A Basic Introduction to Machine Learning

Machine Learning Basics

Sudeshna Sarkar

17 – 18 Sep 2020

Machine Learning

- Provide systems the ability to automatically learn and improve from experience

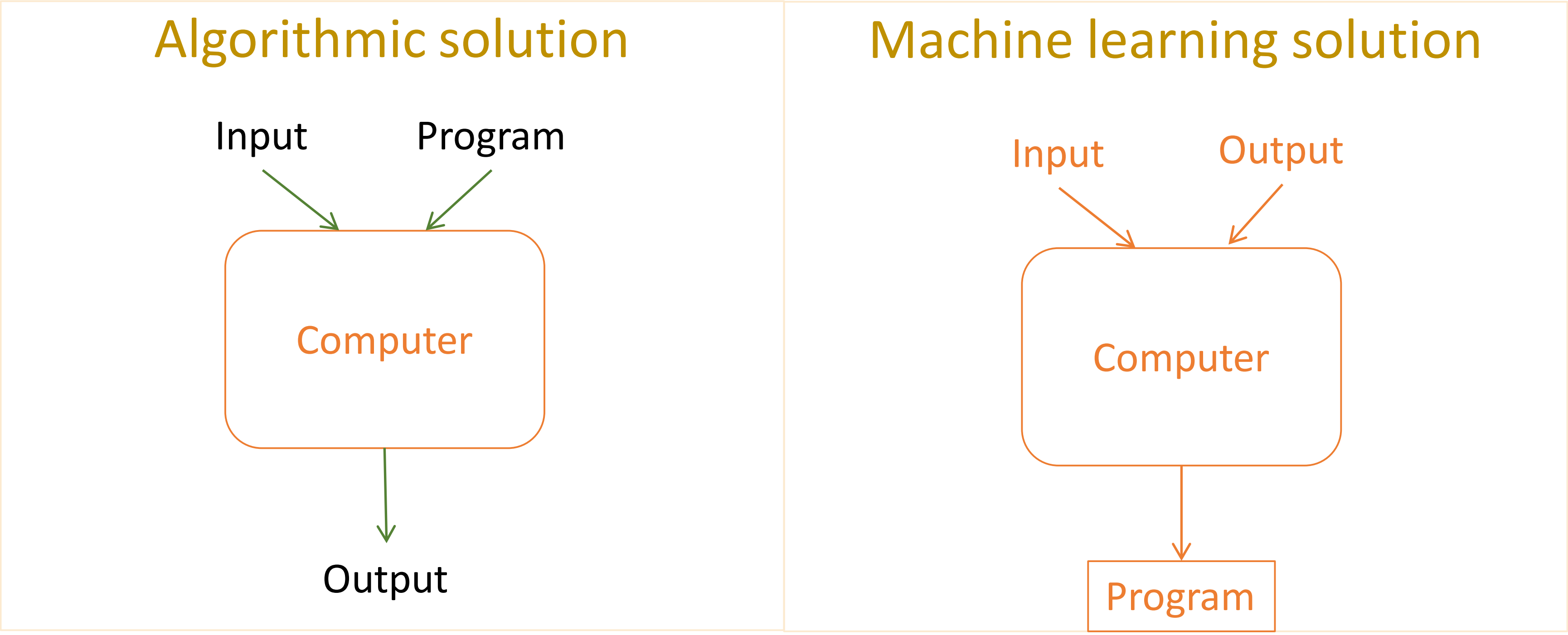
Data

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

- Decision Trees
- Support Vector Machine
- **Neural Networks**

The Machine Learning Solution

- Collect many examples that specify the correct output for a given input
- ML to get the mapping from input to output



