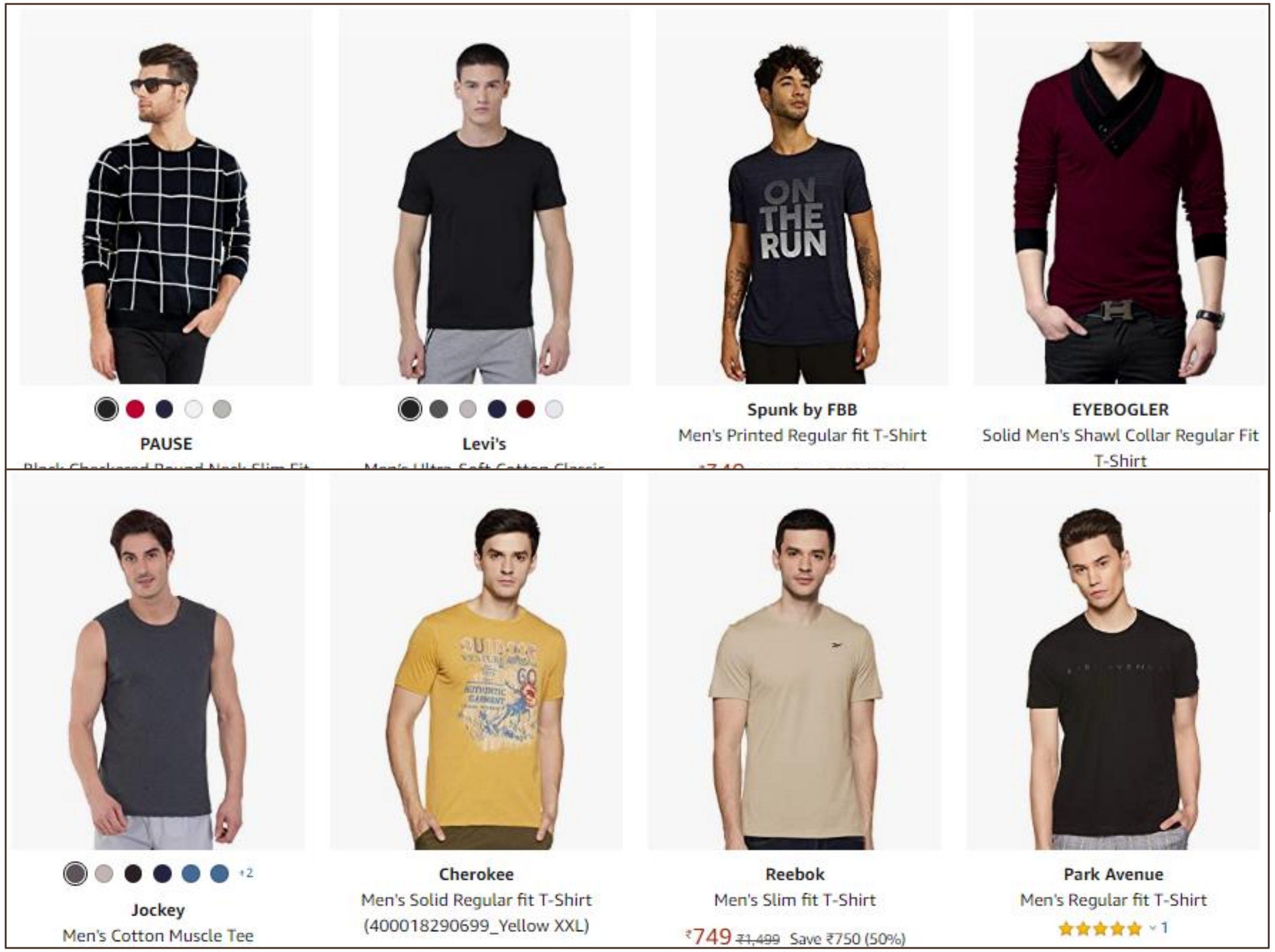
Introduction to Machine Learning and Features

