CS 559: Machine Learning Fundamentals and Applications
1st Set of Notes

Instructor: Philippos Mordohai
Webpage: www.cs.stevens.edu/~mordohai
E-mail: Philippos.Mordohai@stevens.edu
Office: Lieb 215

Objectives

- Obtain hands-on experience with and be able to implement fundamental algorithms
  - Useful for everyday problems
- Be able to use state of the art machine learning and pattern recognition tools for advanced problems

Important Points

- This is an elective course. You chose to be here.
- Expect to work and to be challenged.
- Exams won't be based on recall. They will be open book and you will be expected to solve new problems.

Important Points II

- Always ask:
  - What are we classifying?
  - What is known, what is unknown?
  - Which are the classes/labels/options?
  - What is the objective function?

Logistics

- Office hours: Tuesday 5-6 and by email
- Evaluation:
  - Homework assignments (20%)
  - Project (25%)
  - Pop-up quizzes and participation (10%)
  - Midterm (20%)
  - Final exam (25%)

Project

- Pick topic BEFORE middle of the semester
- I will suggest ideas and datasets in next lectures
- Deliverables:
  - Project proposal
  - Presentation in class
  - Poster in CS department event
  - Final report (around 8 pages)

Project Examples

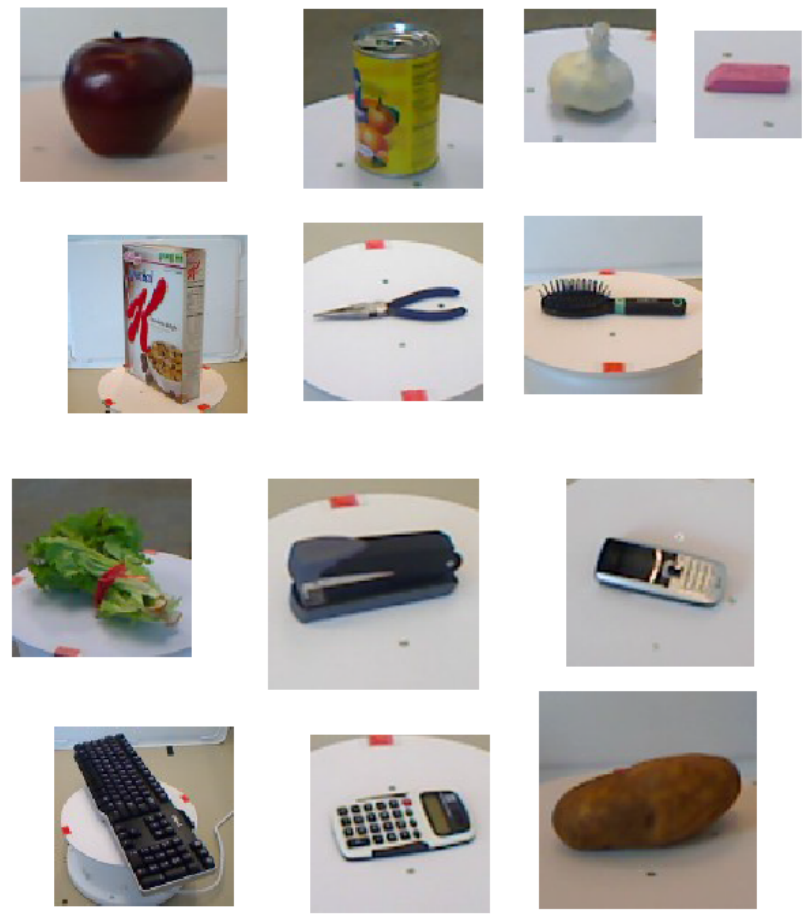
- Face detection

Project Examples

- Spam filtering
- Gender identification from emails
- Author recognition from text
- Handwriting recognition
- Speech recognition
- Malicious website detection

Project Examples

- Object recognition on Kinect data
- More than 250,000 labeled RGB-D images



Prerequisites

- Probability theory
- Some linear algebra
  - Must not be afraid of eigenvalues
- Matlab, python, Java or C/C++ programming
  - This could be "language of your choice", but then you are responsible for debugging etc.
  - I suggest Matlab or python for short development time

- Your grade will be affected by any weaknesses in these

Textbooks

- Bayesian Reasoning and Machine Learning by David Barber, Cambridge University Press, 2012.

