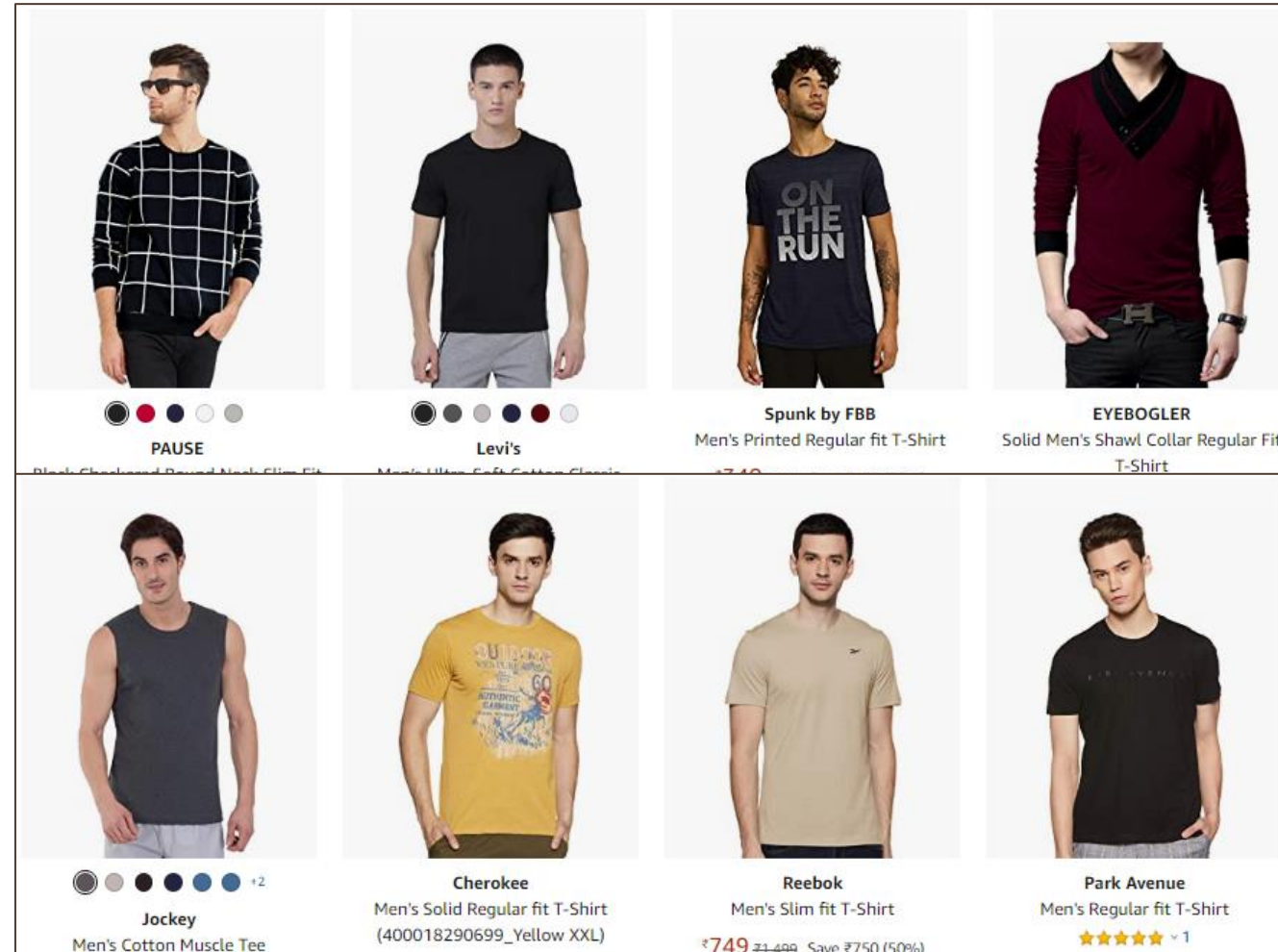


# Introduction to Machine Learning and Features

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

	<p>Bajaj 17 L Solo Microwave Oven (1701 MT, White)</p> <p>★★★★☆ ~ 1,080</p> <p>₹4,099 <del>₹6,500</del> Save ₹2,401 (37%)</p> <p>✓prime FREE delivery by Mon, 27 Jul, 8:00 AM - 5:00 PM</p>
	<p>Samsung 28 L Convection Microwave Oven (CE1041DSB2/TL, Black, SlimFry)</p> <p>★★★★☆ ~ 641</p> <p>₹10,750 <del>₹16,550</del> Save ₹5,800 (35%)</p> <p>✓prime FREE delivery by Mon, 27 Jul, 8:00 AM - 5:00 PM</p>
	<p>Samsung 28 L Convection Microwave Oven (MC28H5013AK, Black)</p> <p>★★★★☆ ~ 189</p> <p>₹10,250 <del>₹14,900</del> Save ₹4,650 (31%)</p> <p>✓prime FREE delivery by Mon, 27 Jul, 8:00 AM - 5:00 PM</p>
	<p>LG 20 L Solo Microwave Oven (MS2043DB, Black)</p> <p>★★★★☆ ~ 209</p> <p>₹5,890 <del>₹6,200</del> Save ₹3,310 (18%)</p> <p>✓prime Get it Monday, July 27 - Monday, August 3</p> <p>More Buying Choices ₹5,490 (4 new offers)</p>
	<p>AmazonBasics 30 L Convection Microwave (Black)</p> <p>★★★★☆ ~ 88</p> <p>₹9,799 <del>₹10,999</del> Save ₹10,200 (51%)</p> <p>✓prime FREE delivery by Mon, 27 Jul, 8:00 AM - 5:00 PM</p>