Machine Learning Unit 2

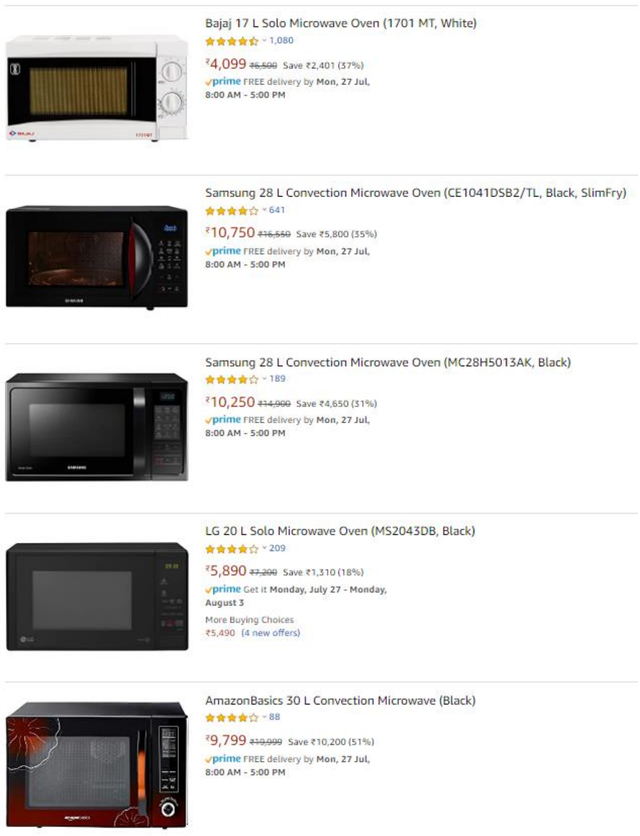
Sudeshna Sarkar

Recognize the type of sleeve/ pattern in the image of a shirt

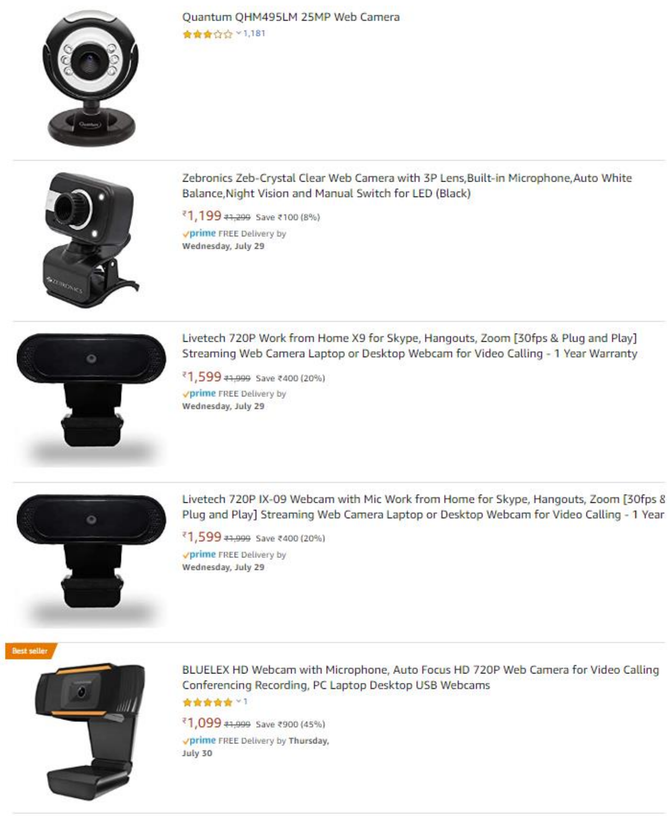


Predict the average rating of a new microwave

Predict the average rating of a new microwave

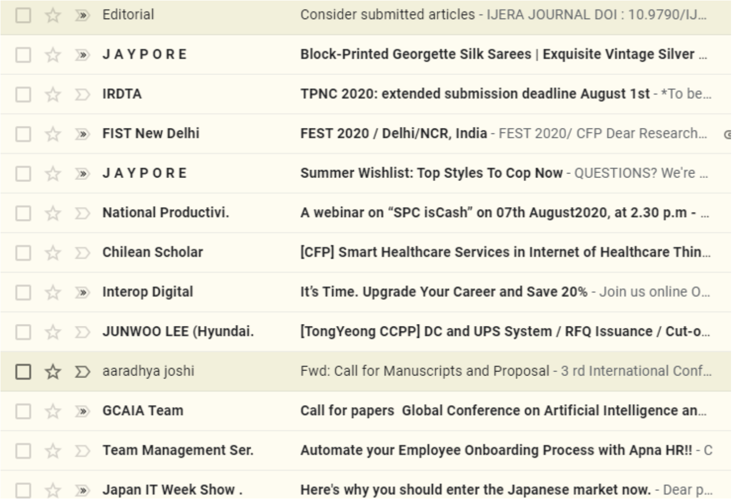


Predict the number of purchases of a webcam

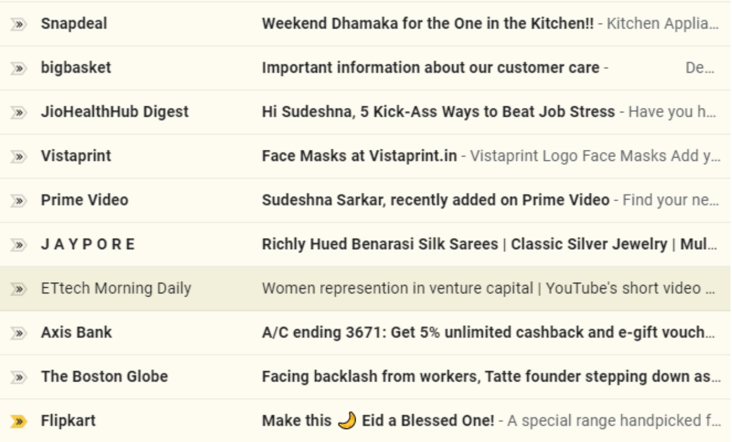


Classify email

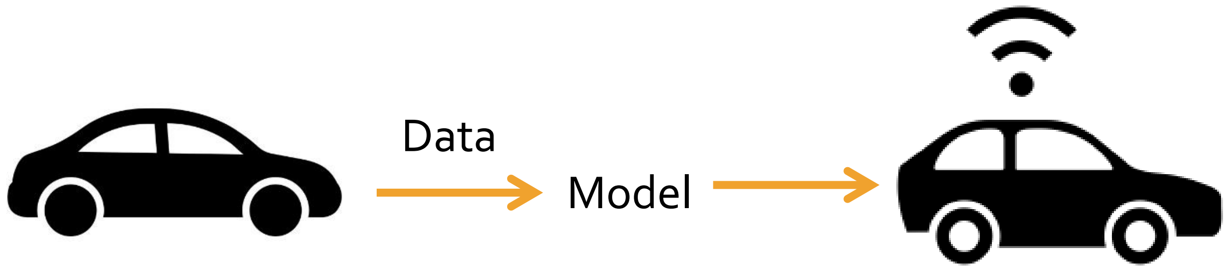
SPAM



PROMOTION



Autonomous Driving

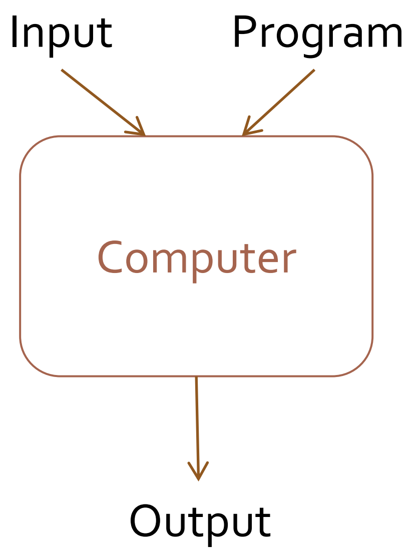


Sensors:
Camera
Radar
Lidar

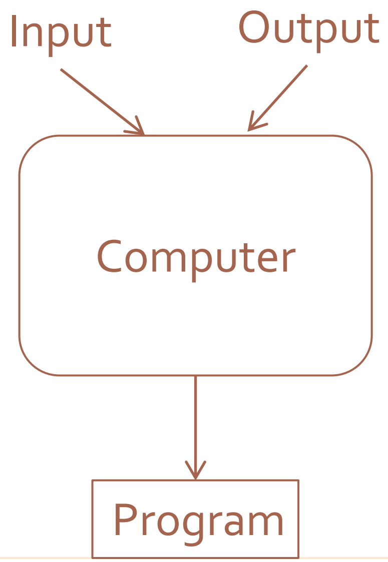
The Machine Learning Solution

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

Algorithmic solution