Machine Learning : Definition

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

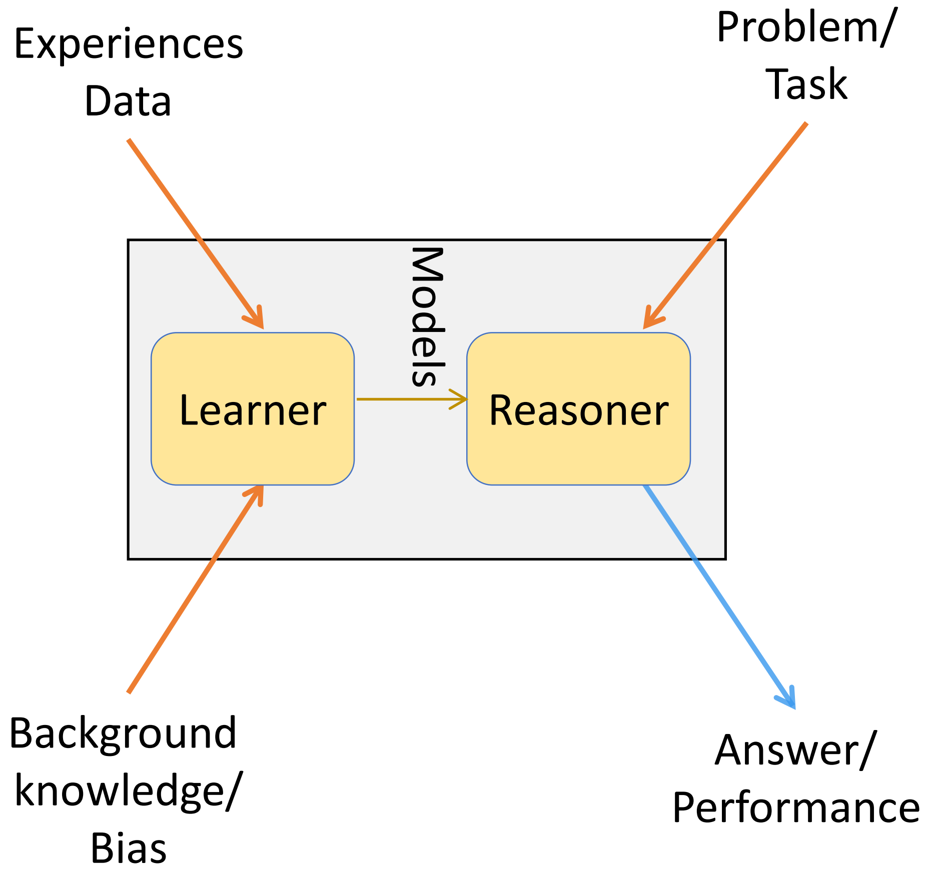
- Machine Learning explores algorithms that can
  - Learn from data such as build a model from data
  - Use the model or experience for prediction, decision making or solving some tasks

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, acquiring new, improved speed and efficiency

Designing a Learner

1. Choose the training experience
2. Choose the target function (that is to be learned)
3. Choose how to represent the target function
4. Choose a learning algorithm to infer the target function



Components of a ML application

Representation
- Features: Data specification
- Function class: Model form

Optimization
- Model Training

Evaluation
- Performance measure

1A. Representation of Data

1. How is the data specified?

   A. Features

   - Feature vector of *n* features

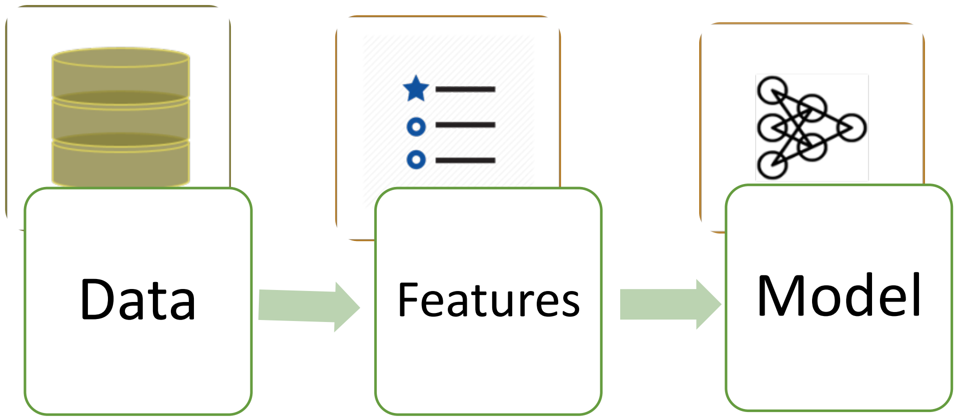
$$\bar{x} = (x_1, x_2, \ldots, x_n)$$

   B. Convert input to a vector of basis functions

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

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



1B. 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.

