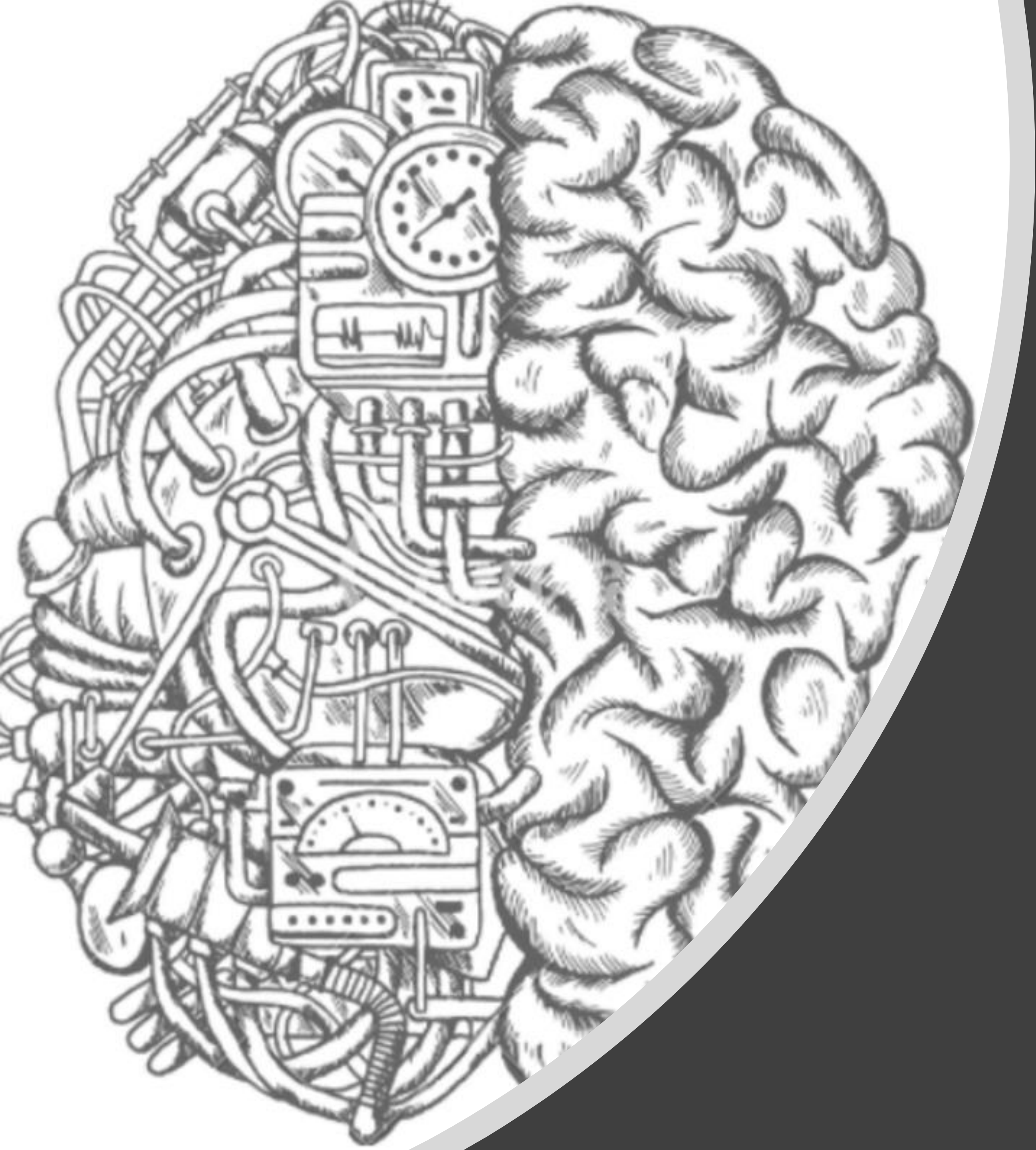


Machine Learning: Models and Applications

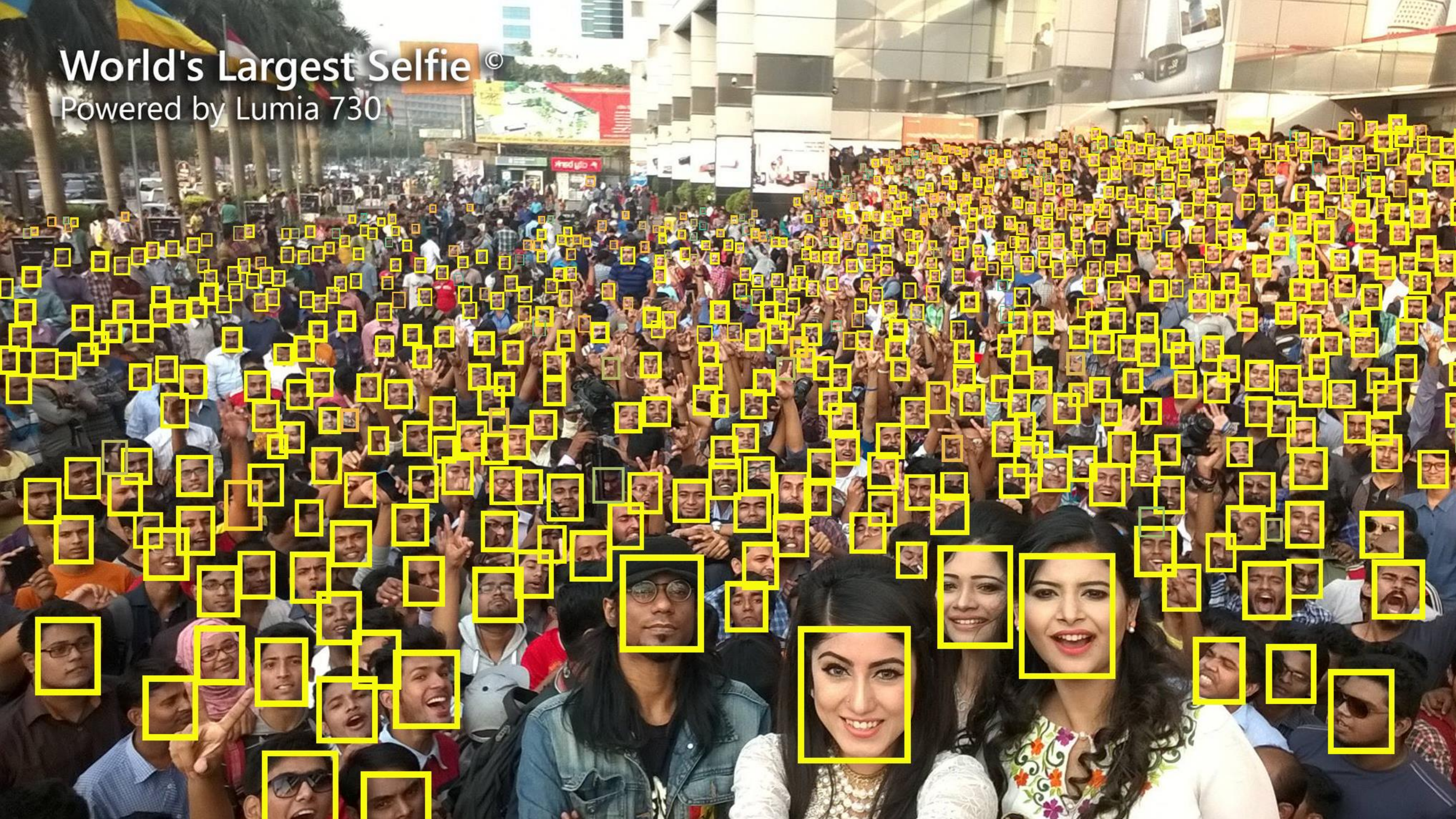
Lecture 1

Lecturer: Xinchao Wang

xinchao@nus.edu.sg

National University of Singapore


World's Largest Selfie ©
Powered by Lumia 730


World's Largest Selfie ©
Powered by Lumia 730

Objective

- Understand basic concepts of Machine Learning
- Obtain hands-on experience and implement basic algorithms
- Develop new algorithms working for real-world problems

Course Information

- Office Hour: appointment by email
- Evaluations:
  - Final Project
  - Final Exam

Prerequisites

- Probability
- Linear Algebra
- Programming
  - Language at your choice, but Matlab or Python preferred

Textbook

- **Pattern Recognition and Machine Learning**, by Christopher Bishop, Springer, 2006
- **Bayesian Reasoning and Machine Learning**, by David Barber, Cambridge University Press, 2012.
- **The Elements of Statistical Learning**, by Trevor Hastie, Robert Tibshirani and Jerome Friedman, Springer, 2009.

Outline

- What is machine learning?
- When do we need machine learning?
- Applications of machine learning
- Types of machine learning
- Walking through a toy example on classification

What is Machine Learning?

Learning is any process by which a system improves performance from experience.

- Herbert Simon

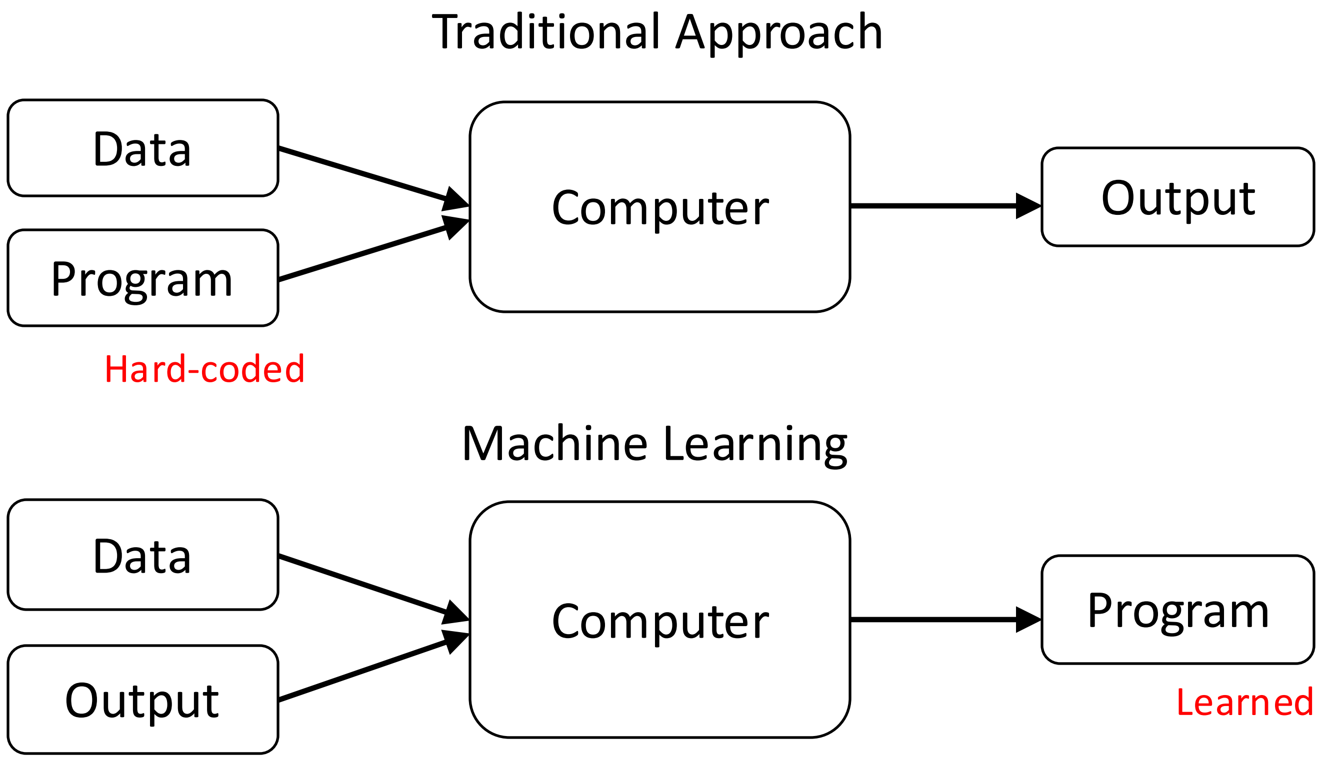
A computer program is said to learn

- from ***experience E***
- with respect to some class of ***tasks T***
- and performance ***measure P***,

if its performance at tasks in T, as measured by P, improves with experience E.