Machine learning solution



Machine Learning : Definition

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

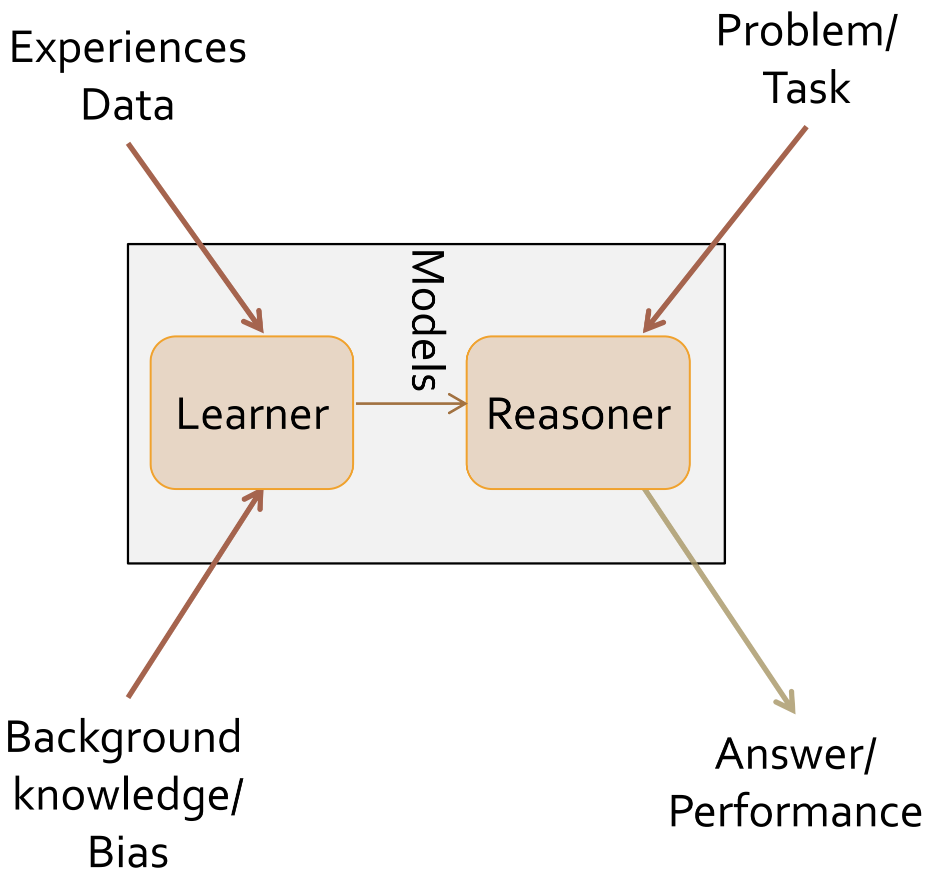
- Machine Learning explores algorithms that can
  - Learn from data such as build a model from data
  - Use the model or experience for prediction, decision making or solving some tasks

Components of a learning problem

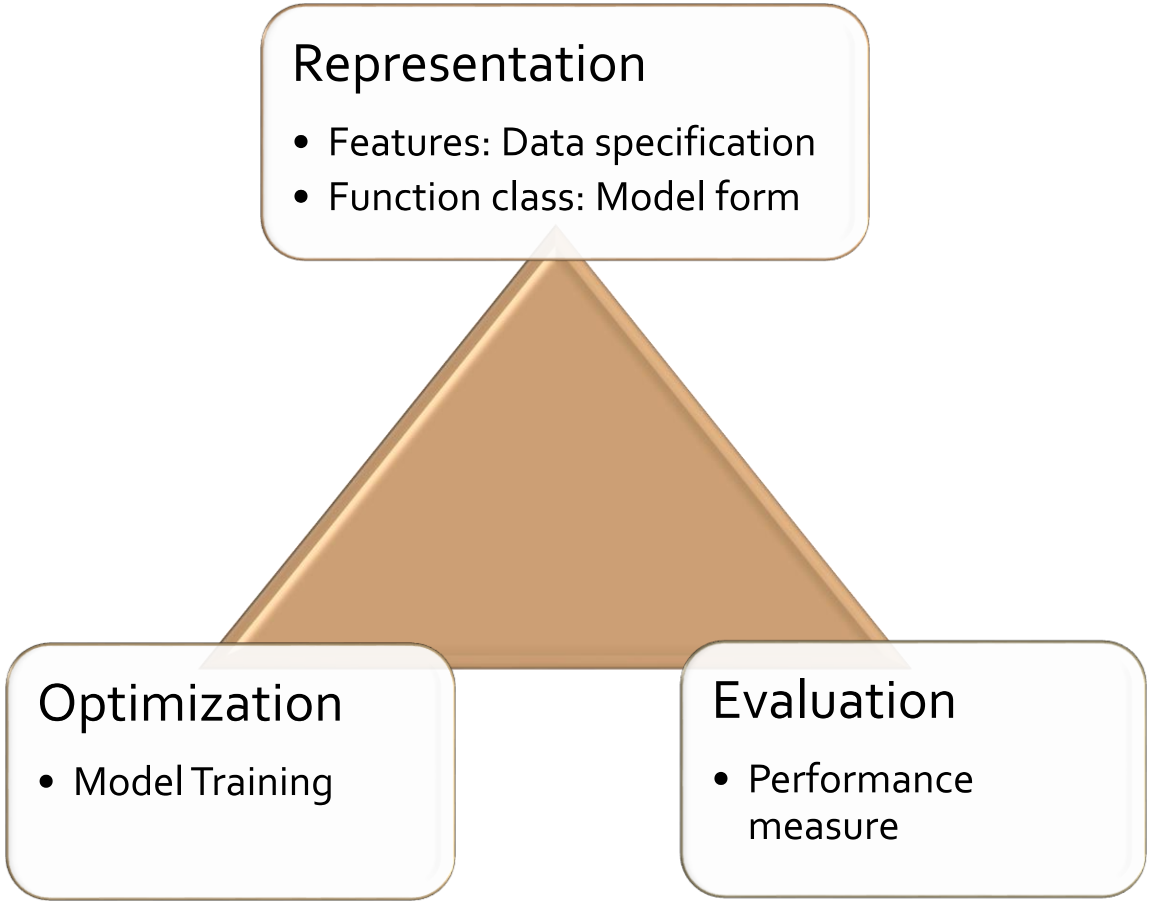
- Task: The behaviour or task being improved.
  - For example: classification, acting in an environment
- Data: The experiences that are being used to improve performance in the task.
- Measure of improvement :
  - For example: increasing accuracy in prediction, acquiring new, improved speed and efficiency