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

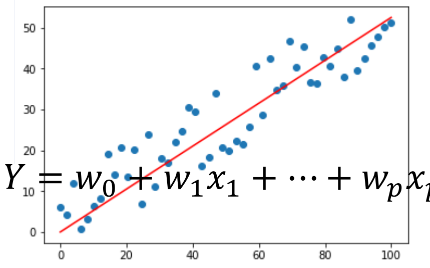
$$y = g(\bar{\phi}(\bar{x}))$$

- Linear function
- Decision Tree
- Graphical Model
- Neural Network

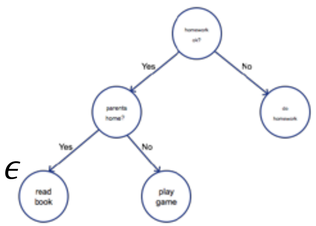
1B. Model Representation
Hypothesis space

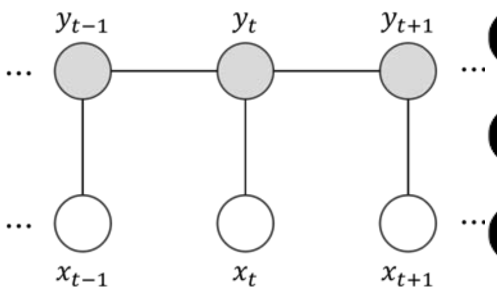
$y = f(\bar{x})$

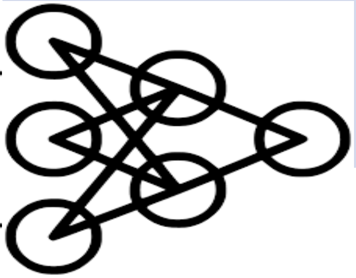
Linear Function	Decision Tree	Graphical Model	Neural Net



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







2. Evaluation

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

- Define loss function
- 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
- Semi-supervised Learning
  - Training Data + some Labels
- 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*

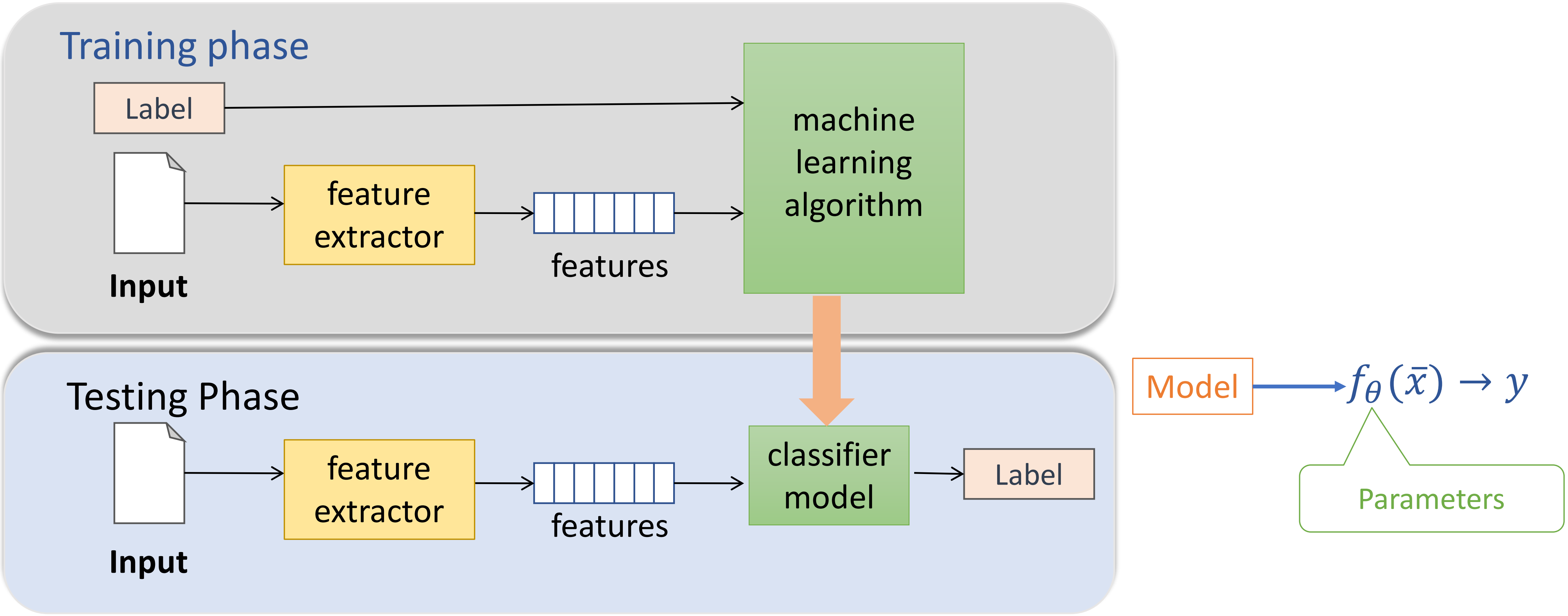
Model $\rightarrow$ $f_\theta(\bar{x}) \rightarrow y$