- The Elements of Statistical Learning (2nd edition) by Trevor Hastie, Robert Tibshirani and Jerome Friedman, Springer, 2009.

- Both are available online
- See http://www.cs.stevens.edu/~mordohai/classes/cs559_f16.html

Introduction

- Slides borrowed or adapted from:
  - David Barber
  - Erik Sudderth
  - Dhruv Batra
  - Pedro Domingos
  - Raquel Urtasun
  - Richard Zemel

Question 1

- What is “machine learning”?

Machine Learning

Machine learning, a branch of artificial intelligence, is a scientific discipline concerned with the design and development of algorithms that take as input empirical data, such as that from sensors or databases, and yield patterns or predictions thought to be features of the underlying mechanism that generated the data. A learner can take advantage of examples (data) to capture characteristics of interest of their unknown underlying probability distribution. Data can be seen as instances of the possible relations between observed variables. A major focus of machine learning research is the design of algorithms that recognize complex patterns and make intelligent decisions based on input data. One fundamental difficulty is that the set of all possible behaviors given all possible inputs is too large to be included in the set of observed examples (training data). Hence the learner must generalize from the given examples in order to produce a useful output in new cases.

Machine Learning

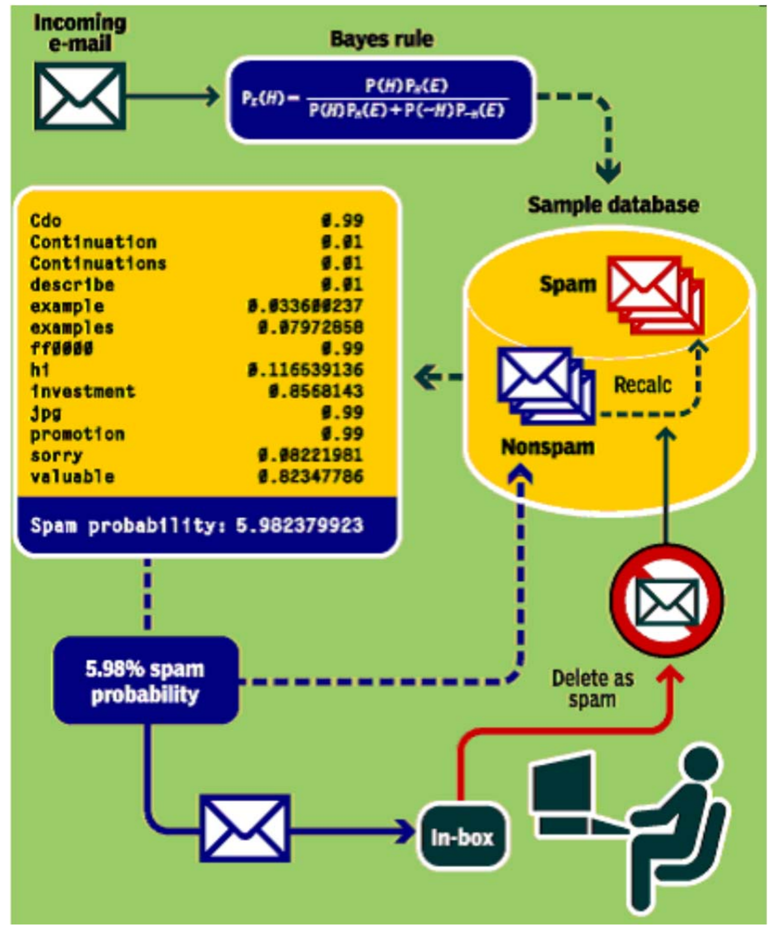
- **The Artificial Intelligence View.** Learning is central to human knowledge and intelligence, and, likewise, it is also essential for building intelligent machines. Years of effort in AI has shown that trying to build intelligent computers by programming all the rules cannot be done; automatic learning is crucial. For example, we humans are not born with the ability to understand language — we learn it — and it makes sense to try to have computers learn language instead of trying to program it all it.