Design 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



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

1. A microwave

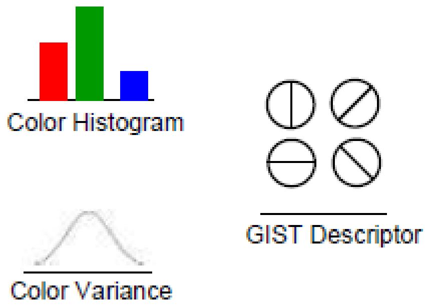
   Attributes:
   - Volume: 17 l, 23 l, …
   - Functions: Micro, Cor
   - Power level
   - Accessories
   - Type of dial
   - Brand
   - Warranty
   - Price

1. Image of shirt
   - Collar style
   - Sleeve type
   - Colour
   - …

Features

Image classification

- Raw pixels
- Histograms
- GIST descriptors



Product Rating (Webcam)

- Frame rate
- Resolution
- Autofocus
- Microphone
- Lens
- Brand

Bank Marketing Dataset

http://archive.ics.uci.edu/ml/datasets/Bank+Marketing

Predict if the client will subscribe (yes/no) a term deposit (variable y).

Input variables:

 # bank client data:

1. age
2. type of job
3. marital status
4. education
5. has credit in default?
6. has housing loan?
7. has personal loan?

 # related with the last contact of the current campaign:

8. contact communication type ('cellular','telephone')
9. last contact date and duration

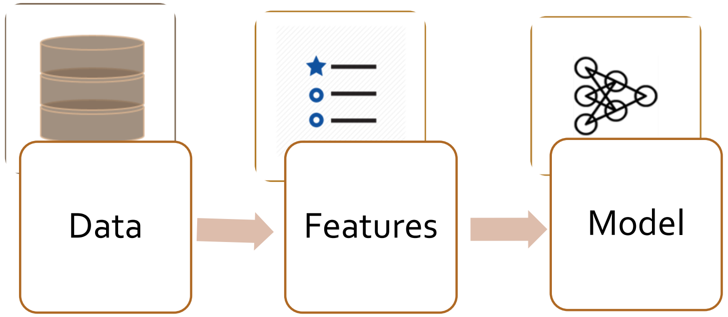
 # other attributes:

12. No of contacts performed for this client
13. No of days after client last contacted
14. No of contacts performed before this campaign and for this client
15. outcome of prev marketing campaign

 # social and economic context attributes

16. employment variation rate - quarterly indicator
17. consumer price index - monthly indicator
18. consumer confidence index - monthly indicator
19. euribor 3 month rate - daily indicator
20. number of employees - quarterly indicator

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

Feature Engineering

Transforming raw data into features that better represent the underlying problem

- Feature Selection
- Feature Extraction
- Missing feature values
- Feature value normalization
- Aggregate Feature values
- Feature Encoding

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.

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

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

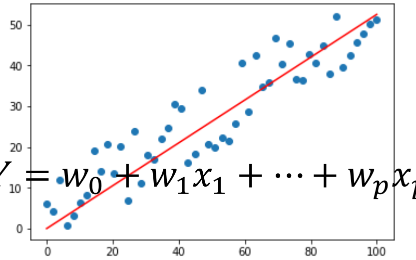
- Linear function
- Decision Tree
- Graphical Model
- Neural Network

1B. Model Representation

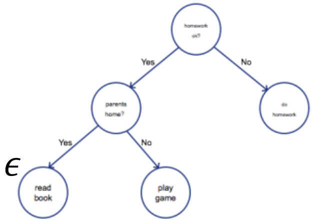
Hypothesis space

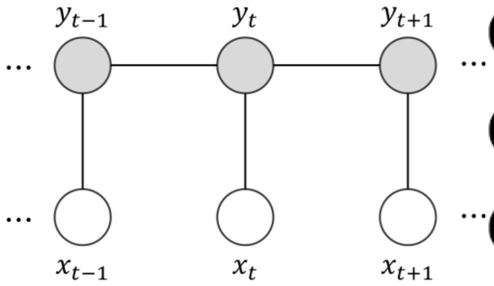
$$y = f(\bar{x})$$

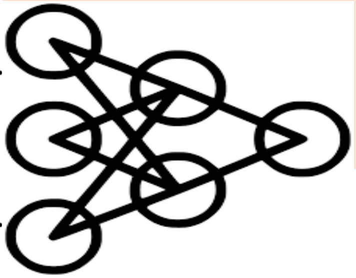
Linear Function	Decision Tree	Graphical Model	Neural Net