- Tom Mitchell

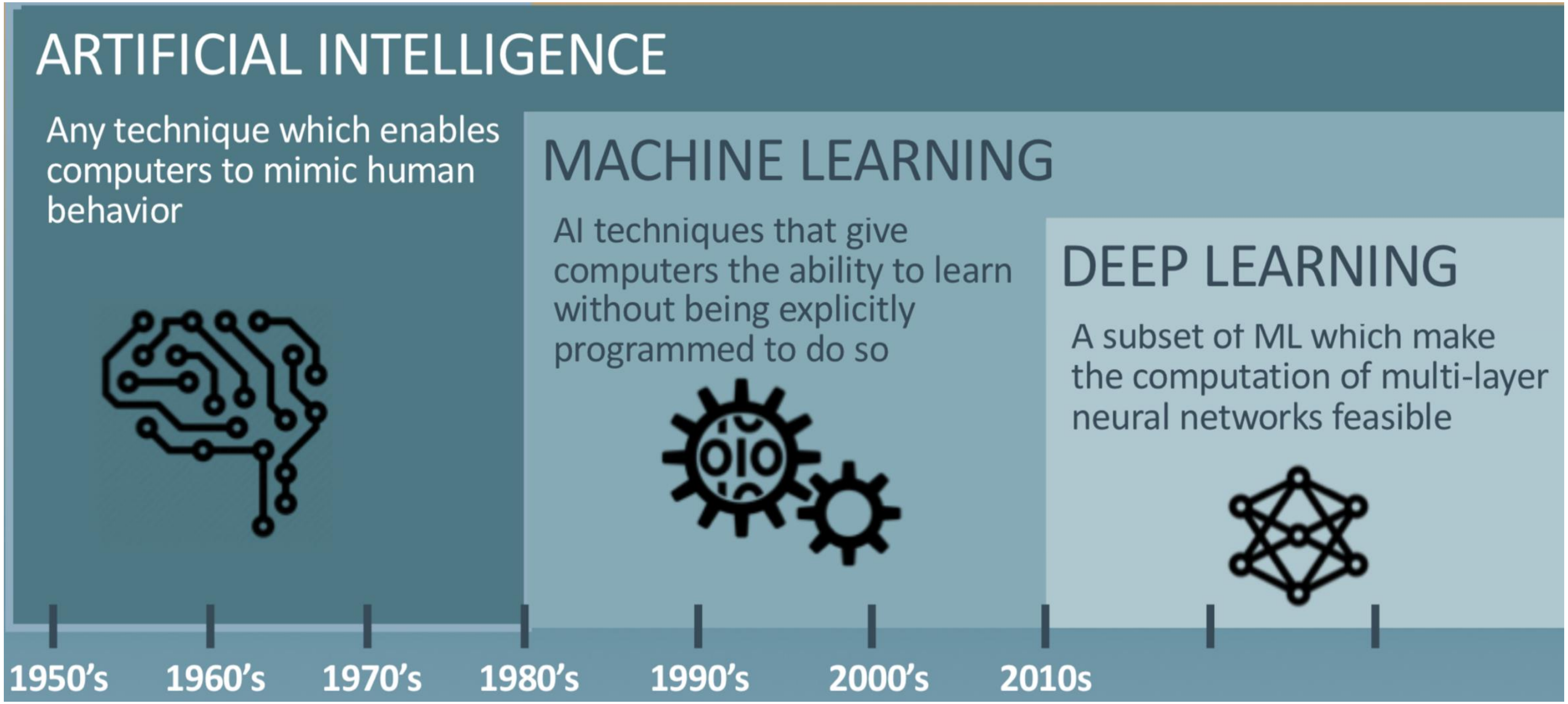
Machine Learning vs Traditional Approach



Machine Learning:
field of study that gives computers the ability to learn without being explicitly programmed

- Arthur Samuel

Artificial Intelligence (AI), Machine Learning, and Deep Learning



When do we need machine learning?

Lack of human expertise
(Navigating on Mars)

Human can’t explain their expertise
(Speech Recognition)

Models must be customized
(Personalized Medicine)

Involves huge amount of data
(Genomics)

Learning is not always useful:

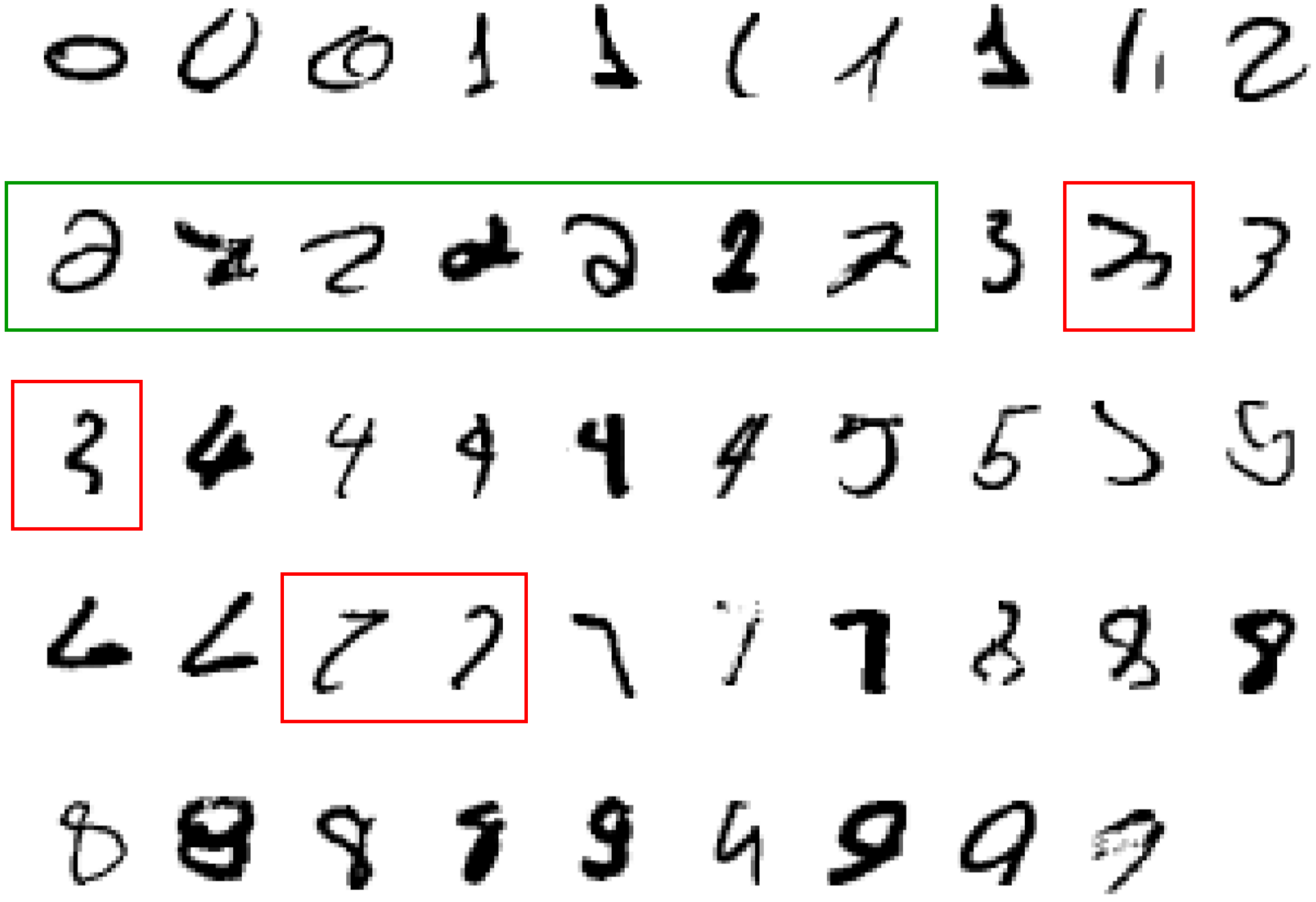
No need to “learn” to calculate payroll!

Application of Machine Learning

Task T, Performance P, Experience E

T: Digit Recognition
P: Classification Accuracy
E: Labelled Images

Labels -> Supervision!



Application of Machine Learning

Task T, Performance P, Experience E

T: Email Categorization
P: Classification Accuracy
E: Email Data, Some Labelled

Application of Machine Learning

Task T, Performance P, Experience E

T: Playing Go Game
P: Chances of Winning
E: Records of Past Games

Application of Machine Learning

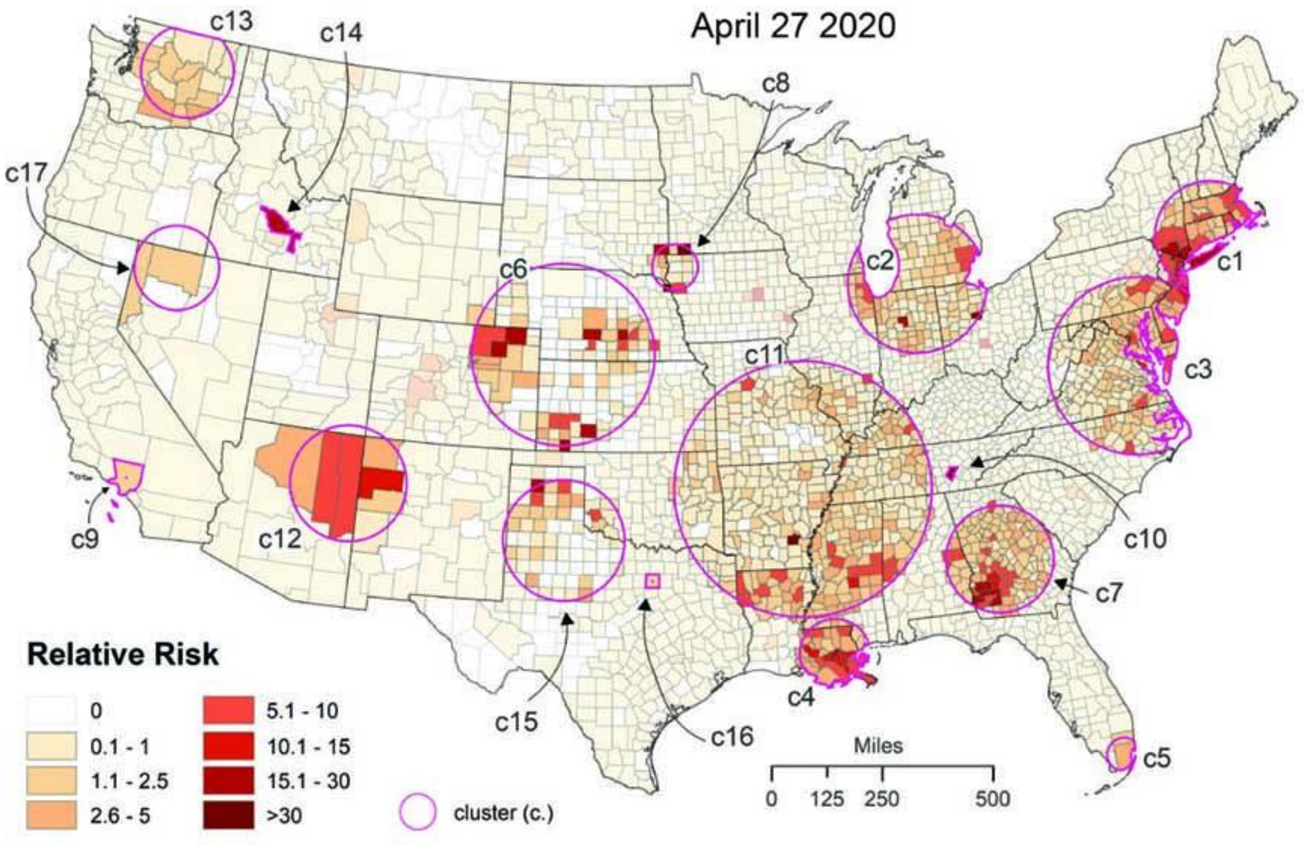
Task T, Performance P, Experience E

T: Recognizing Speech
P: Recognition Accuracy
E: Audio and Labelled Text

Application of Machine Learning

Task T, Performance P, Experience E

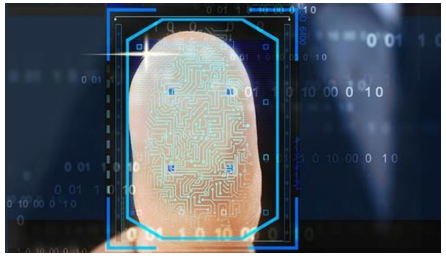
T: Identifying Covid-19 Clusters
P: Small Internal Distances
Larger External Distances
E: Records of Patients



