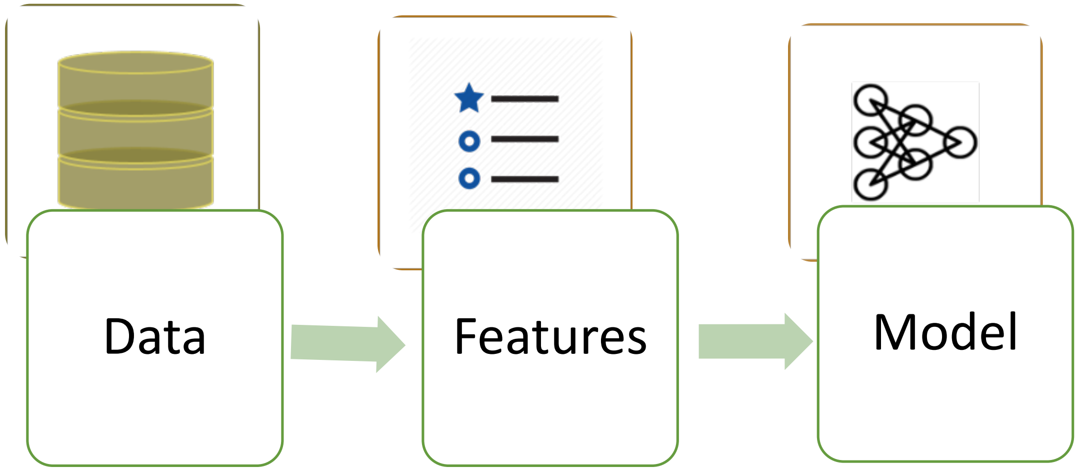


1. Choose the training experience
   - Features
2. Choose how to represent the target function *f*
3. Choose a learning algorithm to infer the target function



Feature Choice

- Input Data comprise features
  - Structured features (numerical or categorical values)
  - Unstructured (text, speech, image, video, etc)
- Use only relevant features
- Too many features?
  - Select feature subset (reduction)
  - Extract features: Transform features



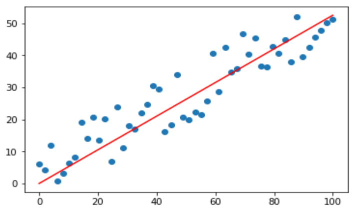
1. Model Representation

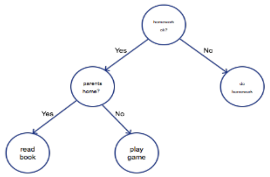
- The richer the representation, the more useful it is for subsequent problem solving.
- The richer the representation, the more difficult it is to learn.

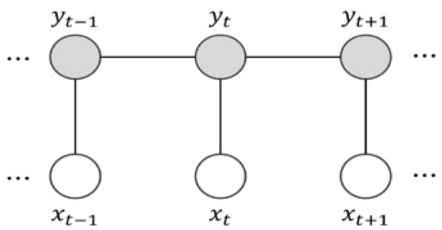
$$y = f(\bar{x})$$

$$y = g(\bar{\phi}(\bar{x}))$$

Linear Function	Decision Tree	Graphical Model	Neural Net







$$Y = w_0 + w_1 x_1 + \cdots + w_p x_p + \epsilon$$

2. Evaluation

1. Accuracy $= \dfrac{\text{\# correctly classified}}{\text{\# all test examples}}$
2. Logarithmic Loss:

$$L_i = -\log(P(Y = y_i | X = x_i))$$

$$L = \sum_{c=1}^{M} y_{oc} \log(p_{oc})$$

3. Mean Squared error

$$MSE = \frac{1}{m} \sum \left(y_{pred} - y_{true}\right)^2$$

3. Optimization

1. Define loss function
2. Optimize loss function

- Stochastic Gradient Descent (Convex functions)
- Combinatorial optimization
  - E.g.: Greedy search
- Constrained optimization
  - E.g.: Linear programming