Parameters

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

Training: Learn the model from the Training Data

Given Test instance $\overline{x'}$, predict $y' = f_\theta(\overline{x'})$

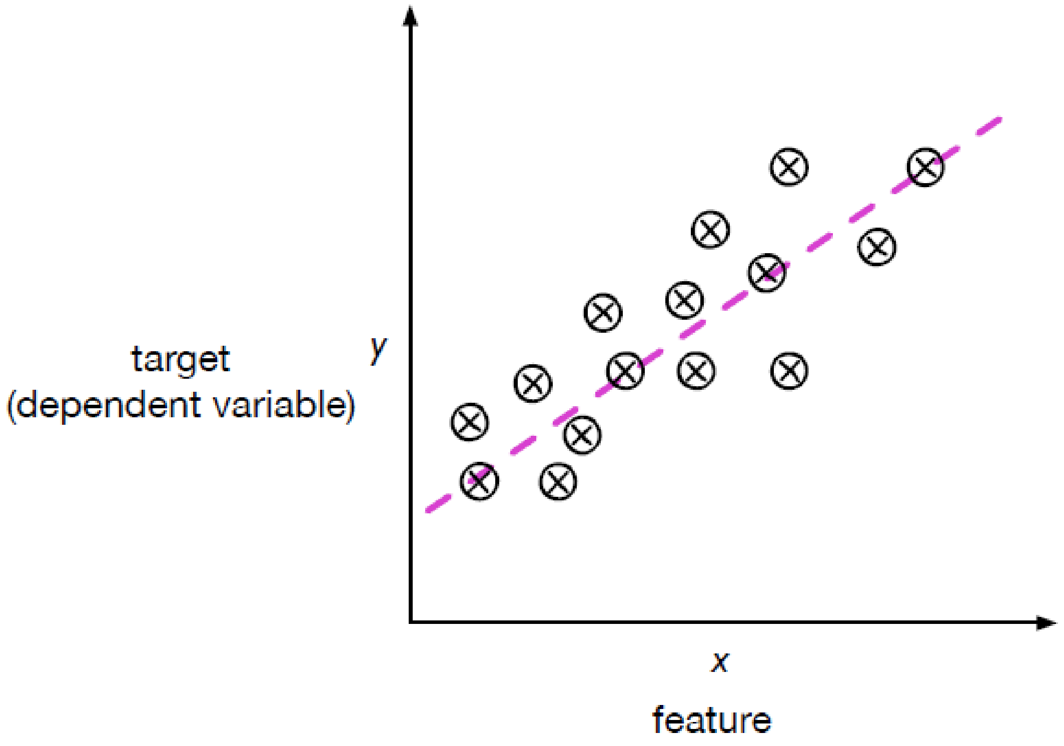


Training phase
Label
machine learning algorithm
feature extractor
features
Input
Testing Phase
feature extractor
features
classifier model
Label
Input
Model
$f_\theta(\bar{x}) \to y$
Parameters