Web Search Engine

Product Recommendation

Language Translation



Fingerprint Recognition

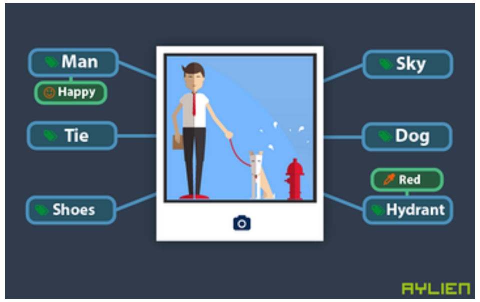


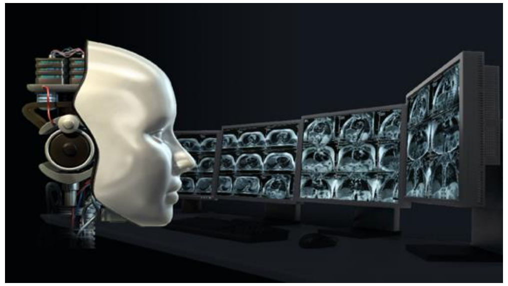
Photo Tagging

Virtual Personal Assistant

Portfolio Management

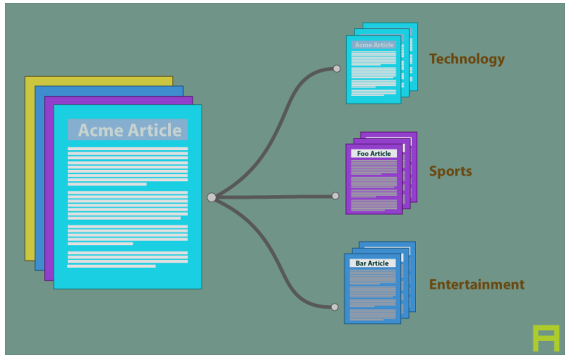
Chatbots

Traffic Prediction



Medical Diagnosis

Algorithmic Trading



Document Analysis

Types of machine learning

Supervised Learning

Input:
1) Training Samples,
2) Desired Output (Teacher/Supervision)

Output:
A rule that maps input to output

Unsupervised Learning

Input:
Samples

Output:
Underlying patterns in data

Reinforcement Learning

Input:
Sequence of States, Actions, and Delayed Rewards

Output:
Action Strategy: a rule that maps the environment to action

Types of machine learning

Supervised Learning

Input:
1) Training Samples,
2) Desired Output (Teacher/Supervision)

Output:
A rule that maps input to output

Unsupervised Learning

Input:
Samples

Output:
Underlying patterns in data

Reinforcement Learning

Input:
Sequence of States, Actions, and Delayed Rewards

Output:
Action Strategy: a rule that maps the environment to action

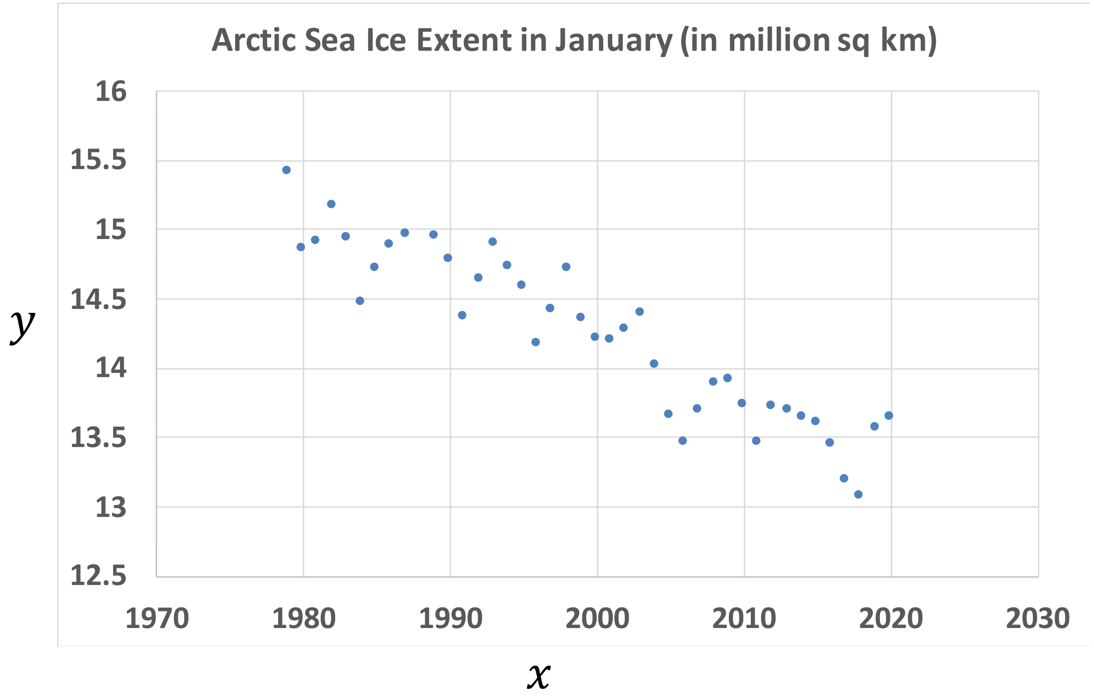
Supervised Learning

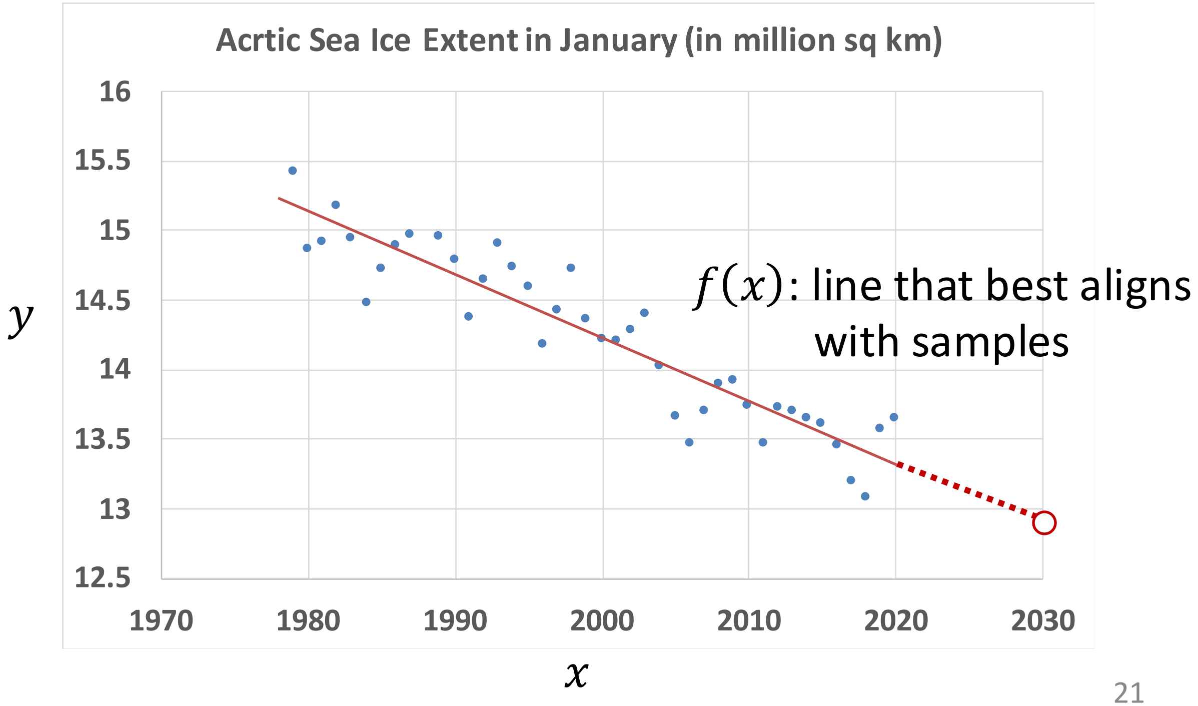
x → Regression → y (Continuous)

x → Classification → y (Categorical)

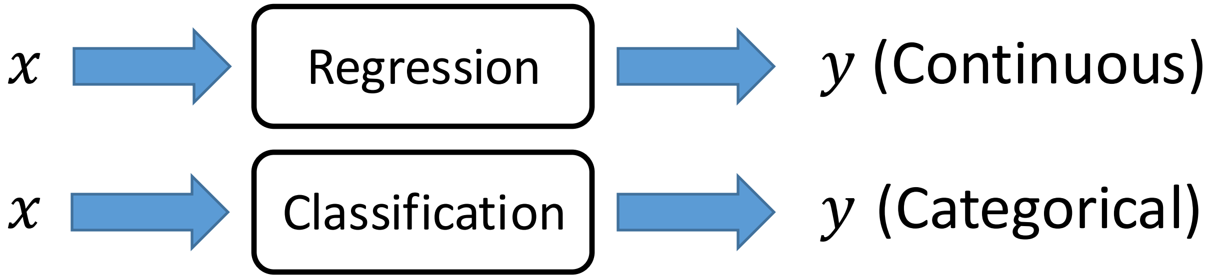
Regression

- Given $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)$
- Learn a function $f(\mathbf{x})$ to predict real-valued y given $\mathbf{x}$