Broad types of machine learning

- **Supervised Learning**
  - Training Data with labels: X,y (pre-classified)
  - Given an observation x, what is the best label for y?
- **Unsupervised learning**
  - Training Data without labels: X
  - Given a set of x's, find hidden structure
- **Reinforcement Learning**
  - Given: observations and periodic rewards as the agent takes sequential action in an environment
  - Determine optimum policy

Supervised Learning

Given data containing the inputs and outputs:

Training Data:

$$\{(\overline{x_1}, y_1), (\overline{x_2}, y_2), \dots, (\overline{x_m}, y_m)\}$$

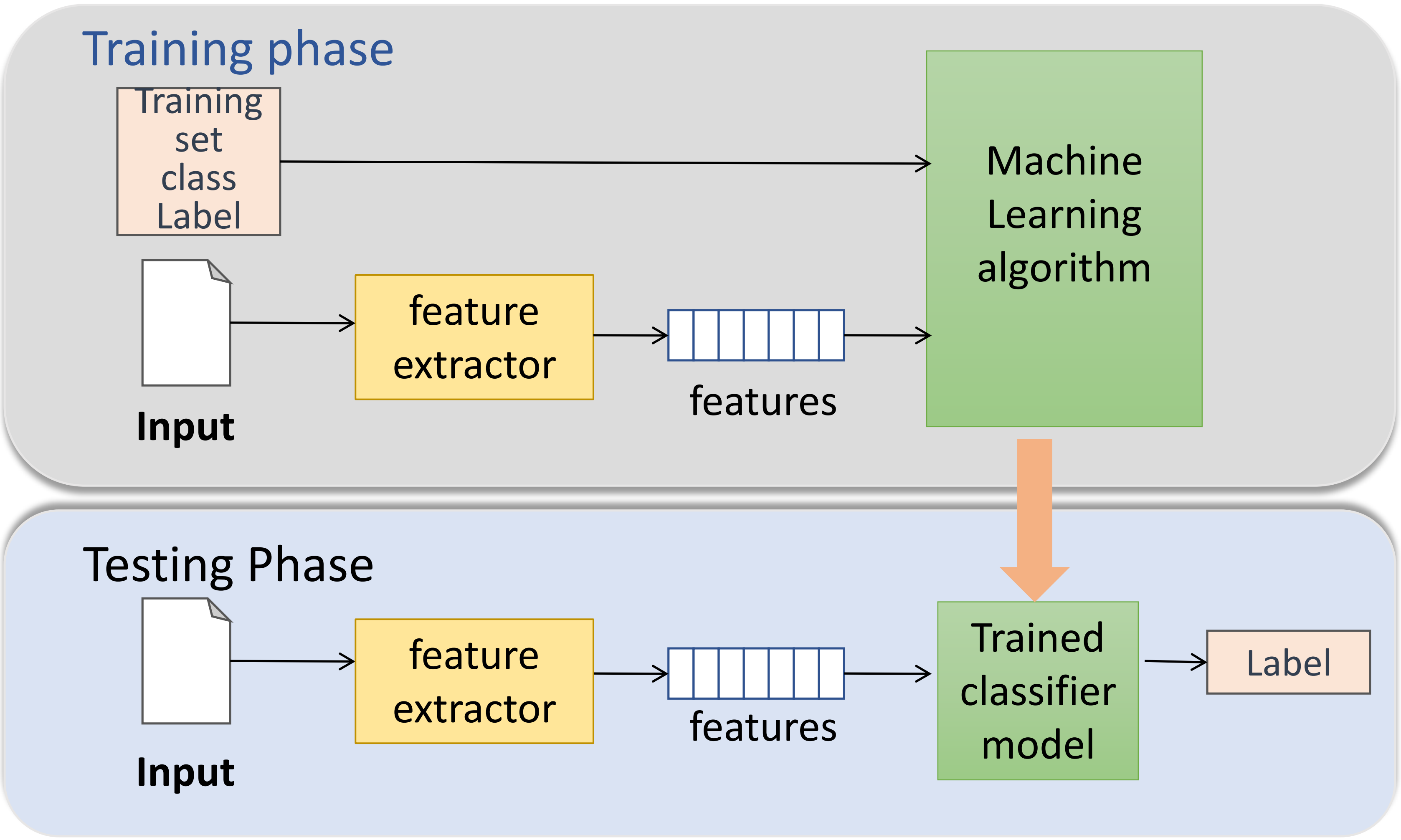
- Learn a function *f*(*x*) to predict *y* given *x*