Supervised Learning

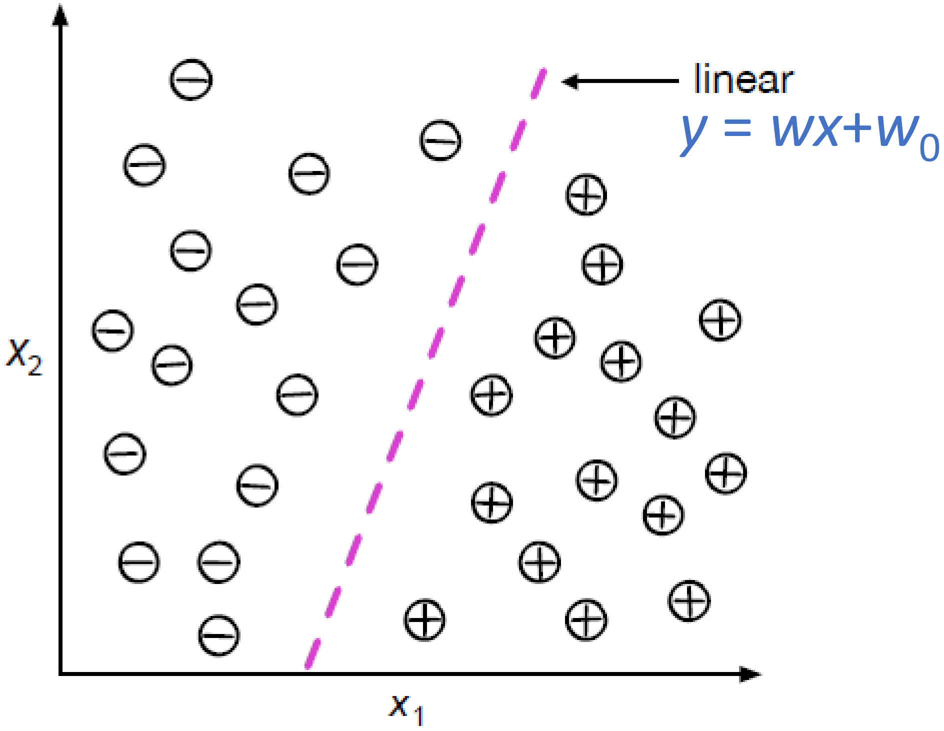
Classification

Y is categorical/ discrete



Regression

Y is numeric / continuous



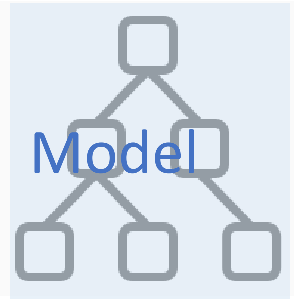
Supervised Learning

Classification Example

Training Samples

x1 (Ave sentence length)	**x2** (personal pronouns)	**…**	**X4** (mentions of slang)	**Category**
15	10		No	F
16	15		Yes	M
…				..
10	12		No	M

Train a model to minimize loss



Test Instances

x1 (Ave sentence length)	**x2 (personal pronouns)**	**…**	**X4 (mentions of slang)**	**Category**
12	10		No	?
18	15		No	?
15	15		Yes	?
9	12		No	?

Probabilistic Classification

x1 (Ave sentence length)	**x2** (personal pronouns)	**…**	**X4** (mentions of slang)	**Category**
15	10		No	SN
16	15		Yes	RK
…				..
10	12		No	PH

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

x1 (Ave sentence length)	**x2 (personal pronouns)**	**…**	**X4 (mentions of slang)**	**SN**	**RK**	**AZ**
12	10		No			
18	15		No			
15	15		Yes			
9	12		No			