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







2. Evaluation

1. Accuracy $= \dfrac{\text{\# correctly classified}}{\text{\# all test examples}}$
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
$$MSE = \frac{1}{m} \sum (y_{pred} - y_{true})^2$$

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

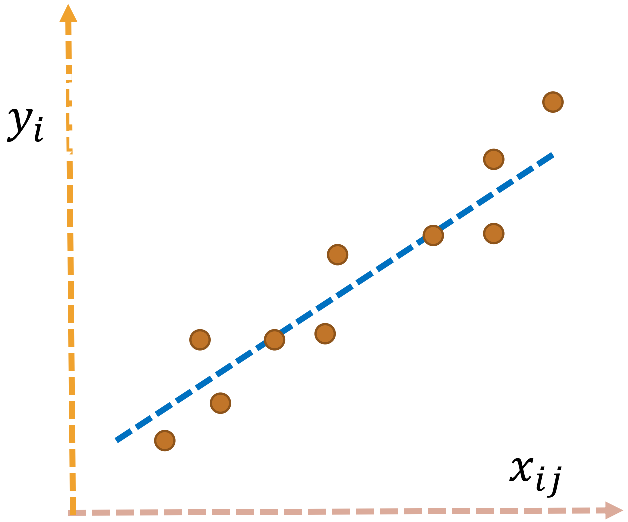
Elements of Optimization

1. Variables
2. Constraints
3. Objective Function

Simple Linear Regression

1. Variables: $w_0, w_1, \dots, w_n$
2. Constraints: none
3. Objective Function: Minimize

$$\sum_{i=1}^{m}\left(y_i\left(w_0+\sum_{j=1}^{n} w_j x_{ij}\right)\right)^2$$



- m data points, n features
  - x_{ij}: jth attribute of ith instance
  - y_i: output of ith instance

Find coefficients $w_0, w_1, \dots, w_n$ to best fit data

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:

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

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

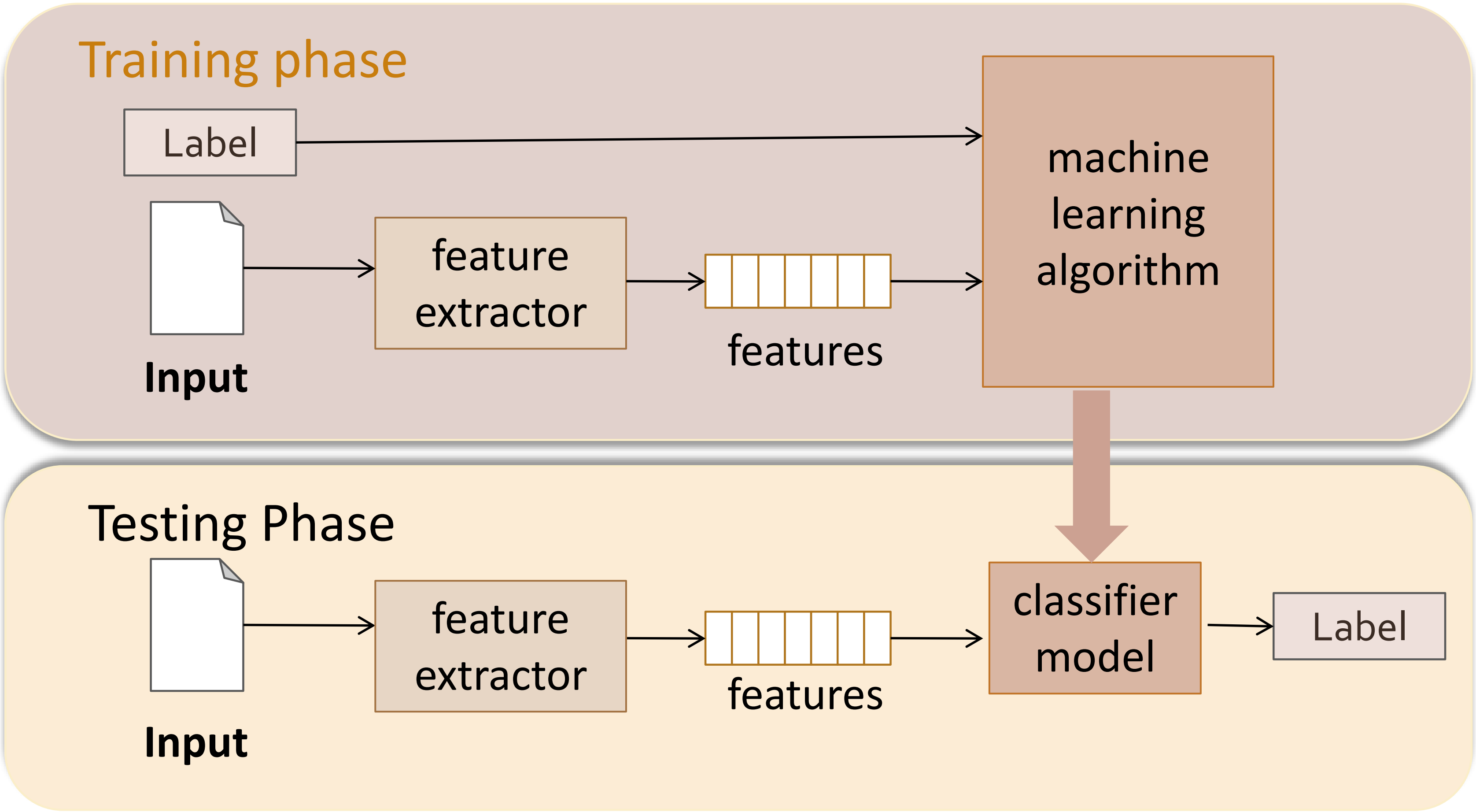
$\bar{X}$	Y
$\overline{x_1}$	y_1
$\overline{x_2}$	y_2
…	..
$\overline{x_m}$	y_m