$\bar{X}$	Y
$\overline{x_1}$	y_1
$\overline{x_2}$	y_2
…	..
$\overline{x_m}$	y_m

Model $\longrightarrow f_\theta(\bar{x}) \to y$

Parameters

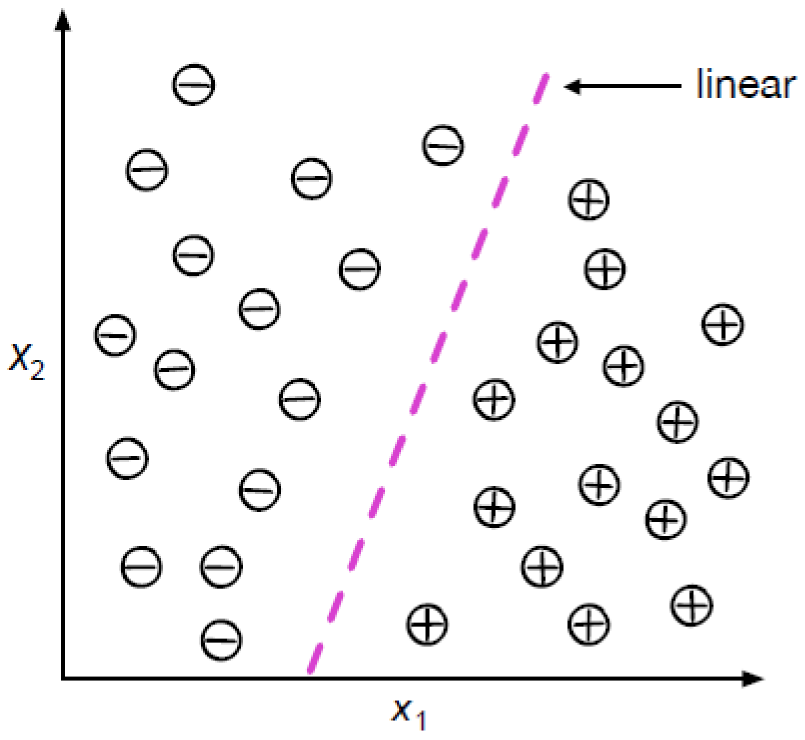
- Training: Learn the model from the Training Data
- Given Test instance $\overline{x'}$, predict $y' = f_\theta(\overline{x'})$


Training phase
Training set class Label
Machine Learning algorithm
feature extractor
features
Input
Testing Phase
feature extractor
features
Trained classifier model
Label
Input