Machine Learning

- **The Software Engineering View.** Machine learning allows us to program computers by example, which can be easier than writing code the traditional way.
- **The Statistics View.** Machine learning is the marriage of computer science and statistics: computational techniques are applied to statistical problems. Machine learning has been applied to a vast number of problems in many contexts, beyond the typical statistics problems. Machine learning is often designed with different considerations than statistics (e.g., speed is often more important than accuracy).

Spam Filtering

- Binary classification problem: Is this e-mail useful or spam?
- Noisy training data: Messages previously marked as spam
- Wrinkle: Spammers evolve to counter filter innovations



Movie Rating Prediction

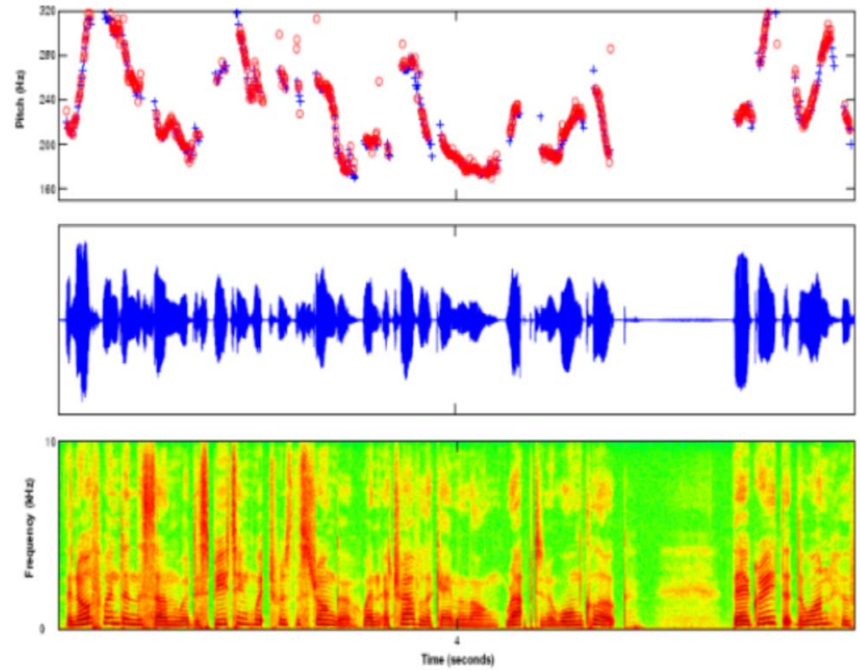
Leaderboard

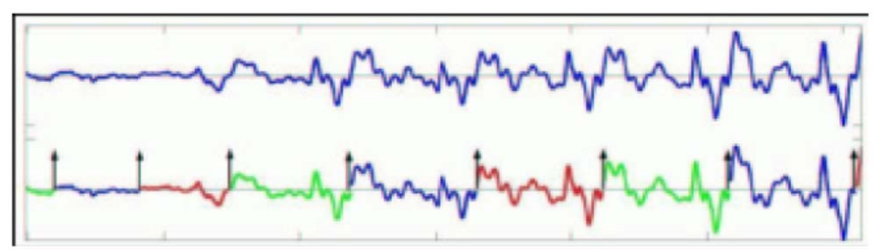
Display top 20 leaders.