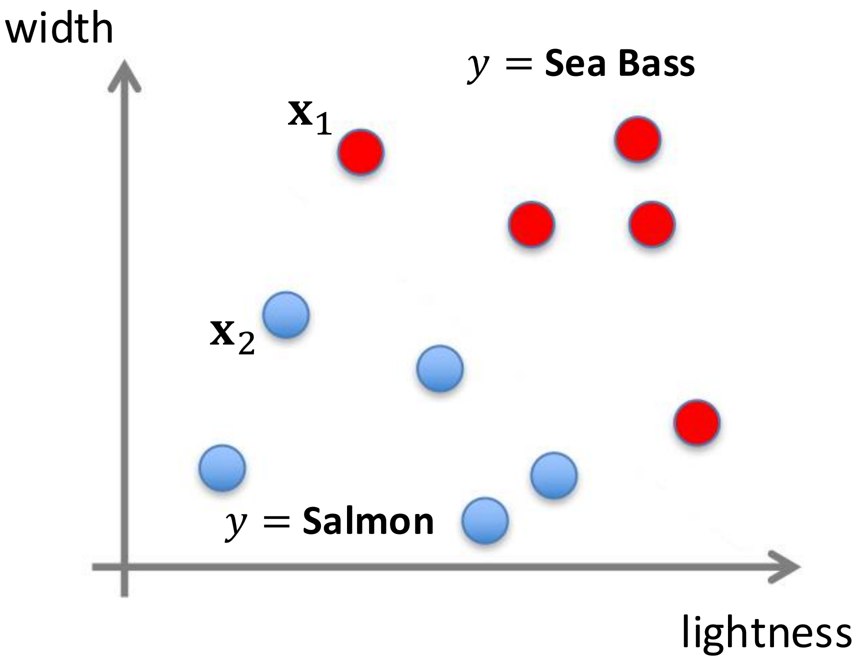
Supervised Learning

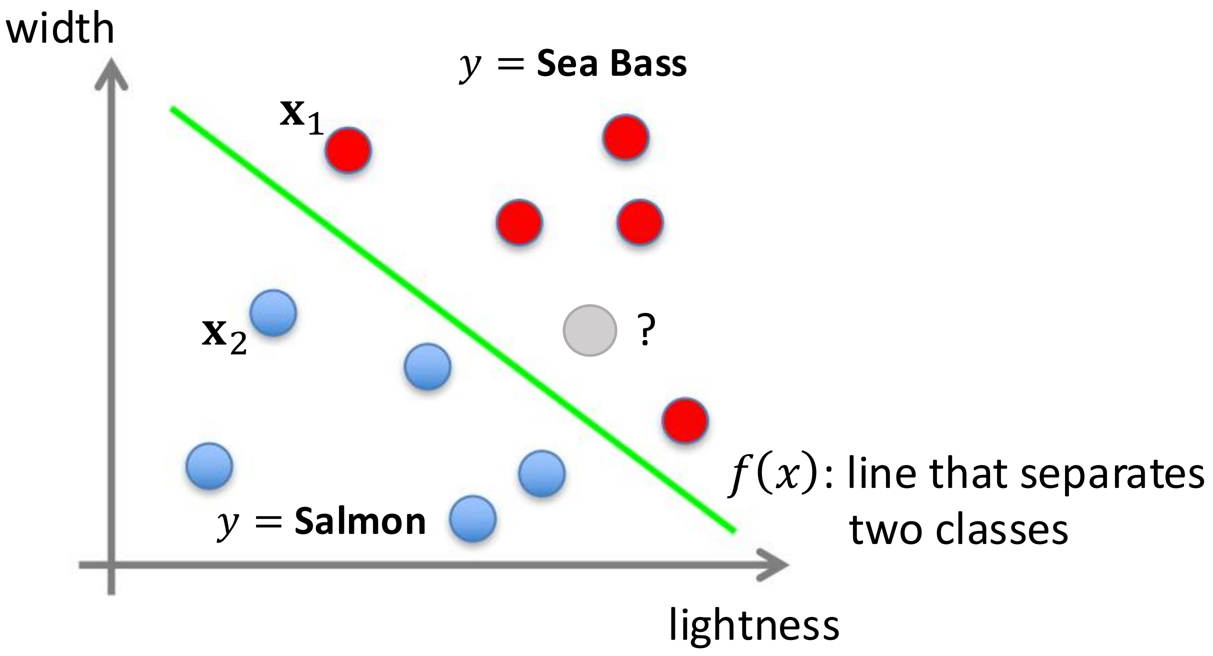


Classification

- Given $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)$
- Learn a function $f(\mathbf{x})$ to predict categorical y given $\mathbf{x}$

Feature Space





Types of machine learning

Supervised Learning

Input:
1) Training Samples,
2) Desired Output (Teacher/Supervision)

Output:
A rule that maps input to output

Unsupervised Learning

Input:
Samples

Output:
Underlying patterns in data

Reinforcement Learning

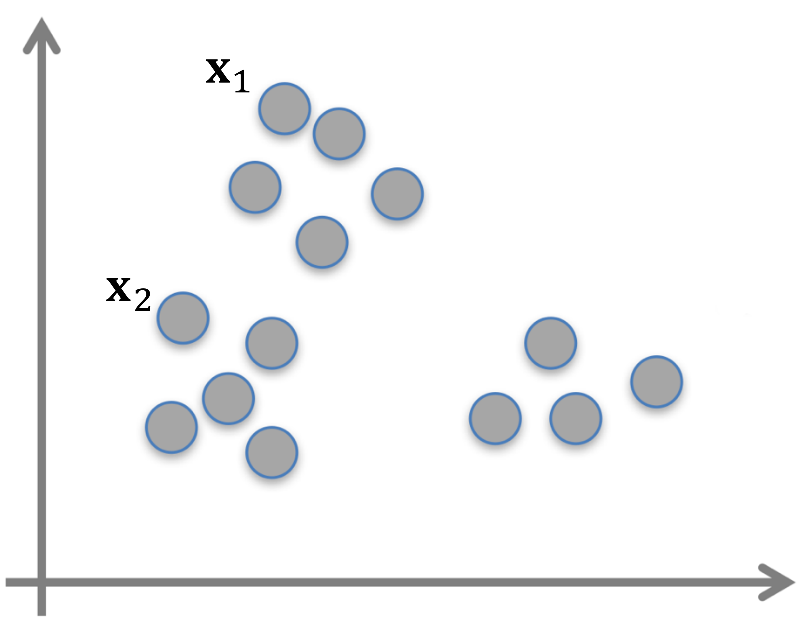
Input:
Sequence of States, Actions, and Delayed Rewards

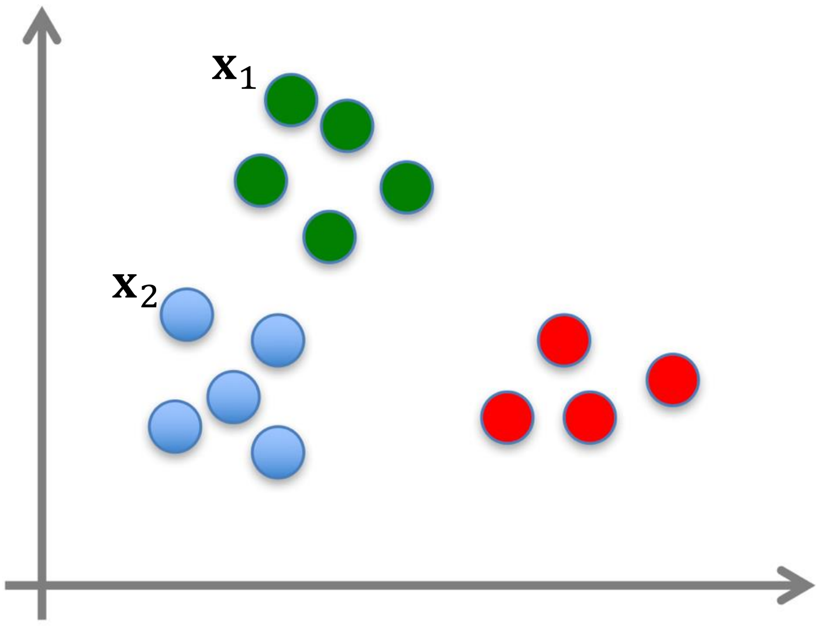
Output:
Action Strategy: a rule that maps the environment to action

Unsupervised Learning

Clustering

- Given $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N$, without labels
- Output Hidden Structure Behind





Types of machine learning

Supervised Learning

Input:
1) Training Samples,
2) Desired Output (Teacher/Supervision)

Output:
A rule that maps input to output

Unsupervised Learning

Input:
Samples

Output:
Underlying patterns in data

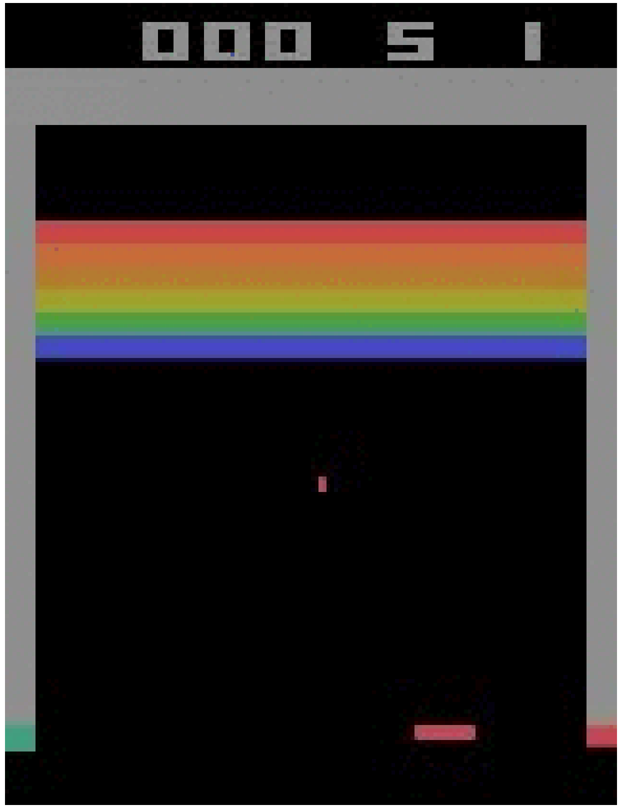
Reinforcement Learning

Input:
Sequence of States, Actions, and Delayed Rewards

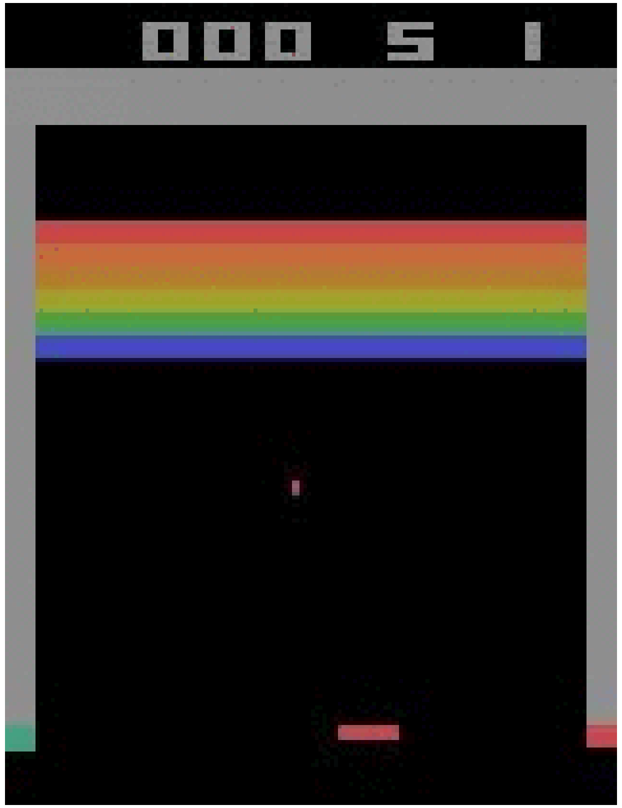
Output:
Action Strategy: a rule that maps the environment to action