Supervised Learning

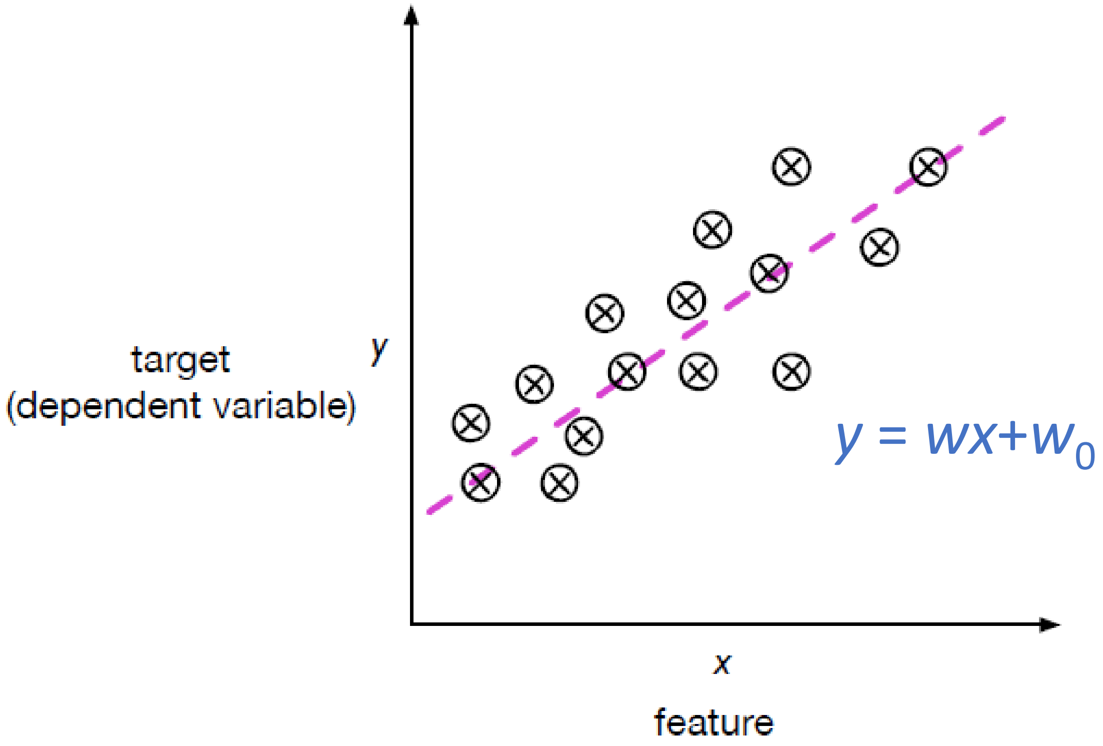
Classification

Y is categorical/ discrete



Regression

Y is numeric / continuous



Example Tasks

Classification

- Object identification from images
- Defect classification
- Credit card transaction fraud or not

Regression

- House price prediction
- Remaining Useful Life Prediction
- Probability of developing cracks
- Demand forecasting

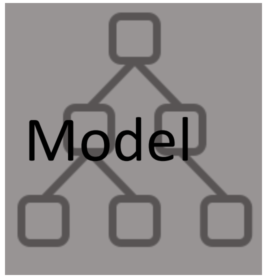
Structured Prediction

- Machine translation: English sentence → Japanese sentence
- Dialogue: conversational history → next utterance
- Image captioning: image → sentence describing image
- Image segmentation: image → segmentation

Supervised Learning

Classification Example

Training
Samples



Train a model to minimize loss

Test
Instances

Probabilistic Classification

X1	X2	X3	X4	Category
				Type 1
				Type 3
				Type 1
				Type 2

Predict a probability distribution over the set of classes **Pr (Y|X)**

X1	X2	X3	X4	Type 1	Type 2	Type 3
				0.4	0.15	0.45
				0.2	0.7	0.1
				0.1	0.2	0.7
				0.5	0.1	0.4

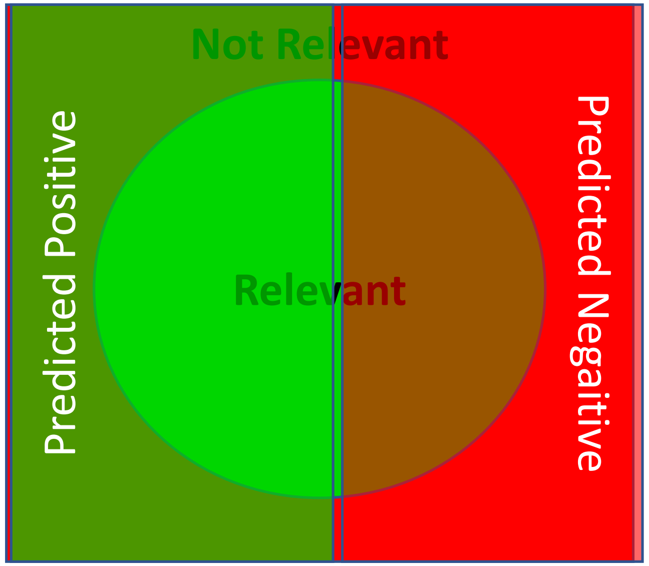
Evaluation for Classification problems

$$\textbf{Accuracy} = \frac{\text{\# correctly classified}}{\text{\# all test examples}}$$

$$= \frac{\#\text{predicted true } pos + \#\text{predicted true } neg}{\#\text{all test examples}}$$

$$\textbf{Precision} = \frac{\text{\# predicted true } pos}{\text{\# predicted pos}}$$

$$\textbf{Recall} = \frac{\text{\# predicted true } pos}{\text{\# True } pos}$$



		True Class	
		Pos	Neg
Predicted Class	Pos	TP	FP
	Neg	FN	TN

Loss Function Classification problems

Loss indicates how bad the model's prediction is.