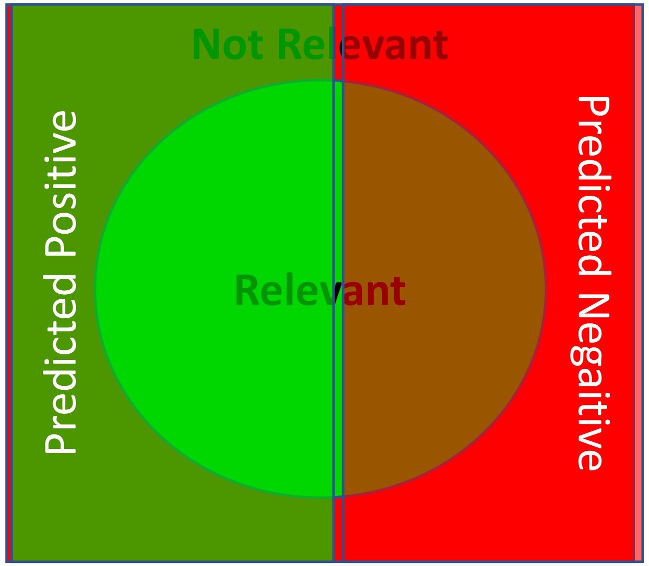
Evaluation for Classification problems

- Accuracy $= \dfrac{\text{\# correctly classified}}{\text{\# all test examples}}$

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

Precision $= \dfrac{\text{\# predicted true pos}}{\text{\# predicted pos}}$

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



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:

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

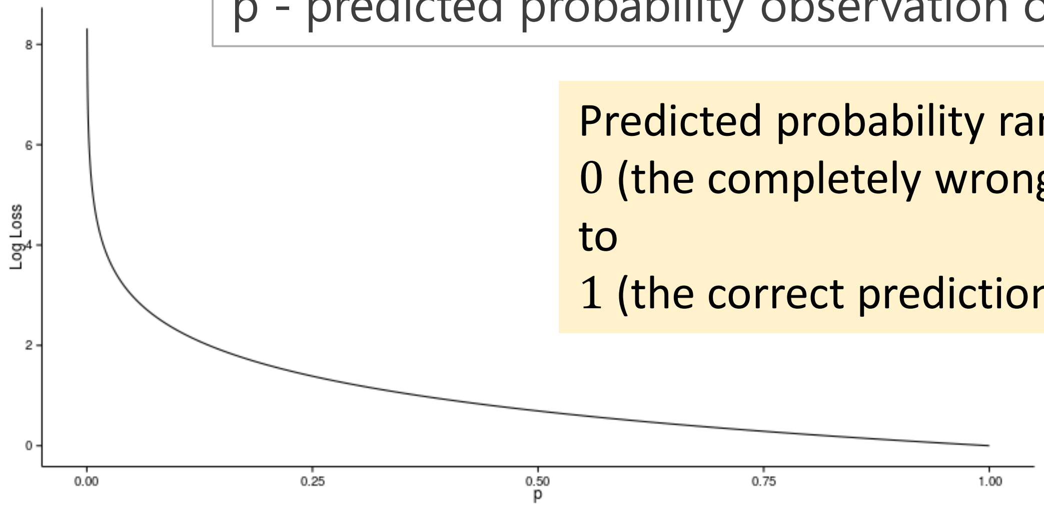
Logarithmic Loss Function

2. Logarithmic Loss:

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

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

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



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

2. Evaluation for regression problem

- Mean Squared error

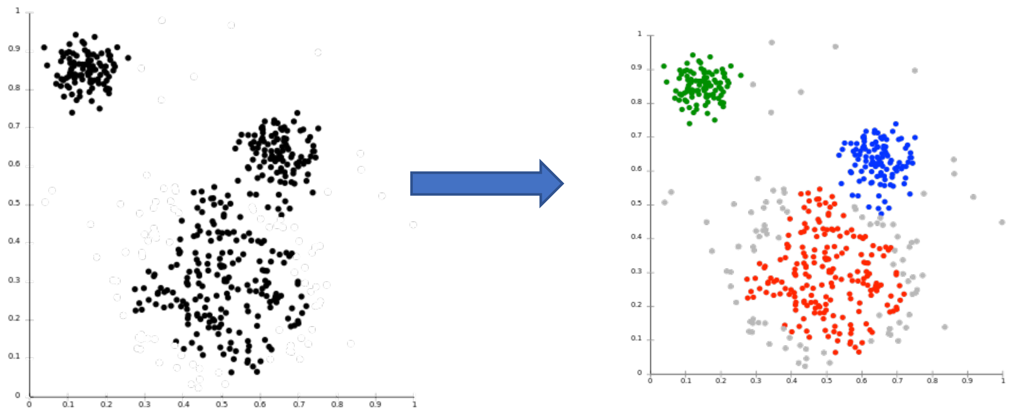
$$MSE = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2$$