Reinforcement Learning

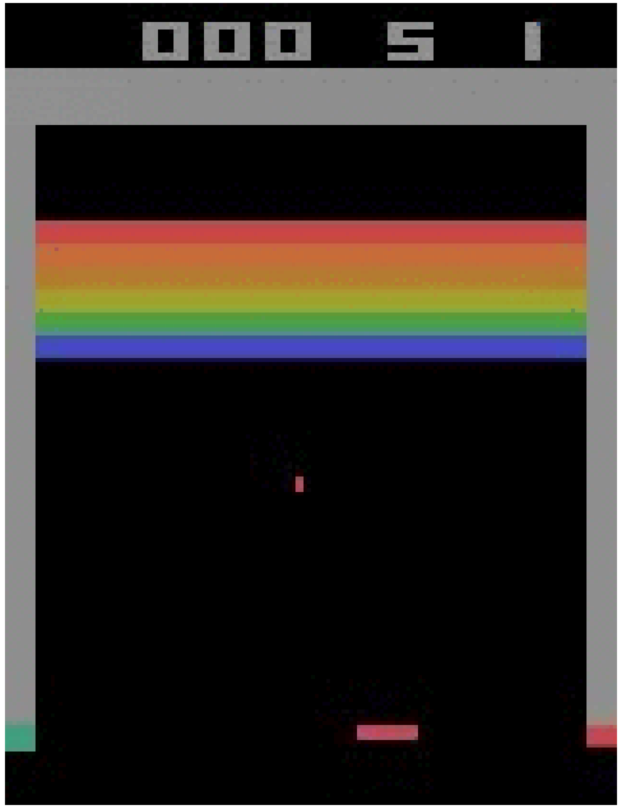
Breakout Game



Initial Performance



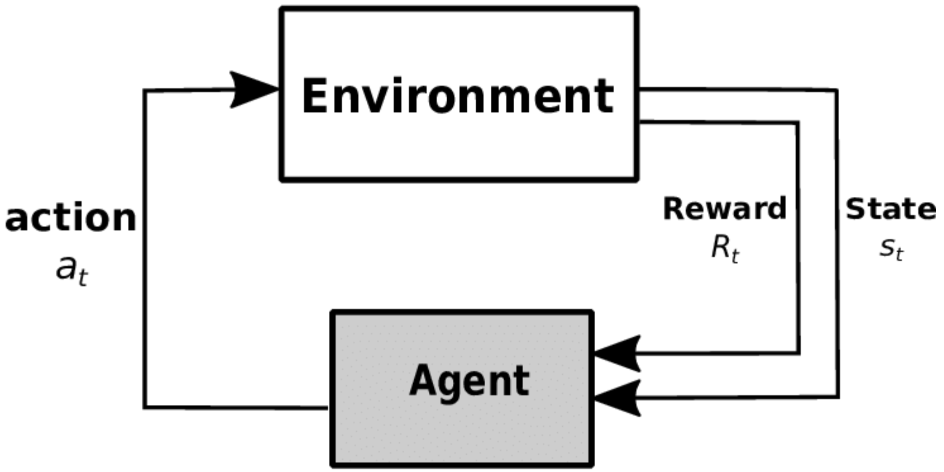
Training 15 minutes

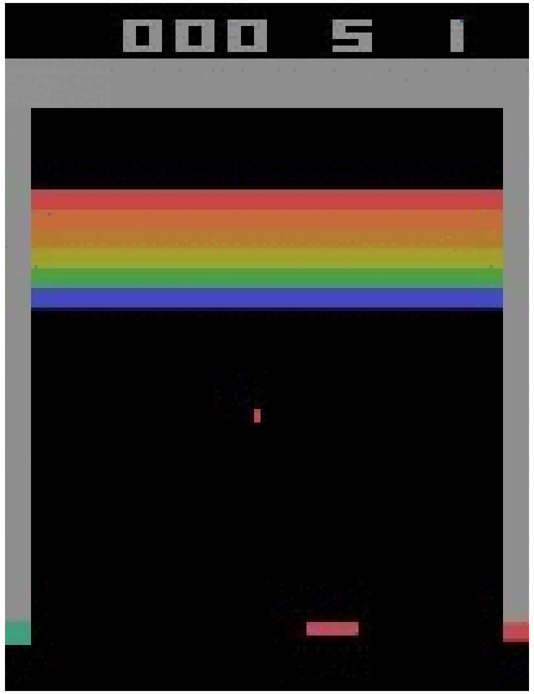


Training 30 minutes

Reinforcement Learning

- Given sequence of states $\boldsymbol{S}$ and actions $\boldsymbol{A}$ with (delayed) rewards $\boldsymbol{R}$
- Output a policy $\pi(a, s)$, to guide us what action a to take in state s





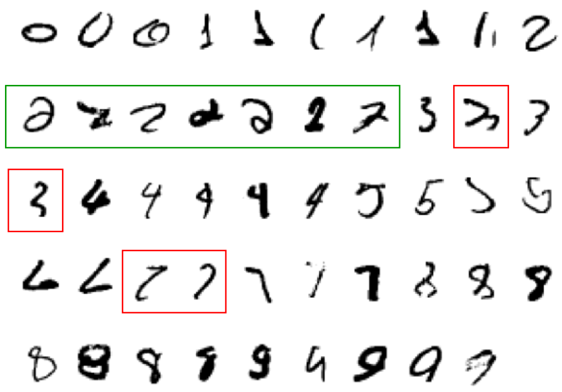
$\boldsymbol{S}$: Ball Location, Paddle Location, Bricks

$\boldsymbol{A}$: left, right

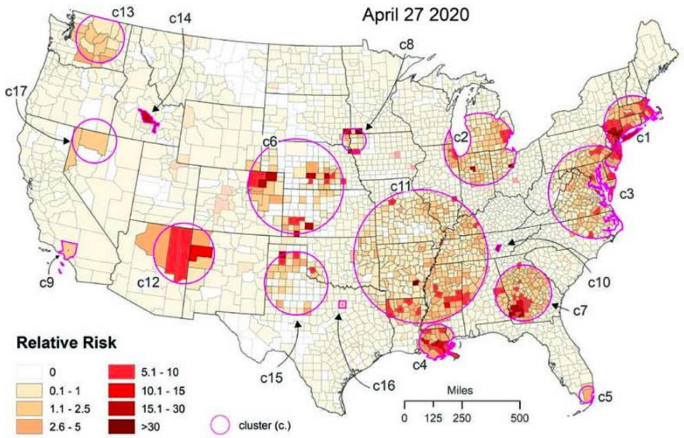
$\boldsymbol{R}$:
positive reward
Knocking a brick, clearing all bricks
negative reward
Missing the ball
zero reward
Cases in between

Quiz Time!

Supervised
Unsupervised
Reinforcement



Supervised



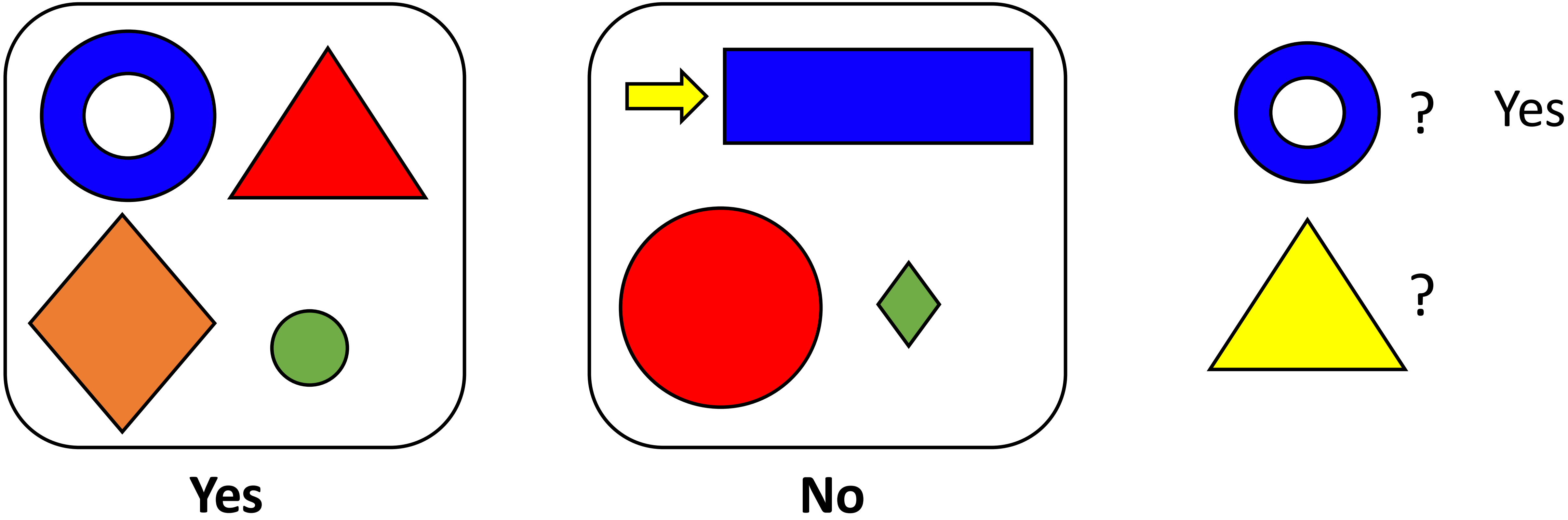
Unsupervised

Supervised

Supervised

Reinforcement

Walking Through A Toy Example: Token Classification

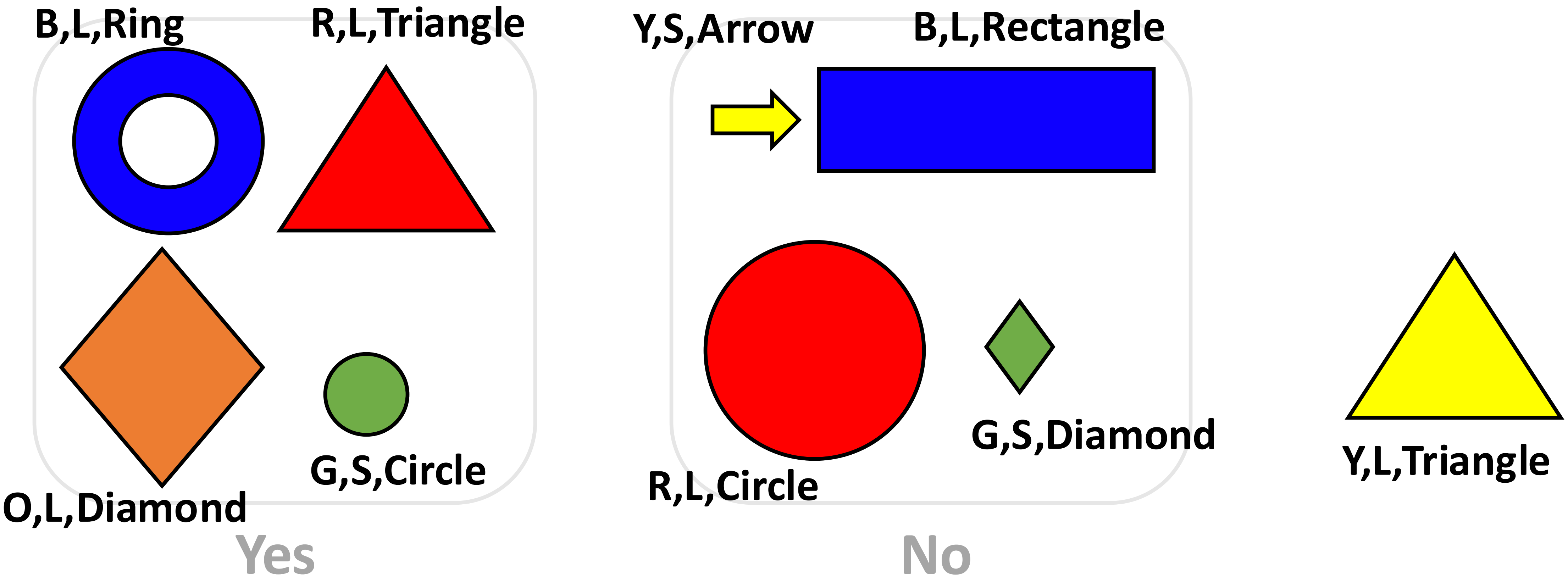


Step1: Feature Extraction
Extract Attributes of Samples

→

Step2: Sample Classification
Decide Label for a Sample

Walking Through A Toy Example: Token Classification

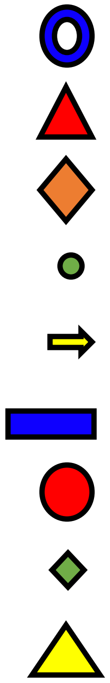


Step 1: Feature Extraction

Color, Size, Shape

Walking Through A Toy Example: Token Classification

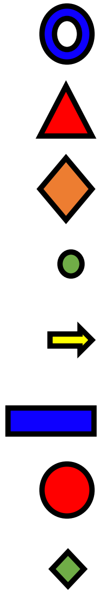
Feature Extraction



Color	Size	Shape	Label
Blue	Large	Ring	Yes
Red	Large	Triangle	Yes
Orange	Large	Diamond	Yes
Green	Small	Circle	Yes
Yellow	Small	Arrow	No
Blue	Large	Rectangle	No
Red	Large	Circle	No
Green	Small	Diamond	No
Yellow	Large	Triangle	?