## Predict the number of purchases of a webcam

	<p>Quantum QHM495LM 25MP Web Camera</p> <p>★★★★☆ ~ 1,181</p>
	<p>Zebtronics Zeb-Crystal Clear Web Camera with 3P Lens,Built-in Microphone,Auto White Balance,Night Vision and Manual Switch for LED (Black)</p> <p>₹1,199 <del>₹1,299</del> Save ₹100 (8%)</p> <p>✓prime FREE delivery by Wednesday, July 29</p>
	<p>Livetech 720P Work from Home X9 for Skype, Hangouts, Zoom [30fps &amp; Plug and Play] Streaming Web Camera Laptop or Desktop Webcam for Video Calling - 1 Year Warranty</p> <p>₹1,599 <del>₹2,099</del> Save ₹400 (20%)</p> <p>✓prime FREE delivery by Wednesday, July 29</p>
	<p>Livetech 720P IX-09 Webcam with Mic Work from Home for Skype, Hangouts, Zoom [30fps &amp; Plug and Play] Streaming Web Camera Laptop or Desktop Webcam for Video Calling - 1 Year</p> <p>₹1,599 <del>₹2,099</del> Save ₹400 (20%)</p> <p>✓prime FREE delivery by Wednesday, July 29</p>
	<p>BLUELEX HD Webcam with Microphone, Auto Focus HD 720P Web Camera for Video Calling Conferencing Recording, PC Laptop Desktop USB Webcams</p> <p>★★★★☆ ~ 1</p> <p>₹1,099 <del>₹1,499</del> Save ₹900 (45%)</p> <p>✓prime FREE delivery by Thursday, July 30</p>