1. Fraction of Misclassifications

$$Error = \sum_{i=1}^{m} \frac{I(y_i \neq \widehat{y_i})}{m}$$

2. Logarithmic Loss: Maximize the log likelihood. For a loss function, minimize the negative log likelihood of the correct class:

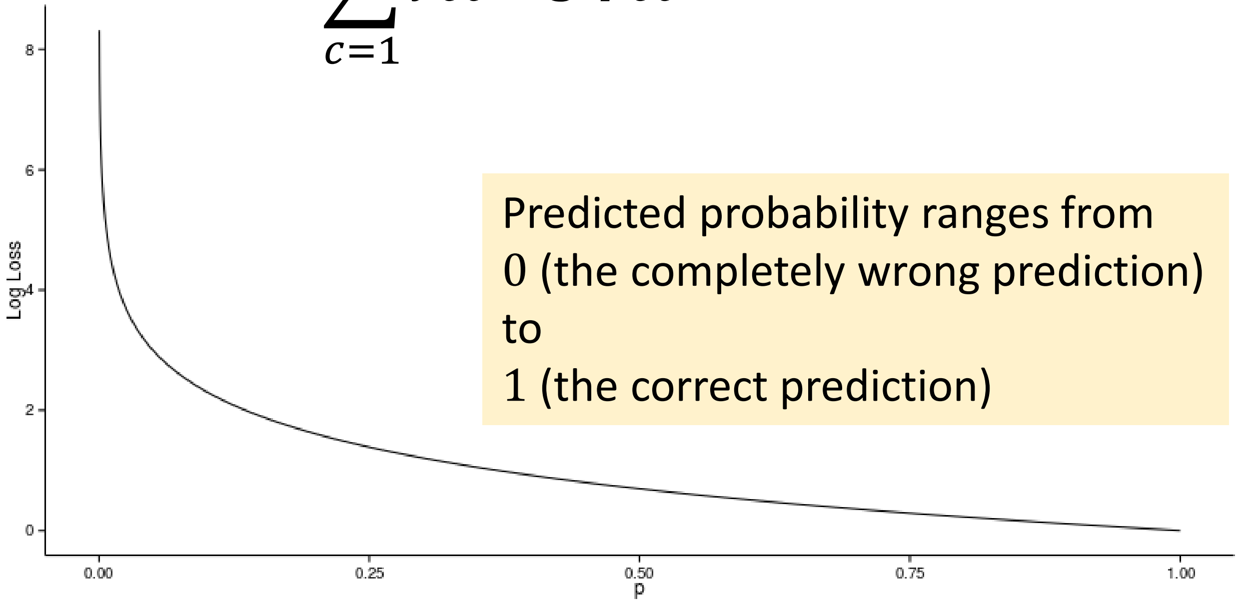
$$L_i = -\log(P(Y = y_i | X = x_i))$$

Logarithmic Loss Function

Logarithmic Loss:

$$L_i = -\log(P(Y = y_i | X = x_i))$$

$$L = \sum_{c=1}^{M} y_{oc} \log(p_{oc})$$



Predicted probability ranges from
0 (the completely wrong prediction)
to
1 (the correct prediction)

M - number of classes
y - binary indicator (0 or 1) if class label ***c*** is the correct classification for observation ***o***
p - predicted probability observation ***o*** is of class ***c***