Walking Through A Toy Example: Token Classification

Feature Extraction



Color	**Size**	**Shape**	**Label**
Blue	Large	Ring	Yes
Red	Large	Triangle	Yes
Orange	Large	Diamond	Yes
Green	Small	Circle	Yes
Yellow	Small	Arrow	No
Blue	Large	Rectangle	No
Red	Large	Circle	No
Green	Small	Diamond	No

Walking Through A Toy Example: Token Classification

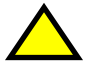
Feature Extraction

	Color	Size	Shape	Label
	Blue	Large	Ring	Yes
	Red	Large	Triangle	Yes
	Orange	Large	Diamond	Yes
	Green	Small	Circle	Yes
	Yellow	Small	Arrow	No
	Blue	Large	Rectangle	No
	Red	Large	Circle	No
	Green	Small	Diamond	No

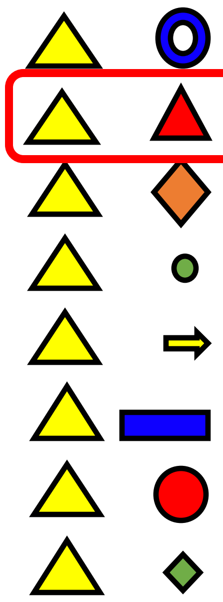
Similarity

		Color	Size	Shape	Total
		0	1	0	1
		0	1	1	2
		0	1	0	1
		0	0	0	0
		1	0	0	1
		0	1	0	1
		0	1	0	1
		0	0	0	0

Walking Through A Toy Example: Token Classification

: Yes!

Similarity



Color	Size	Shape	Total
0	1	0	1
0	1	1	2
0	1	0	1
0	0	0	1
1	0	0	1
0	1	0	1
0	1	0	1
0	0	0	0

Nearest Neighbor Classifier:

1) Find the “nearest neighbor” of a sample in the feature space
2) Assign the label of the nearest neighbor to the sample

Summary by Quick Quiz

Three Components in ML Definition

Task T, Performance P, Experience E

Three Types of in ML

Supervised Learning
Unsupervised Learning
Reinforcement Learning

Two Types of Supervised Learning

Classification, Regression

One Type of Unsupervised Learning

Clustering

Example of a Classifier Model

Nearest Neighbor Classifier

Probability Review

The Axioms of Probability

- 0 <= P(A) <= 1
- P(True) = 1
- P(False) = 0
- P(A or B) = P(A) + P(B) – P(A and B)

Overview

- Discrete Random Variables
- Expected Value
- Pairs of Discrete Random Variables
  - Conditional Probability
  - Bayes Rule
- Continuous Random Variables

Overview

- Discrete Random Variables
- Expected Value
- Pairs of Discrete Random Variables
  - Conditional Probability
  - Bayes Rule
- Continuous Random Variables

Discrete Random Variables

- A **Random Variable**, denoted by r.v. x, is a measurement on an outcome of a random event

- **Discrete Random Variable**:
  - The r.v. can assume a finite or countably infinite number of values.

- **Continuous Random Variable**:
  - The r.v. can assume all values in an interval.

Examples

- Which of the following r.v. are discrete and which are continuous?
  - X = Number of houses sold per week?
  - X = Number of heads in ten tosses of a coin?
  - X = Weight of a child at birth?
  - X = Time required to run 100 meters?

Examples

- Dice
  - Probability of rolling 5-6 or two 6s with two dice

- Examples of Probability Distribution

Probability Distribution Example:

Suppose that you have a die which, when thrown, takes the numbers from 1 to 6 with equal probability.

red	1 2 3 4 5 6

Probability Distribution Example:

Suppose that you also have a green die that can take the numbers from 1 to 6 with equal probability.

red / green	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

Probability Distribution Example: X is the Sum of Two Dice

We will define a random variable X as the sum of the numbers when the dice are thrown.

red green	1	2	3	4	5	6
1						
2						
3						
4						
5						
6						

Probability Distribution Example: X is the Sum of Two Dice

For example, if the red die is 4 and the green one is 6, X is equal to 10.

red green	1	2	3	4	5	6
1						
2						
3						
4						
5						
6				10		

Probability Distribution Example: X is the Sum of Two Dice

Similarly, if the red die is 2 and the green one is 5, X is equal to 7.

red green	1	2	3	4	5	6
1						
2						
3						
4						
5		7				
6						

Probability Distribution Example:
X is the Sum of Two Dice

This table shows all the possible outcomes.

red green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

Probability Distribution Example: X is the Sum of Two Dice

If you look at the table, you can see that X can be any of the numbers from 2 to 12.

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X
2
3
4
5
6
7
8
9
10
11
12

Probability Distribution Example: X is the Sum of Two Dice

We will now define f, the frequencies associated with the possible values of X.

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	

Probability Distribution Example: X is the Sum of Two Dice

For example, there are four outcomes which make X equal to 5.

red green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f
2	
3	
4	
5	4
6	
7	
8	
9	
10	
11	
12	

Probability Distribution Example: X is the Sum of Two Dice

Similarly you can work out the frequencies for all the other values of X.

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f
2	1
3	2
4	3
5	4
6	5
7	6
8	5
9	4
10	3
11	2
12	1

Probability Distribution Example: X is the Sum of Two Dice

Finally we will derive the probability of obtaining each value of X.

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f	p
2	1	
3	2	
4	3	
5	4	
6	5	
7	6	
8	5	
9	4	
10	3	
11	2	
12	1	

Probability Distribution Example: X is the Sum of Two Dice

If there is 1/6 probability of obtaining each number on the red die, and the same on the green die, each outcome in the table will occur with 1/36 probability.

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f	p
2	1	
3	2	
4	3	
5	4	
6	5	
7	6	
8	5	
9	4	
10	3	
11	2	
12	1	

Probability Distribution Example: X is the Sum of Two Dice

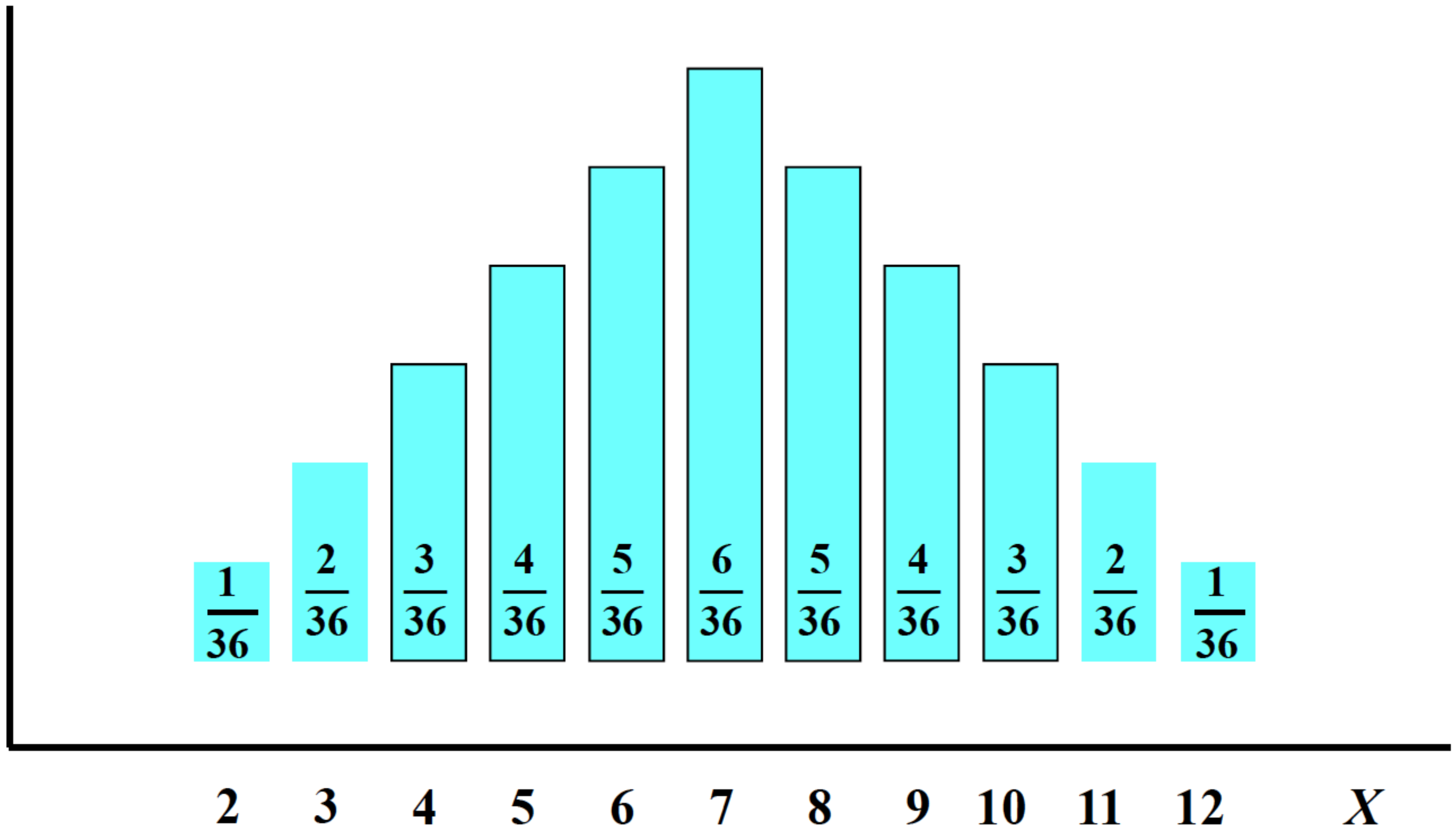
Hence to obtain the probabilities associated with the different values of X, we divide the frequencies by 36.

red / green	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

X	f	p
2	1	1/36
3	2	2/36
4	3	3/36
5	4	4/36
6	5	5/36
7	6	6/36
8	5	5/36
9	4	4/36
10	3	3/36
11	2	2/36
12	1	1/36

Probability Distribution Example: X is the Sum of Two Dice

The distribution is shown graphically.



Overview

- Discrete Random Variables
- Expected Value
- Pairs of Discrete Random Variables
  - Conditional Probability
  - Bayes Rule
- Continuous Random Variables

Expected Value

- Definition of E(X), the expected value of X:

$$E(X) = x_1 p_1 + \ldots + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

- Also known as the ***population mean***
- Weighted average of its possible values, the weights being the probabilities attached to the values

Expected Value Example

x_i	p_i	$x_i p_i$
x_1	p_1	$x_1 p_1$
x_2	p_2	$x_2 p_2$
x_3	p_3	$x_3 p_3$
x_4	p_4	$x_4 p_4$
x_5	p_5	$x_5 p_5$
x_6	p_6	$x_6 p_6$
x_7	p_7	$x_7 p_7$
x_8	p_8	$x_8 p_8$
x_9	p_9	$x_9 p_9$
x_{10}	p_{10}	$x_{10} p_{10}$
x_{11}	p_{11}	$x_{11} p_{11}$
		$\Sigma\, x_i p_i = E(X)$

x_i	p_i	$x_i p_i$
2	1/36	2/36
3	2/36	6/36
4	3/36	12/36
5	4/36	20/36
6	5/36	30/36
7	6/36	42/36
8	5/36	40/36
9	4/36	36/36
10	3/36	30/36
11	2/36	22/36
12	1/36	12/36
		252/36 = 7

Expected Value Properties

- Linear
  - $E(X + Y) = E(X) + E(Y)$
  - $E(bX) = bE(X)$, b is a constant
  - $E(b) = b$
  - $Y = b_1 + b_2X$, we have $E(Y) = E(b_1 + b_2X) = E(b_1) + E(b_2X) = b_1 + b_2E(X)$

- Also often referred as μ

Variance

- Definition:

$$Var(X) = E[(X-\mu)^2] = \sum (x_i - \mu)^2 P(X = x_i)$$

- One property:

$$Var(X) = E[(X-\mu)^2] = E[X^2] - (E[X])^2$$

Overview

- Discrete Random Variables
- Expected Value
- Pairs of Discrete Random Variables
  - Conditional Probability
  - Bayes Rule
- Continuous Random Variables

Pairs of Discrete Random Variables

- Let x and y be two discrete r.v.
- For each possible pair of values, we can define a joint probability

$$P_{ij} = P(X = x_i, Y = y_j)$$

- Marginal Distribution:

$$P_x(x) = \sum_{y \in Y} P(x, y)$$

$$P_y(y) = \sum_{x \in X} P(x, y)$$

Statistical Independence

- Two events A and B are independent if their joint probability equals the product of their probabilities:

$$P(A \cap B) = P(A)P(B)$$

- Two random variables X and Y are independent iff, for every a and b, the events $\{X \leq a\}$ and $\{Y \leq b\}$ are independent events.