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

Parameters

Training: Learn the model from the Training Data

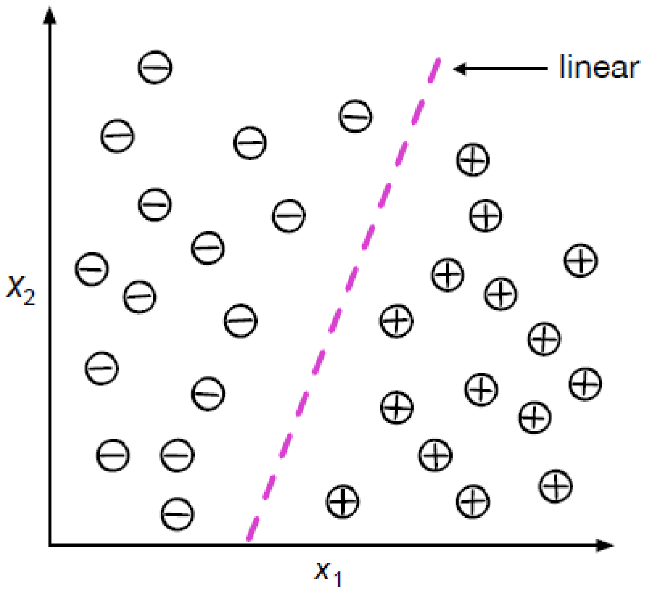
Given Test instance $\overline{x'}$, predict $y' = f_\theta(\overline{x'})$


Training phase
Label
machine learning algorithm
feature extractor
features
Input
Testing Phase
feature extractor
features
classifier model
Label
Input

Supervised Learning

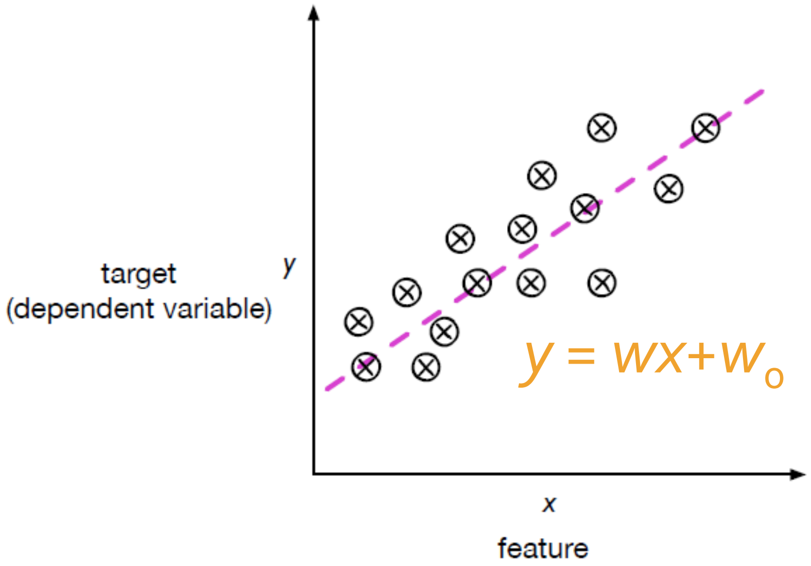
Classification

Y is categorical/ discrete



Regression

Y is numeric / continuous



Example Tasks

Classification

- Object identification from images
- Email classification
- Whether stock price will be up or down next month

Regression

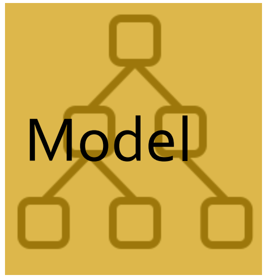
- House price prediction
- Stock price prediction
- Precipitation forecasting
- What is the probability that the user will buy this phone?

Supervised Learning
Classification Example

Training Samples

Color	Sleeve	Pattern	Collar	Like
Red	Half	Printed	No	Y
White	Half	Solid	Yes	N
…				..
White	Full	Solid	Yes	Y

Train a model to minimize loss



Test Instances

Color	Sleeve	Pattern	Collar	Like
Blue	Full	Printed	No	?
White	Half	Solid	No	?
Red	Half	Stripe	Yes	?
Red	Full	Solid	Yes	?

Probabilistic Classification

Color	Sleeve	Pattern	Collar	Category
Red	Half	Printed	No	Casual
White	Half	Solid	Yes	Lounge
...				..
White	Full	Solid	Yes	Formal

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

Color	Sleeve	Pattern	Collar	Formal	Casual	Party
Blue	Full	Printed	No	0.4	0.15	0.45
White	Half	Solid	No	0.2	0.7	0.1
Red	Half	Stripe	Yes	0.1	0.2	0.7
Red	Full	Solid	Yes	0.5	0.1	0.4

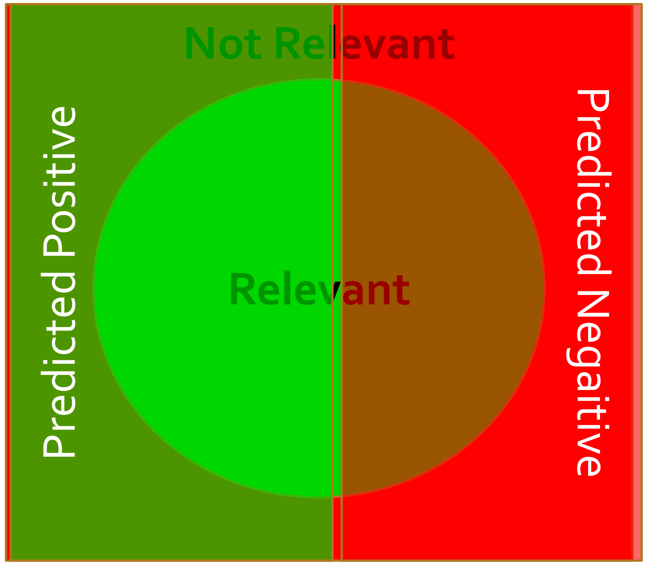
Evaluation for Classification problems

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

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

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

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



Loss Function
Classification problems

Loss indicates how bad the model's prediction is.