# Classify email

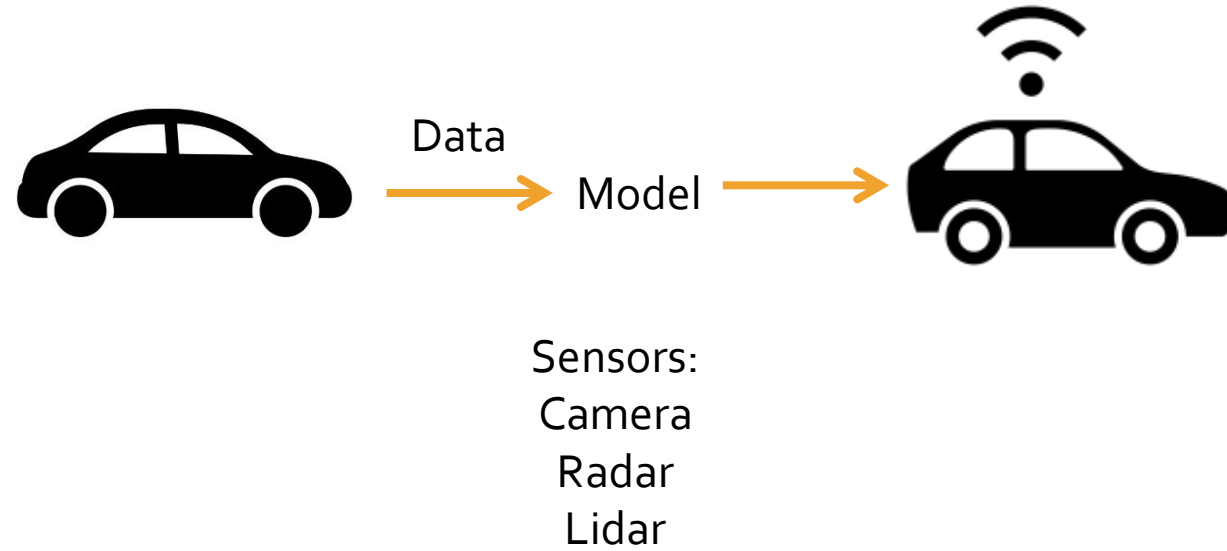
## SPAM

<input type="checkbox"/>	☆	»	Editorial	Consider submitted articles - IJERA JOURNAL DOI : 10.9790/IJ...
<input type="checkbox"/>	☆	»	J A Y P O R E	Block-Printed Georgette Silk Sarees   Exquisite Vintage Silver ...
<input type="checkbox"/>	☆	»	IRDTA	TPNC 2020: extended submission deadline August 1st - *To be...
<input type="checkbox"/>	☆	»	FIST New Delhi	FEST 2020 / Delhi/NCR, India - FEST 2020/ CFP Dear Research...
<input type="checkbox"/>	☆	»	J A Y P O R E	Summer Wishlist: Top Styles To Cop Now - QUESTIONS? We're ...
<input type="checkbox"/>	☆	»	National Productivi.	A webinar on "SPC isCash" on 07th August2020, at 2.30 p.m - ...
<input type="checkbox"/>	☆	»	Chilean Scholar	[CFP] Smart Healthcare Services in Internet of Healthcare Thin...
<input type="checkbox"/>	☆	»	Interop Digital	It's Time. Upgrade Your Career and Save 20% - Join us online O...
<input type="checkbox"/>	☆	»	JUNWOO LEE (Hyundai.	[TongYeong CCPP] DC and UPS System / RFQ Issuance / Cut-o...
<input type="checkbox"/>	☆	»	aaradhy joshi	Fwd: Call for Manuscripts and Proposal - 3 rd International Conf...
<input type="checkbox"/>	☆	»	GCAIA Team	Call for papers Global Conference on Artificial Intelligence an...
<input type="checkbox"/>	☆	»	Team Management Ser.	Automate your Employee Onboarding Process with Apna HR!! - C
<input type="checkbox"/>	☆	»	Japan IT Week Show .	Here's why you should enter the Japanese market now. - Dear p...

## PROMOTION

»	Snapdeal	Weekend Dhamaka for the One in the Kitchen!! - Kitchen Applia...
»	bigbasket	Important information about our customer care - De...
»	JioHealthHub Digest	Hi Sudeshna, 5 Kick-Ass Ways to Beat Job Stress - Have you h...
»	Vistaprint	Face Masks at Vistaprint.in - Vistaprint Logo Face Masks Add y...
»	Prime Video	Sudeshna Sarkar, recently added on Prime Video - Find your ne...
»	J A Y P O R E	Richly Hued Benarasi Silk Sarees   Classic Silver Jewelry   Mul...
»	ETtech Morning Daily	Women representation in venture capital   YouTube's short video ...
»	Axis Bank	A/C ending 3671: Get 5% unlimited cashback and e-gift vouch...
»	The Boston Globe	Facing backlash from workers, Tattle founder stepping down as...
»	Flipkart	Make this 🌸 Eid a Blessed One! - A special range handpicked f...