Conditional Probability

- When two r.v. are not independent, knowing one allows better estimate of other (e.g. outside temperature, season)

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$

- If independent, P(X|Y) = P(X)

Sum and Product Rules

- Sum Rule

$$P(X) = \sum_{Y} P(X, Y)$$

- Product Rule

$$P(X, Y) = P(Y|X)P(X)$$

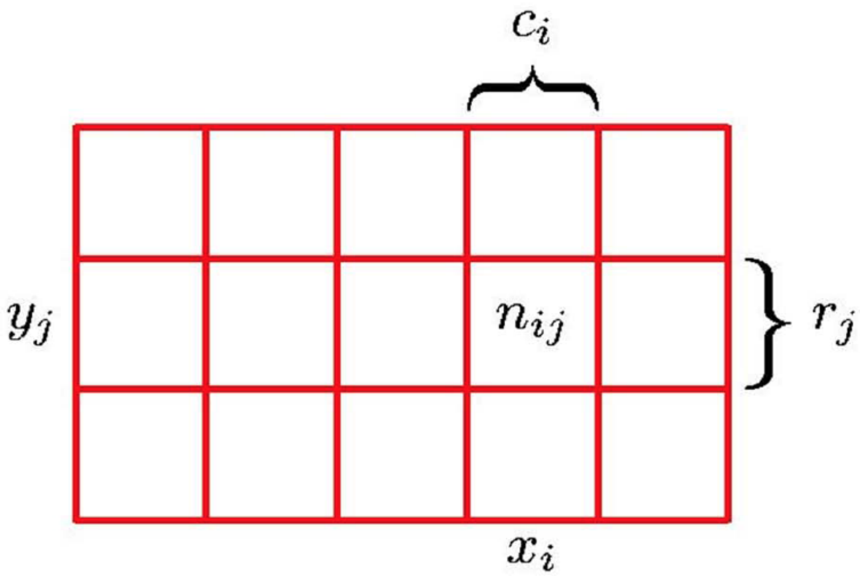
Law of Total Probability

$$P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j P(X = x_i | Y = y_j) P(Y = y_j)$$

Sum Rule

Product Rule

Sum and Product Rules



Marginal Probability

$$p(X = x_i) = \frac{c_i}{N}.$$

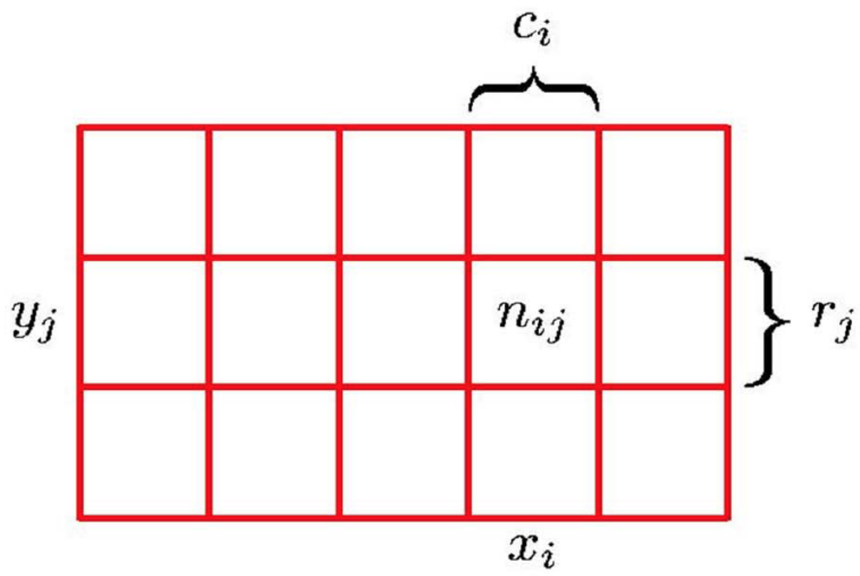
Joint Probability

$$p(X = x_i, Y = y_j) = \frac{n_{ij}}{N}$$

Conditional Probability

$$p(Y = y_j | X = x_i) = \frac{n_{ij}}{c_i}$$

Sum and Product Rules



Sum Rule

$$p(X = x_i) = \frac{c_i}{N} = \frac{1}{N}\sum_{j=1}^{L} n_{ij}$$

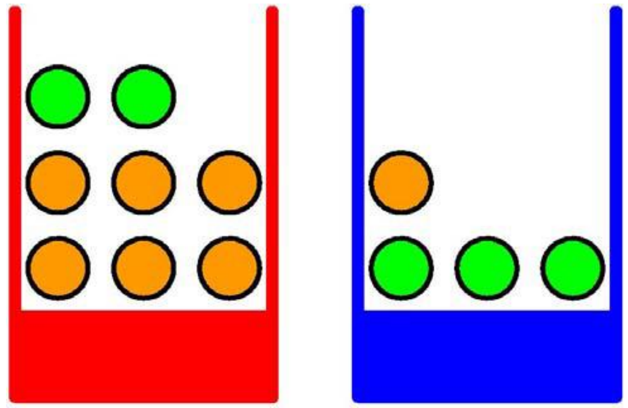
$$= \sum_{j=1}^{L} p(X = x_i, Y = y_j)$$

Product Rule

$$\begin{aligned} p(X = x_i, Y = y_j) &= \frac{n_{ij}}{N} = \frac{n_{ij}}{c_i} \cdot \frac{c_i}{N} \\ &= p(Y = y_j | X = x_i) p(X = x_i) \end{aligned}$$

Example

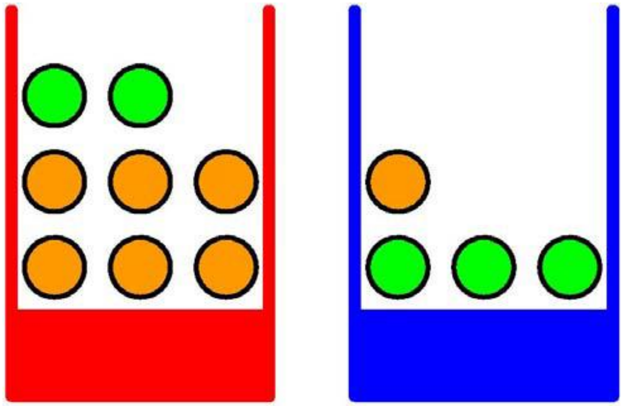
- Let us look at the following example:
  - We have two boxes, one red and one blue
  - Red box: 2 apples and 6 oranges
  - Blue box: 3 apples and 1 orange
  - Pick red box 40% of the time and blue box 60% of the time, then pick one item of fruit



C.M. Bishop, “Pattern Recognition and Machine Learning”, 2006

Example

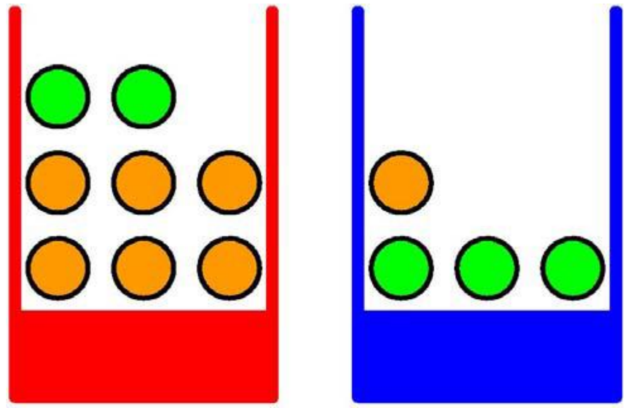
- Define:
  - B random variable for box picked
    - B = {blue(b), red(r)}
  - F identity of fruit
    - F = {apple(a), orange(o)}
  - P(B=r)=0.4 and P(B=b)=0.6
  - Events are mutually exclusive and include all possible outcomes
  - Their probabilities must sum to 1



Example

- P(B=r) = 0.4, P(B=b) = 0.6
- P(B=r) + P(B=b) = 1.0

- Conditional Probabilities
  - P(F=a|B=r) = 2/8 = 0.25
  - P(F=o|B=r) = 6/8 = 0.75
  - P(F=a|B=b) = 3/4 = 0.75
  - P(F=o|B=b) = 1/4 = 0.25



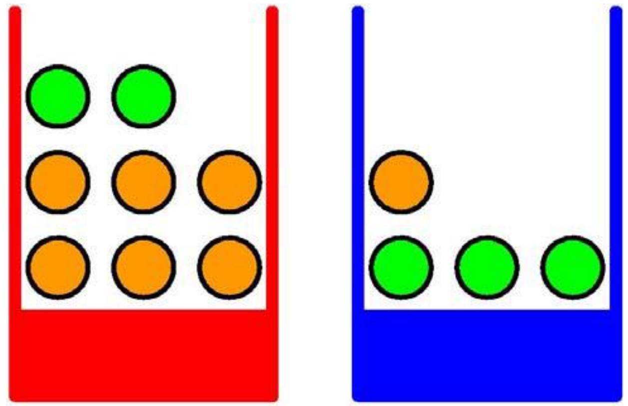
Example

- Note: $P(F=a|B=r)+P(F=o|B=r) = 1$

- $P(F=a) = P(F=a|B=r)P(B=r)+P(F=a|B=b)P(B=b)$

  $= 0.25 * 0.4 + 0.75 * 0.6 = 0.55$

- $P(F=o) = ?$



$P(F=a|B=r) = 2/8 = 0.25$
$P(F=o|B=r) = 6/8 = 0.75$
$P(F=a|B=b) = 3/4 = 0.75$
$P(F=o|B=b) = 1/4 = 0.25$

Bayes Rule

- Recall Product Rule:

$$P(X, Y) = P(Y|X)P(X)$$

$$P(X) = \sum_Y P(X, Y)$$

- Bayes Rule -- rewrite Product Rule and Sum Rule:

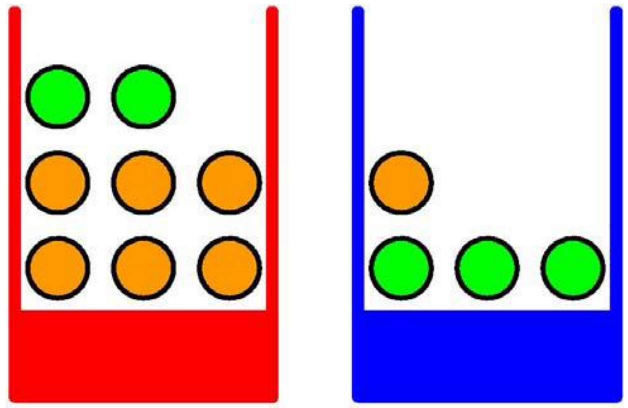
$$\underbrace{P(X|Y)}_{\text{Posterior}} = \frac{P(X, Y)}{P(Y)} = \frac{\overbrace{P(Y|X)}^{\text{Likelihood}}\overbrace{P(X)}^{\text{Prior}}}{\underbrace{\sum_X P(X, Y)}_{\text{Evidence}}}$$