Rank	Team Name	Best Score	% Improvement	Last Submit Time
1	The Ensemble	0.8553	10.10	2009-07-26 18:38:22
2	BellKor's Pragmatic Chaos	0.8554	10.09	2009-07-26 18:18:28
Grand Prize - RMSE <= 0.8563				
3	Grand Prize Team	0.8571	9.91	2009-07-24 13:07:49
4	Opera Solutions and Vandelay United	0.8573	9.89	2009-07-25 20:05:52
5	Vandelay Industries !	0.8579	9.83	2009-07-26 02:49:53
6	PragmaticTheory	0.8582	9.80	2009-07-12 15:09:53
7	BellKor in BigChaos	0.8590	9.71	2009-07-26 12:57:25
8	Dace_	0.8603	9.58	2009-07-24 17:18:43
9	Opera Solutions	0.8611	9.49	2009-07-26 18:02:08
10	BellKor	0.8612	9.48	2009-07-26 17:19:11
11	BigChaos	0.8613	9.47	2009-06-23 23:06:52
12	Feeds2	0.8613	9.47	2009-07-24 20:06:46
Progress Prize 2008 - RMSE = 0.8616 - Winning Team: BellKor in BigChaos				
13	xiangliang	0.8633	9.26	2009-07-21 02:04:40
14	Gravity	0.8634	9.25	2009-07-26 15:58:34
15	Ces	0.8642	9.17	2009-07-25 17:42:38
16	Invisible Ideas	0.8644	9.14	2009-07-20 03:26:12
17	Just a guy in a garage	0.8650	9.08	2009-07-22 14:10:42
18	Craig Carmichael	0.8656	9.02	2009-07-25 16:00:54
19	J Dennis Su	0.8658	9.00	2009-03-11 09:41:54
20	acmehill	0.8659	8.99	2009-04-16 06:29:35
Progress Prize 2007 - RMSE = 0.8712 - Winning Team: KorBell				
Cinematch score on quiz subset - RMSE = 0.9514				

Speech Recognition

- Given an audio waveform, robustly extract & recognize any spoken words
- Statistical models can be used to
  – Provide greater robustness to noise
  – Adapt to accent of different speakers
  – Learn from training





S. Roweis, 2004

What is Machine Learning?