2. Evaluation for regression problem

Mean Squared error

$$MSE = \frac{1}{m}\sum_{i=1}^{m}(\widehat{y_i} - y_i)^2$$

Predicted value

True value

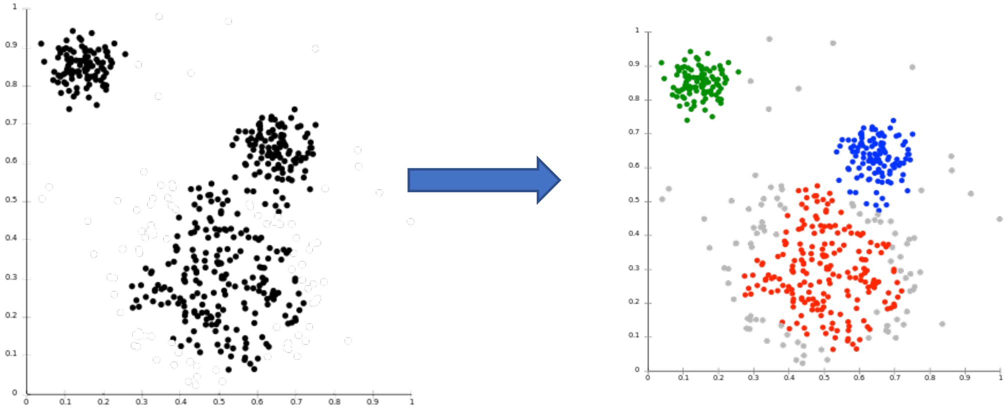
Unsupervised Learning (Clustering)

Given $\{\overline{x_1}, \overline{x_2}, \dots \overline{x_m}, \}$ without labels

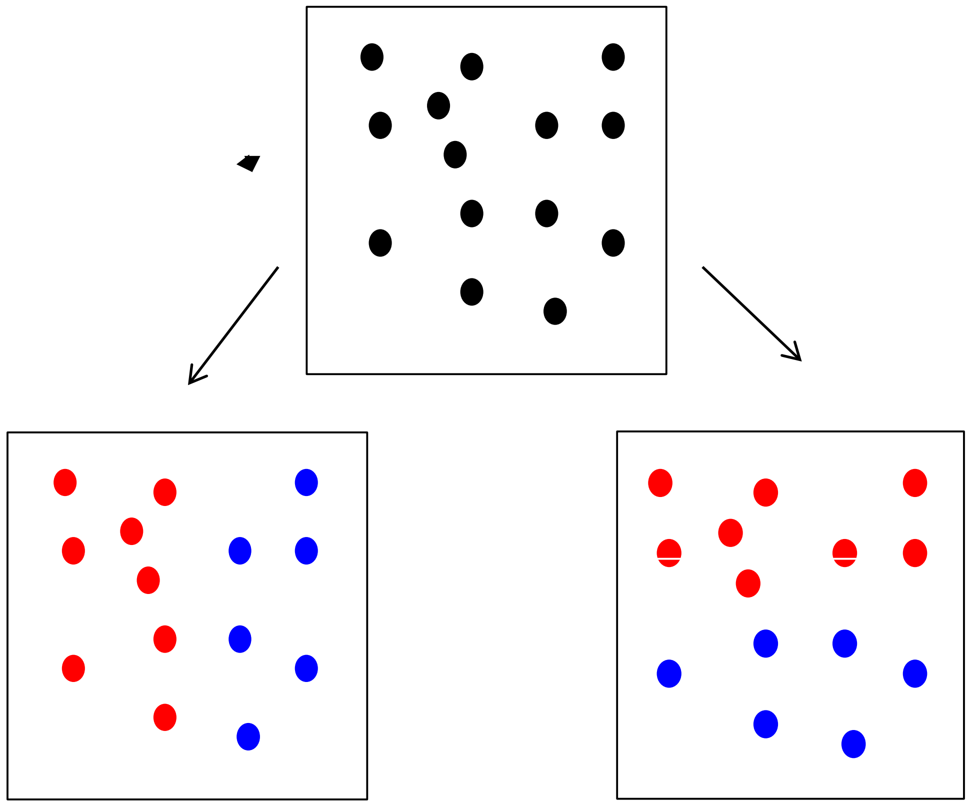
Find hidden structure in the data

- Clustering
- Dimensionality Reduction

Clustering: Grouping similar objects



Clustering Problems



How to evaluate clustering?

- Internal Evaluation:
  - Intra-cluster distances are minimized
  - Inter-cluster distances are maximized
- External Evaluation