1. Fraction of Misclassifications

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

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

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

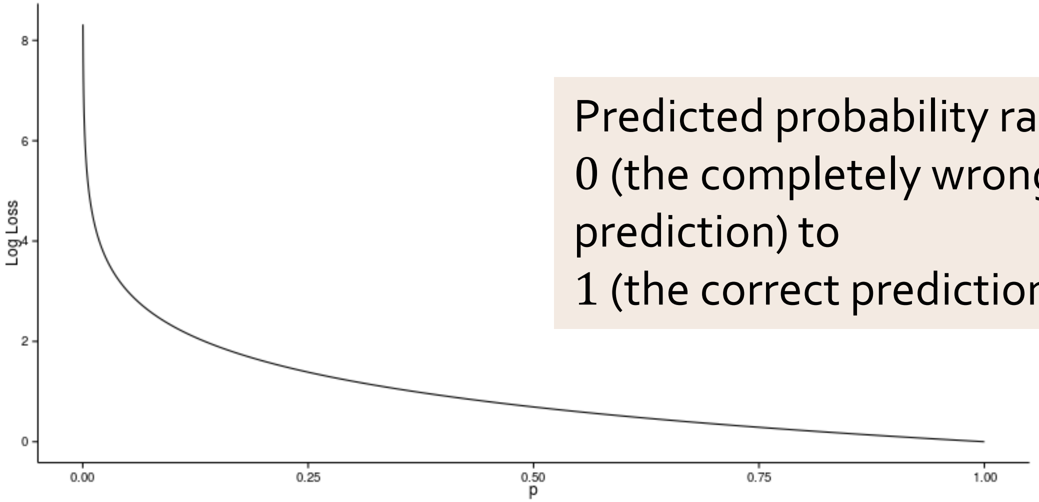
Logarithmic Loss Function

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

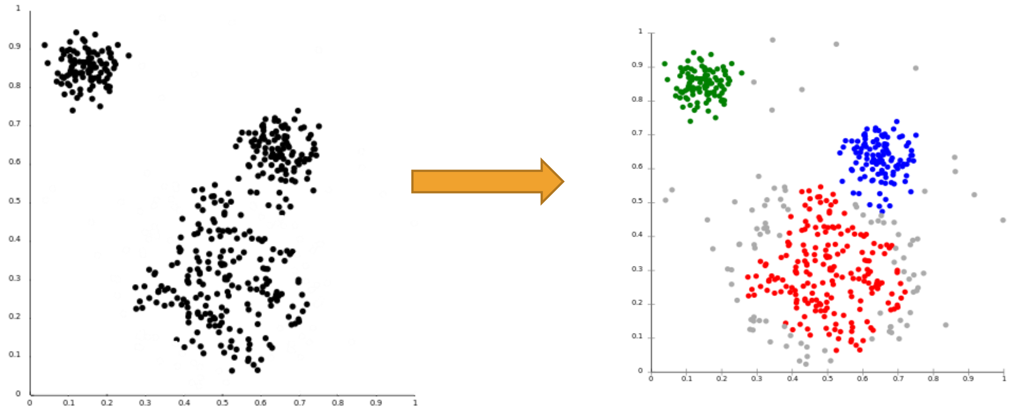
$$MSE = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)^2$$