Bayes Rule on the Fruit Example

- Assume we have now picked an orange
- Ask: which is the probability that it was from the red box?

$$P(B = r|F = o) = \frac{P(F = o|B = r)P(B = r)}{P(F = o)} = \frac{0.75 \times 0.4}{0.45} = \frac{2}{3}$$

P(B=r) = 0.4, P(B=b) = 0.6
P(F=o) = 0.45,
P(F=o|B=r) = 6/8 = 0.75



Overview

- Discrete Random Variables
- Expected Value
- Pairs of Discrete Random Variables
  - Conditional Probability
  - Bayes Rule
- Continuous Random Variables

Continuous Random Variables

- Examples:
  - Room temperature
  - Time to run 100m
  - Height of a person
- Cannot talk about probability of that X has a particular value
- Instead, probability X falls into an interval

$$Pr[x \in (a,b)] = \int_a^b p(x)dx$$

$$p(x) \geq 0$$

$$\int_{-\infty}^{\infty} p(x)dx = 1$$

Continuous Random Variables

$$E[x] = \mu = \int_{-\infty}^{\infty} x p(x) dx$$

$$E[f(x)] = \int_{-\infty}^{\infty} f(x) p(x) dx$$

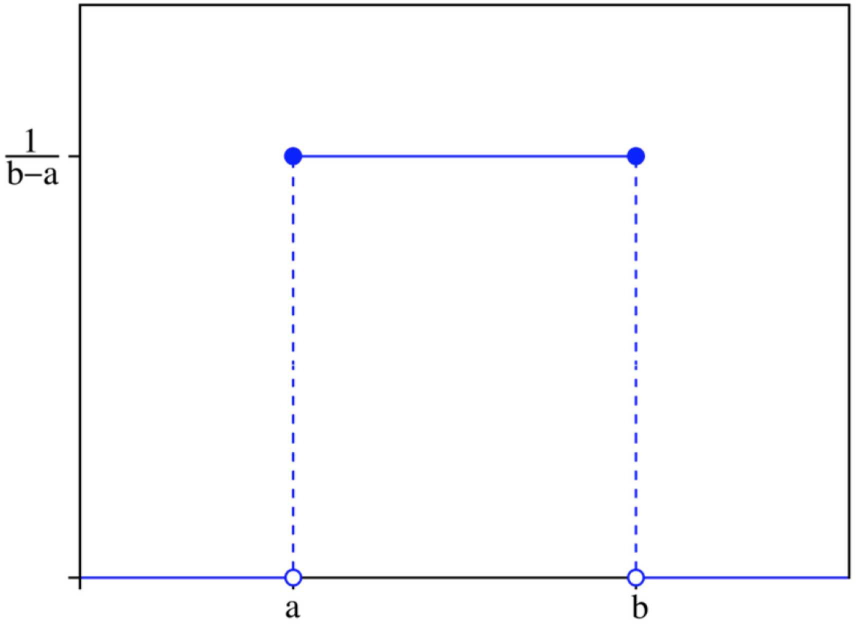
$$Var[x] = \sigma^2 = \int_{-\infty}^{\infty} (x-\mu)^2 p(x) dx$$

$$p(x \mid y) = \frac{p(y \mid x) p(x)}{\int_{-\infty}^{\infty} p(y \mid x) p(x) dx}$$

$$\text{posterior} = \frac{\text{likelihood} * \text{prior}}{\text{evidence}}$$

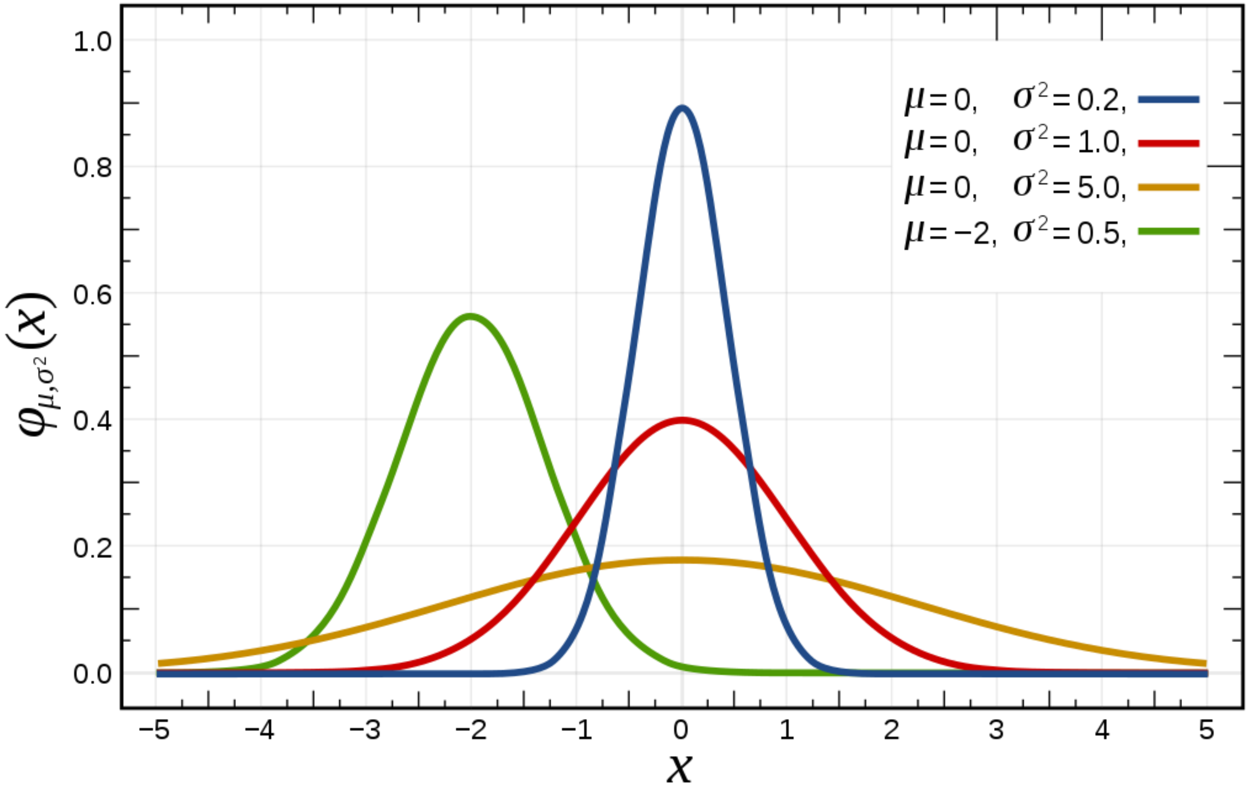
Uniform Distribution

$$p(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a <= x <= b \\ 0 & \text{for } x < a \text{ or } x > b \end{cases}$$



Normal (Gaussian) Distribution

$$p(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = N(\mu, \sigma^2)$$



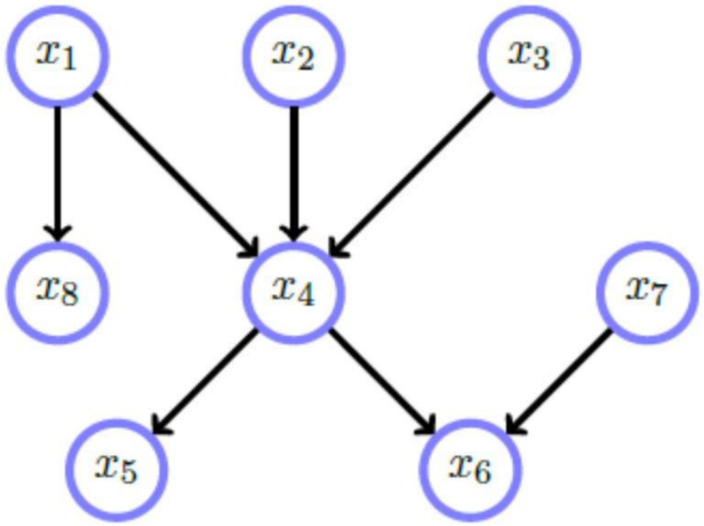
Quick Intro to Graphical Model

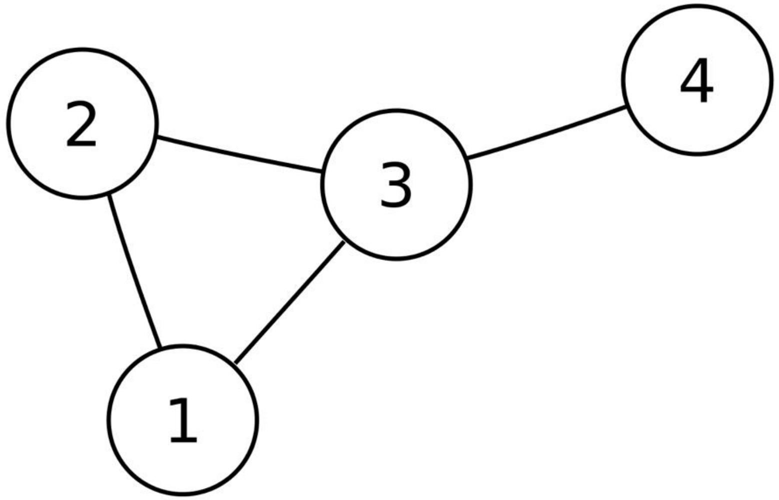
Graphical Models

- GMs are graph-based representations of various ***factorization assumptions*** of distributions
  - These factorizations are typically equivalent to **independence** statements amongst (sets of) variables in the distribution

Definitions

- A graph G contains:
  - Nodes, also known as vertices
  - Edges, also know as links between nodes
- Edges may be directed or undirected
  - They may also have associated weights
- Directed Graphs:
  - All edges are directed
- Undirected Graphs:
  - All edges are undirected





More Definitions

- Path:
  - The Path A->B from node A to node B is a sequence of nodes that connects A to B

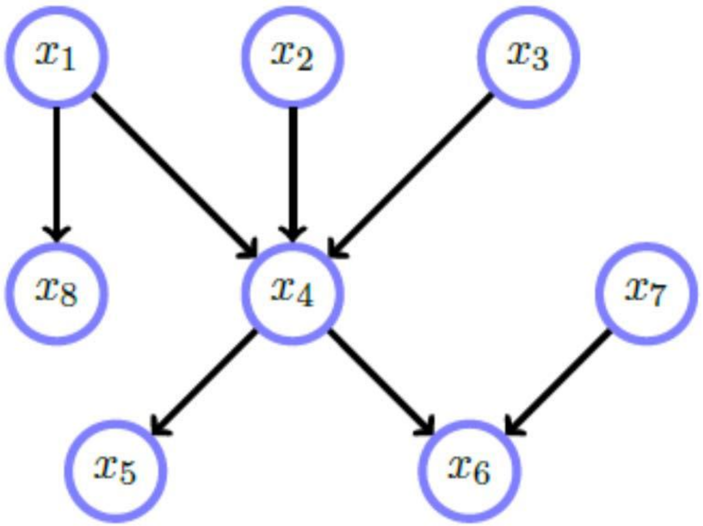
- Cycle:
  - A cycle is a directed path that starts and returns to the same node

- Directed Acyclic Graph (DAG):
  - A DAG is a graph G with directed edges (arrows on each link) between the nodes such that by following a path of nodes from one node to another along the direction of each edge, no path will revisit a node
  - Directed Graph without cycles

More Definitions

- Parent
  - Parent of x_4 are $pa(x_4) = \{x_1, x_2, x_3\}$
- Children
  - Children of x_4 are $ch(x_4) = \{x_5, x_6\}$

- Graphs can be represented using:
  - The edge list L={(1,8),(1,4),(2,4) …}
  - The adjacency matrix



Slides borrowed or adopted from

- Special thanks to
  - Philippos Mordohai
- Also thanks to
  - David Barber
  - Erik Sudderth
  - Dhruv Batra
  - Pedro Domingos
  - Raquel Urtasun
  - Richard Zemel