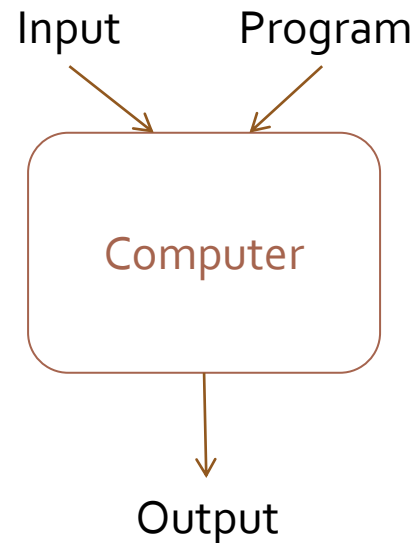
# Autonomous Driving



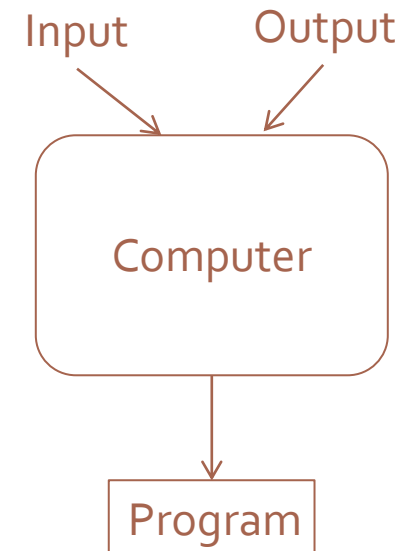
# The Machine Learning Solution

- 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

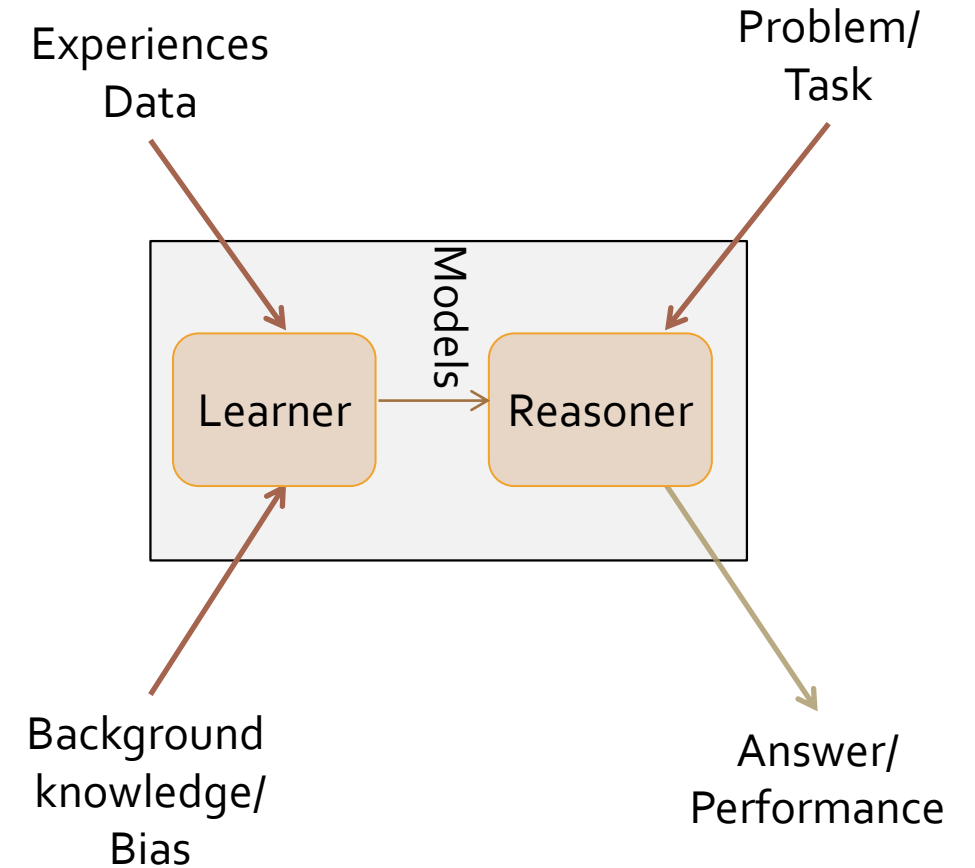
# Components of a learning problem

- **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

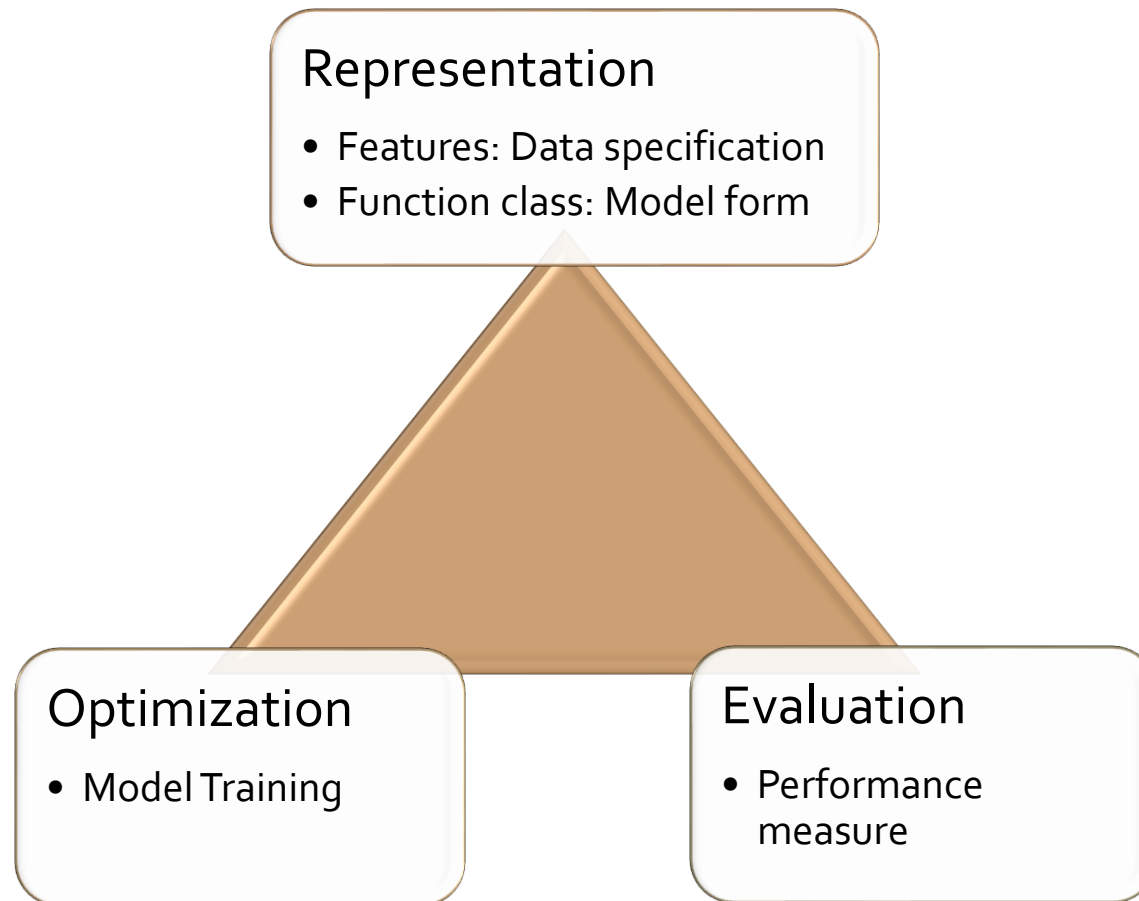


# Design a Learner

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

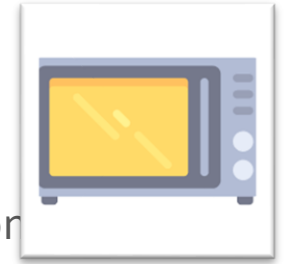
B. Convert input to a vector of basis functions

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

1. A microwave

Attributes:

- Volume: 17 l, 23 l, ...
- Functions: Micro, Con
- Power level
- Accessories
- Type of dial
- Brand
- Warranty
- Price



1. Image of shirt

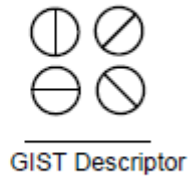
- Collar style
- Sleeve type
- Colour
- ...