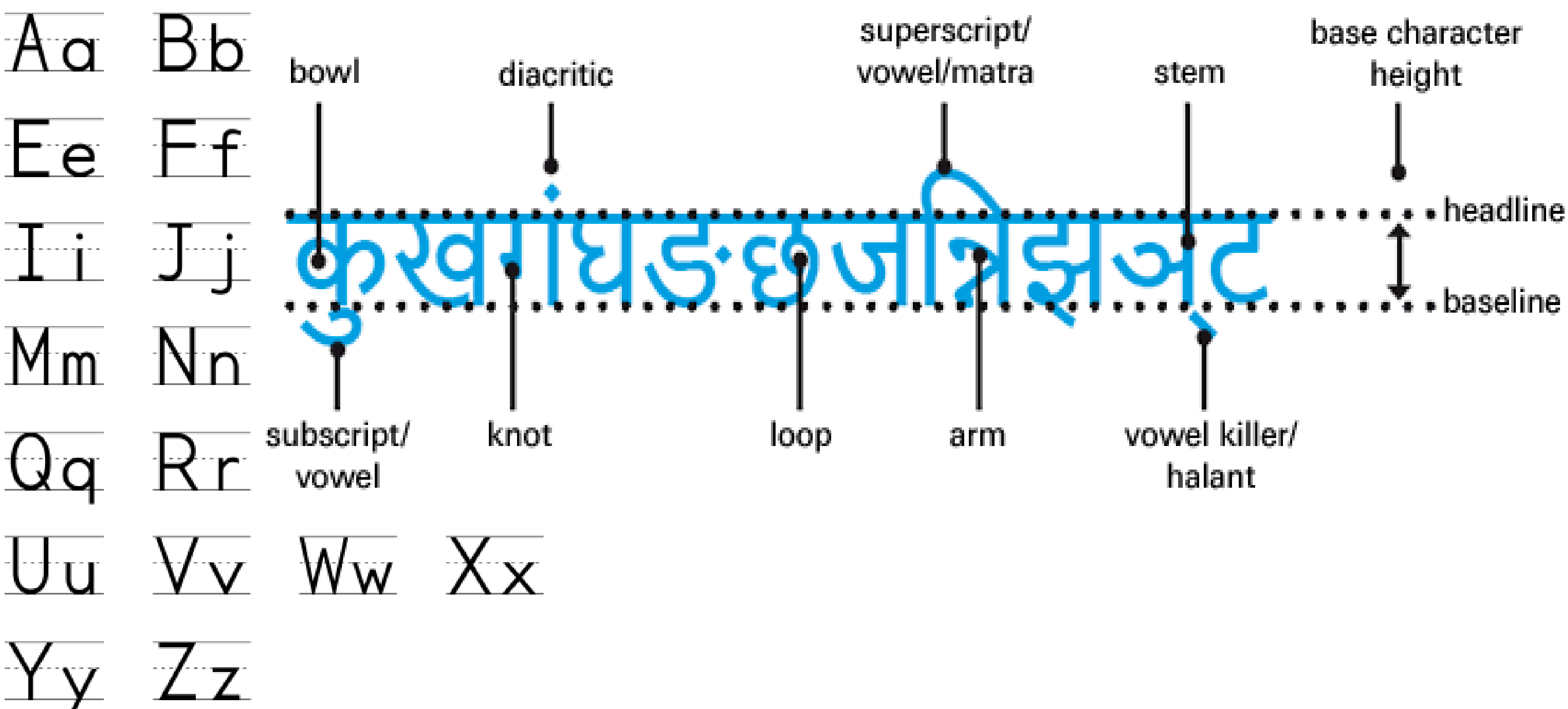
- Given a collection of examples (called "training data"), we want to predict something about novel examples
  - The novel examples are usually incomplete
- Examples:
  - Labeling: Spam or ham? How many stars?
  - Interpretation:
    - What sentence was just spoken?
    - Where are the objects moving in this video?
    - When and where have seismic events (earthquakes or explosions) occurred?

What do we actually do?

- Build idealized models of the application area we're working
  - Probabilistic models with explicit randomness
- Derive algorithms and implement in code
- Use historical data to learn numeric parameters, and sometimes model structure
- Use test data to validate the learned model, quantitatively measure its predictions
- Assess errors and repeat…