Predicted value

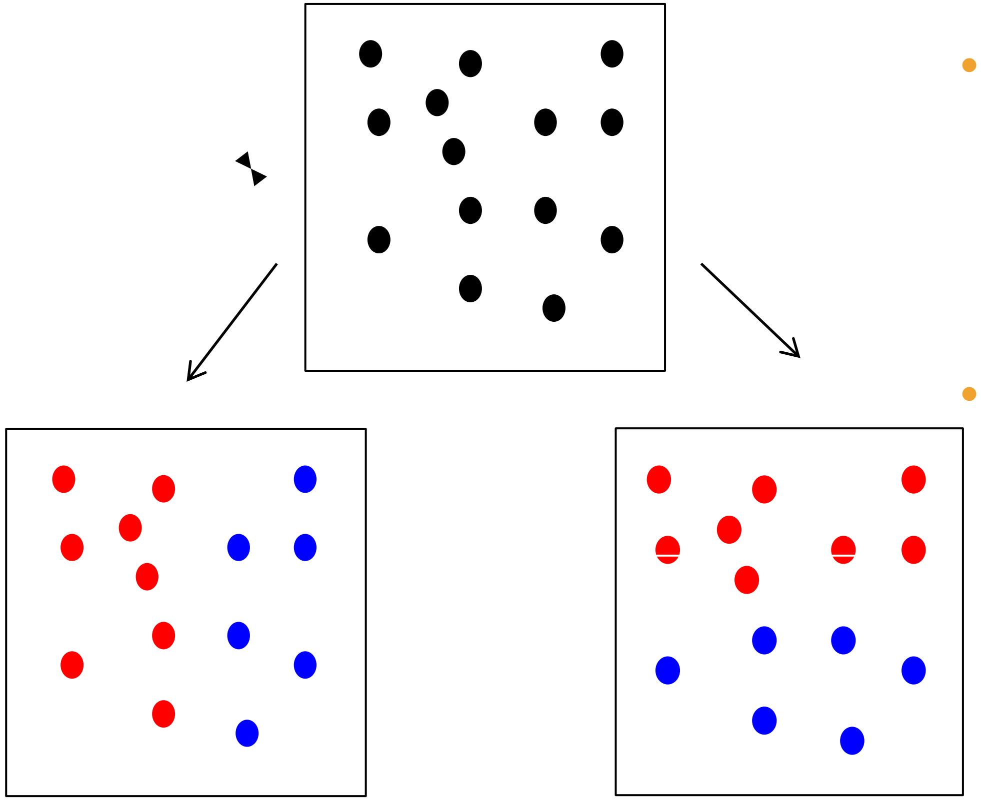
True value

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

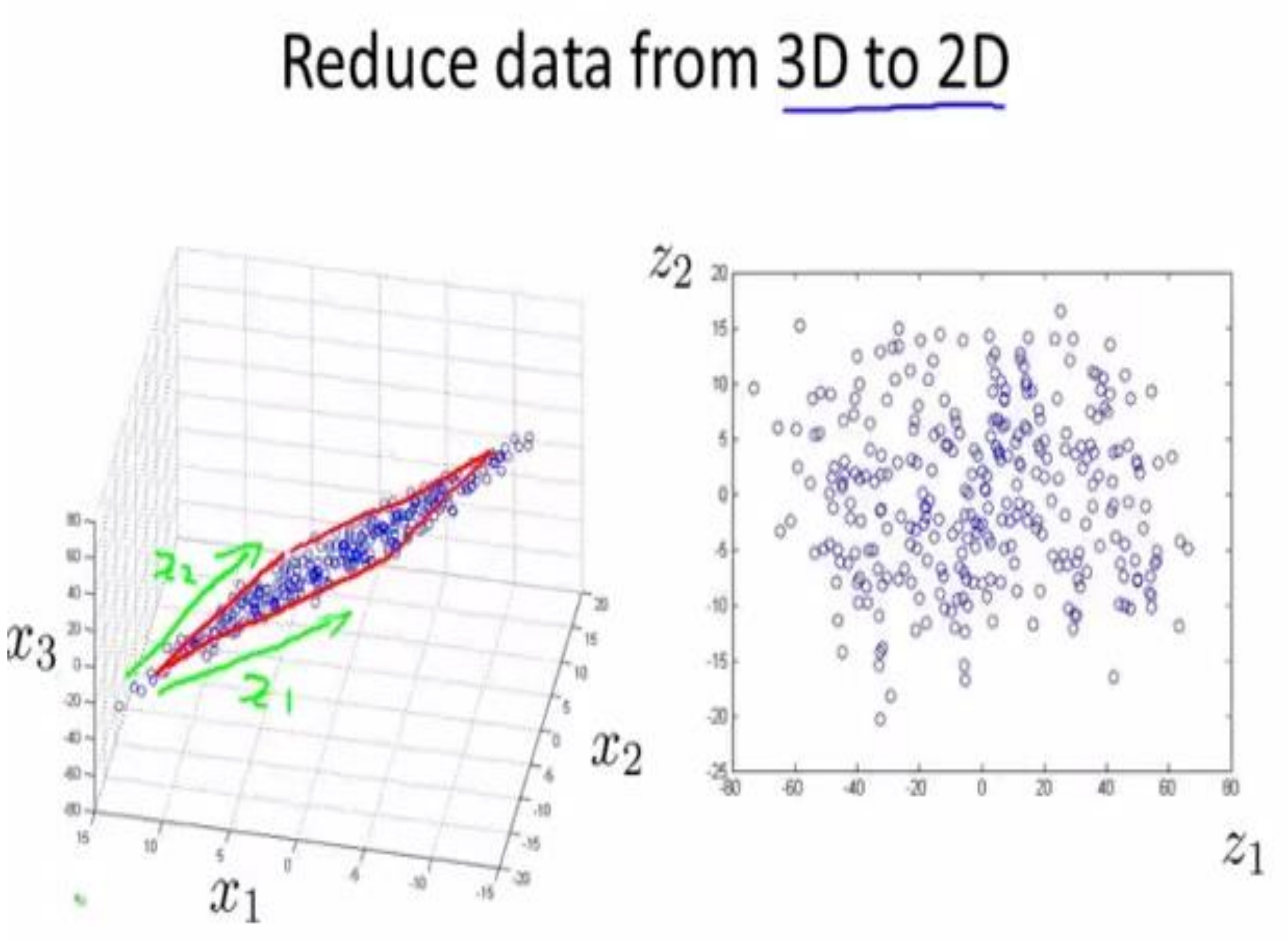


Clustering Problems



- How to evaluate clustering?
- Internal Evaluation:
  - Intra-cluster distances are minimized
  - Inter-cluster distances are maximized
- External Evaluation

Dimensionality Reduction



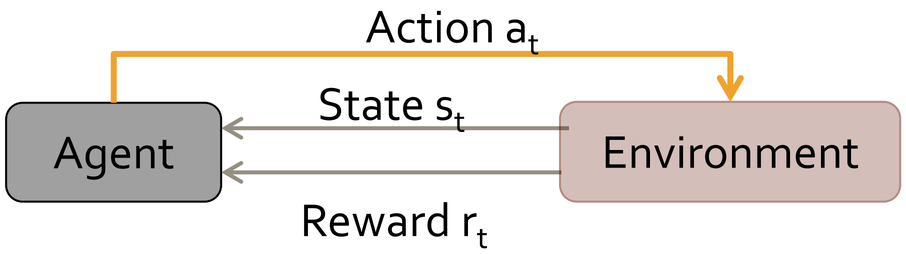
By Andrew Ng

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:
  - Game playing (Go)
  - Robot grasping
  - Controlling aircraft and robotic motion
  - Dynamic Pricing
  - Personalized recommendation

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

Why Machine Learning?

- Human expertise not sacrosanct
  - Pricing, Promotions
- Human expertise cannot be coded
  - face/handwriting/speech recognition
  - driving a car, flying a plane
- Rapidly changing situation
  - Credit scoring, financial modeling
  - Fraud detection
- Need for personalization
  - Product recommendation
- Too much Data