Predicted value

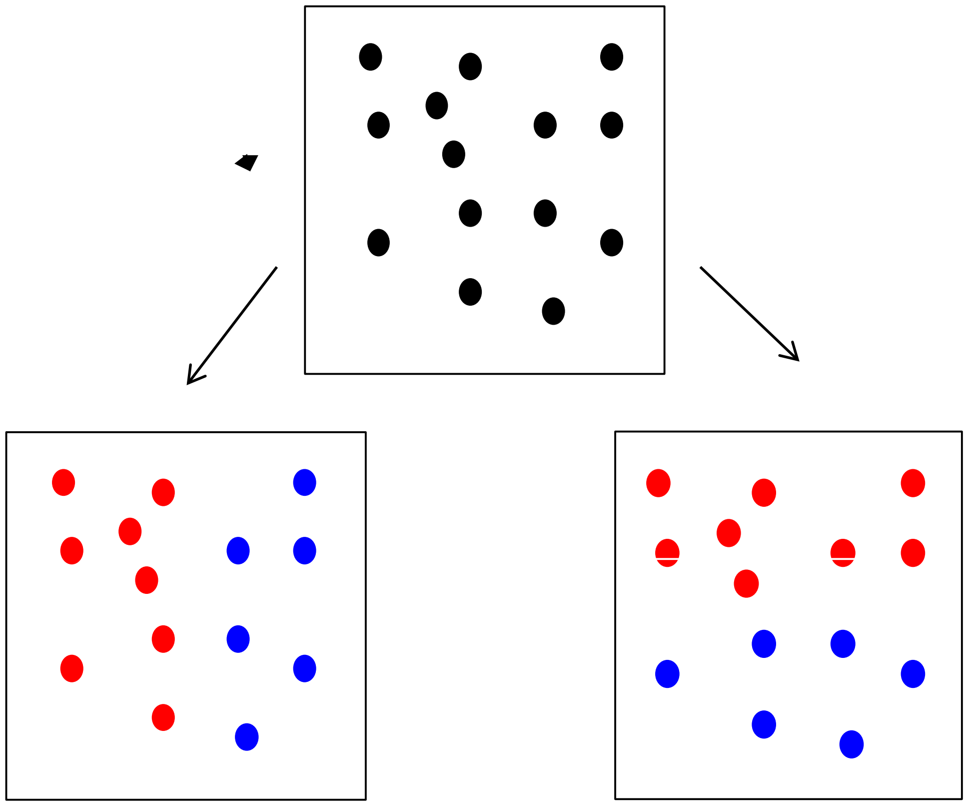
True value

Unsupervised Learning (Clustering)

- Given $\{\overline{x_1}, \overline{x_2}, \ldots \overline{x_m}, \}$ without labels
- Find hidden structure in the data
  - Clustering
  - Dimensionality Reduction
- Clustering: Grouping similar objects