Optical Character Recognition

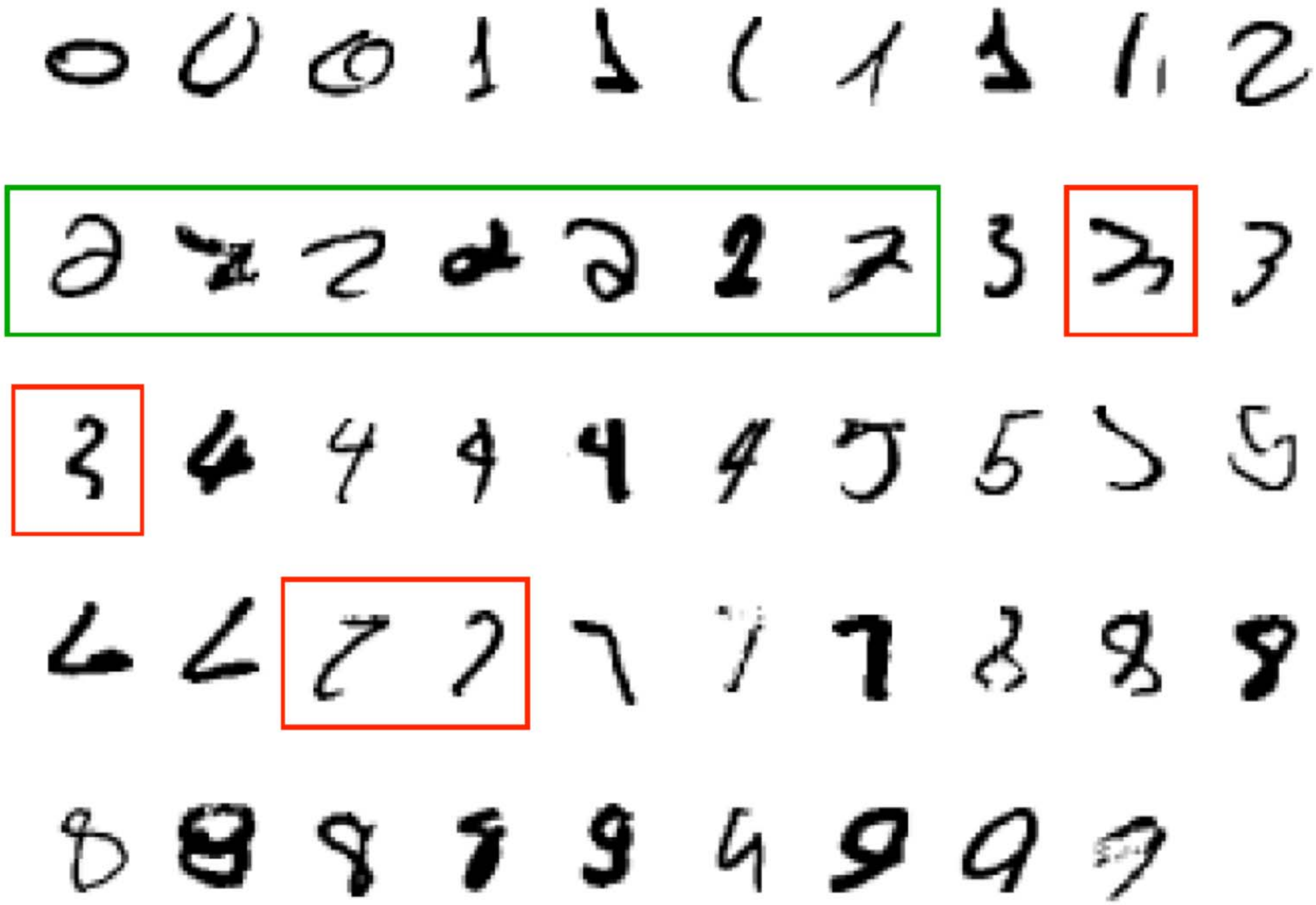
- Hard way: Understand handwriting/characters



Optical Character Recognition

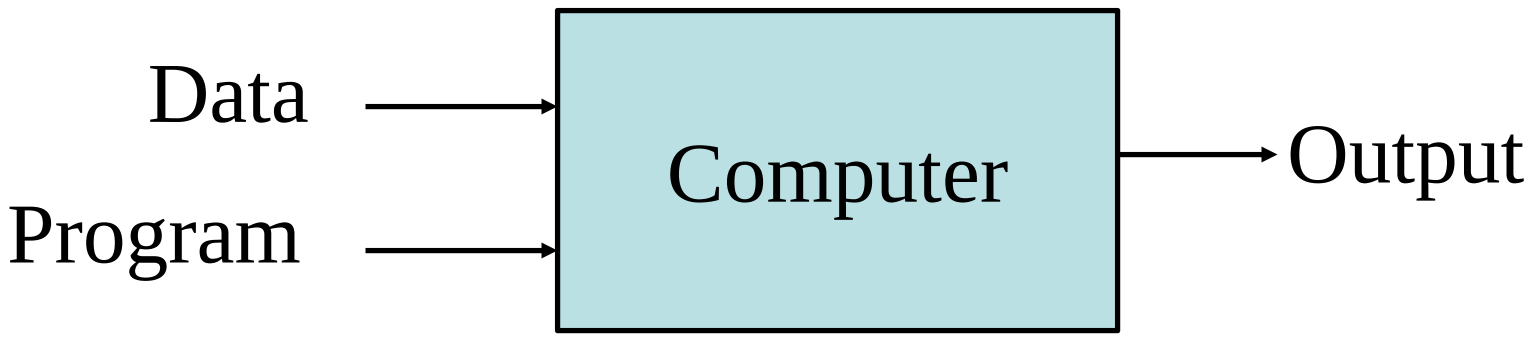
- Hard way: Understand handwriting/characters
- Lazy way: use more data!