# 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

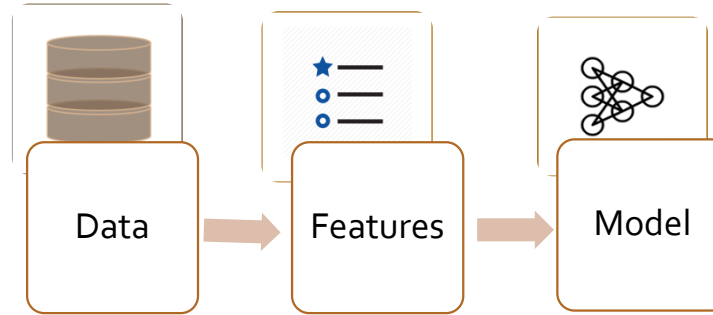
## # other attributes:

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

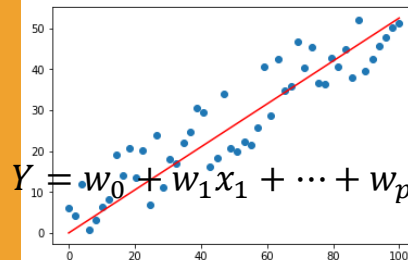


# 1B. Model Representation

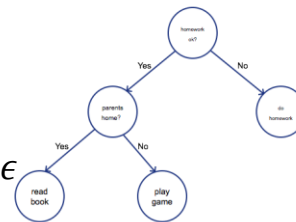
Hypothesis space

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

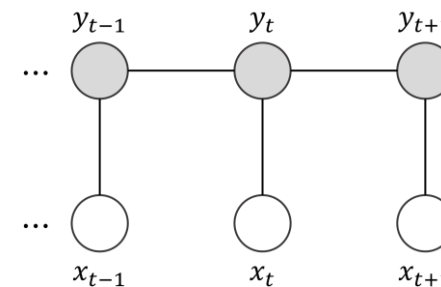
Linear Function



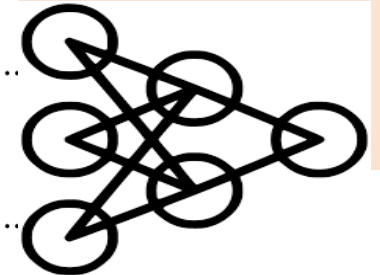
Decision Tree



Graphical Model



Neural Net



## 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

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

# 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

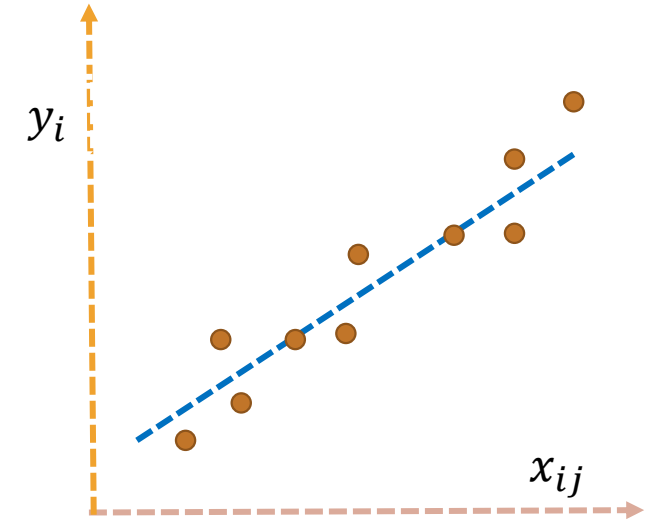
# Elements of Optimization

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

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



- $m$  data points,  $n$  features
  - $x_{ij}$ :  $j$ th attribute of  $i$ th instance
  - $y_i$ : output of  $i$ th instance

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