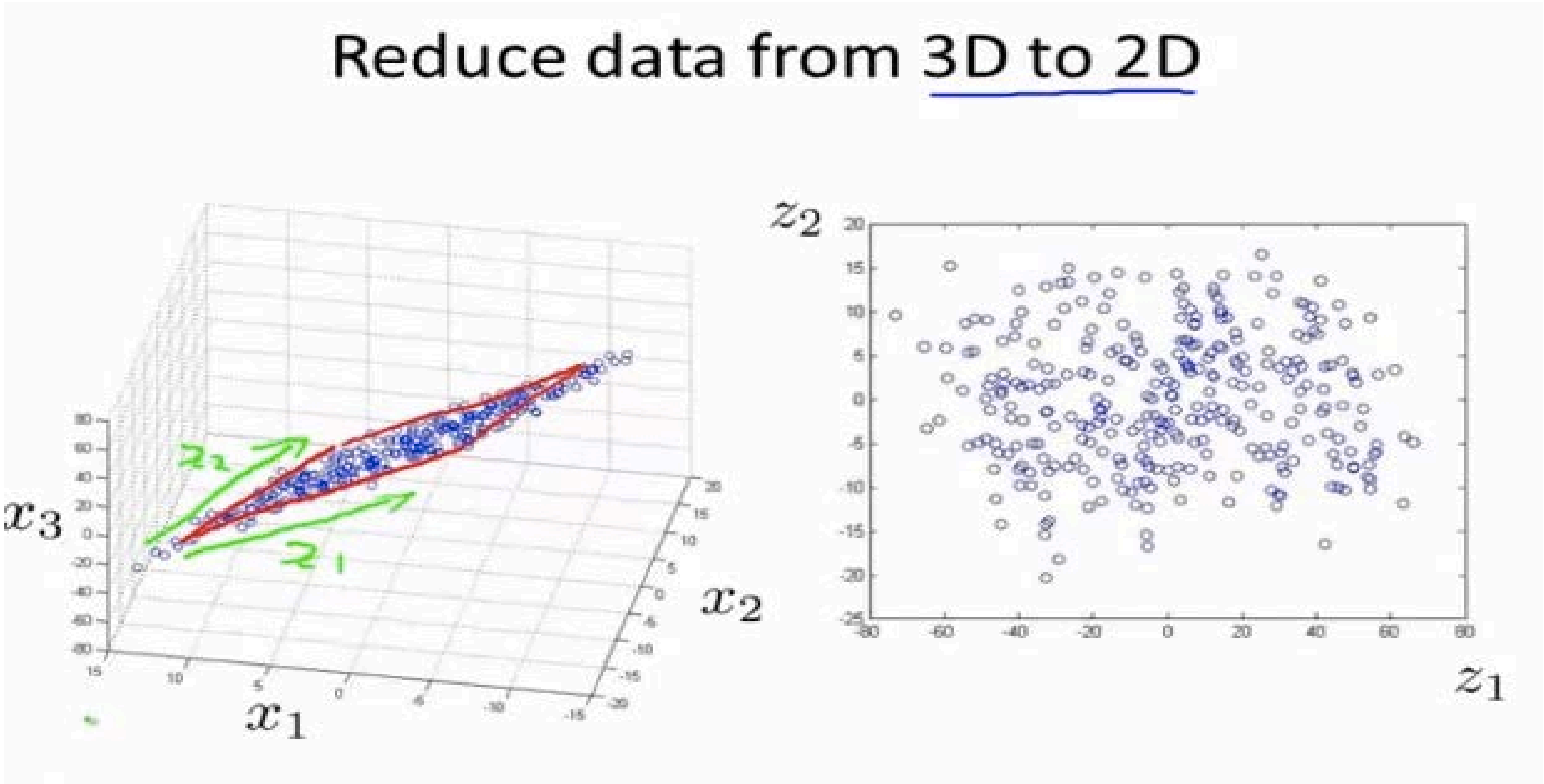
Clustering Problems



How to evaluate clustering?

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

Dimensionality Reduction



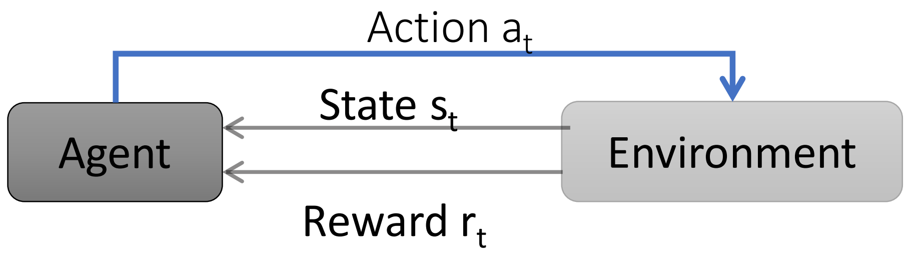
By Andrew Ng

Semi-Supervised Learning

- Supervised learning + Additional unlabeled data
- Unsupervised learning + Additional labeled data
- Learning Algorithm:
  - Start from the labeled data to build an initial classifier
  - Use the unlabeled data to enhance the model

Reinforcement Learning

- Given a sequence of states and actions with (delayed) rewards, output a policy.



  - Receive feedback in the form of rewards
  - Agent's utility is defined by the reward function
  - Must (learn to) act so as to maximize expected rewards

- Examples:
  - Dialog systems
  - Information retrieval
  - Personalized recommendation

Goal: Constantly learn to make 'optimal' predictions based on real-time feedback from past predictions