What Makes a 2?

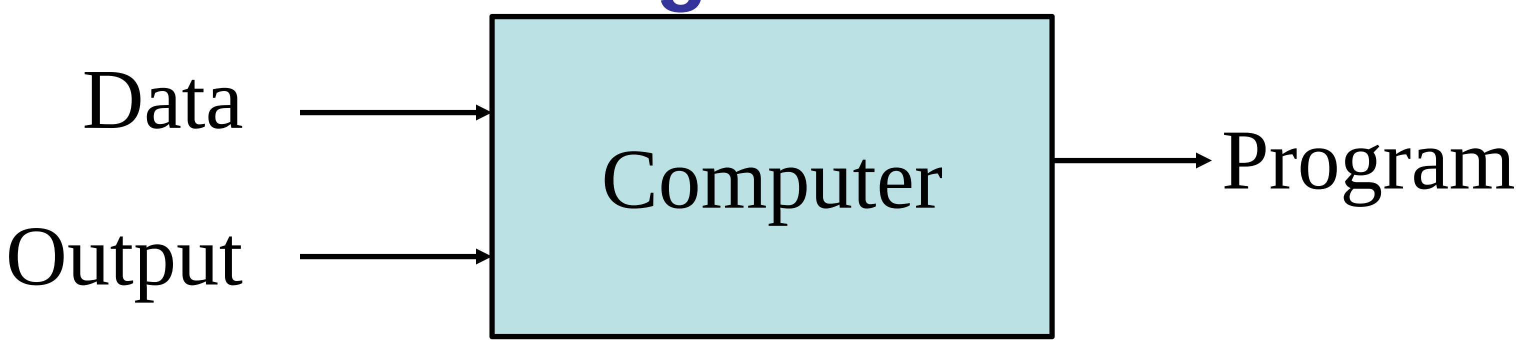


ML vs Traditional Approach

- **Traditional Programming**



- **Machine Learning**



ML in a Nutshell

- Tens of thousands of machine learning algorithms
  - Hundreds new every year

- Decades of ML research oversimplified:
  - All of Machine Learning:
  - Learn a mapping from input to output f: X → Y
  - X: emails, Y: {spam, notspam}

ML in a Nutshell

- Input: x (images, text, emails…)
- Output: y (spam or non-spam…)
- (Unknown) Target Function
  - f: X → Y (the "true" mapping / reality)
- Data
  - $(x_1,y_1), (x_2,y_2), \ldots, (x_N,y_N)$
- Model / Hypothesis
  - g: X → Y
  - $y = g(x) = \text{sign}(w^T x)$

ML in a Nutshell

- Every machine learning algorithm has three components:
  - Representation / Model Class
  - Evaluation / Objective Function
  - Optimization

Representation / Model Class

- Decision trees
- Sets of rules / Logic programs
- Instances
- Graphical models (Bayes/Markov nets)
- Neural networks
- Support vector machines
- Model ensembles
- Etc.

Evaluation / Objective Function

- Accuracy
- Precision and recall
- Squared error
- Likelihood
- Posterior probability
- Cost / Utility
- Margin
- Entropy
- K-L divergence
- Etc.

Types of Learning

- Supervised learning
  - Training data include desired outputs
  - Test data only have features, must predict outputs
- Unsupervised learning
  - Training data do not include desired outputs
- Semi-supervised learning
  - Training data include a few desired outputs
- Reinforcement learning
  - Rewards from sequence of actions
  - Out of scope in this course

Types of Learning

Supervised Learning

x → Classification → y Discrete

x → Regression → y Continuous

Unsupervised Learning

x → Clustering → y Discrete ID

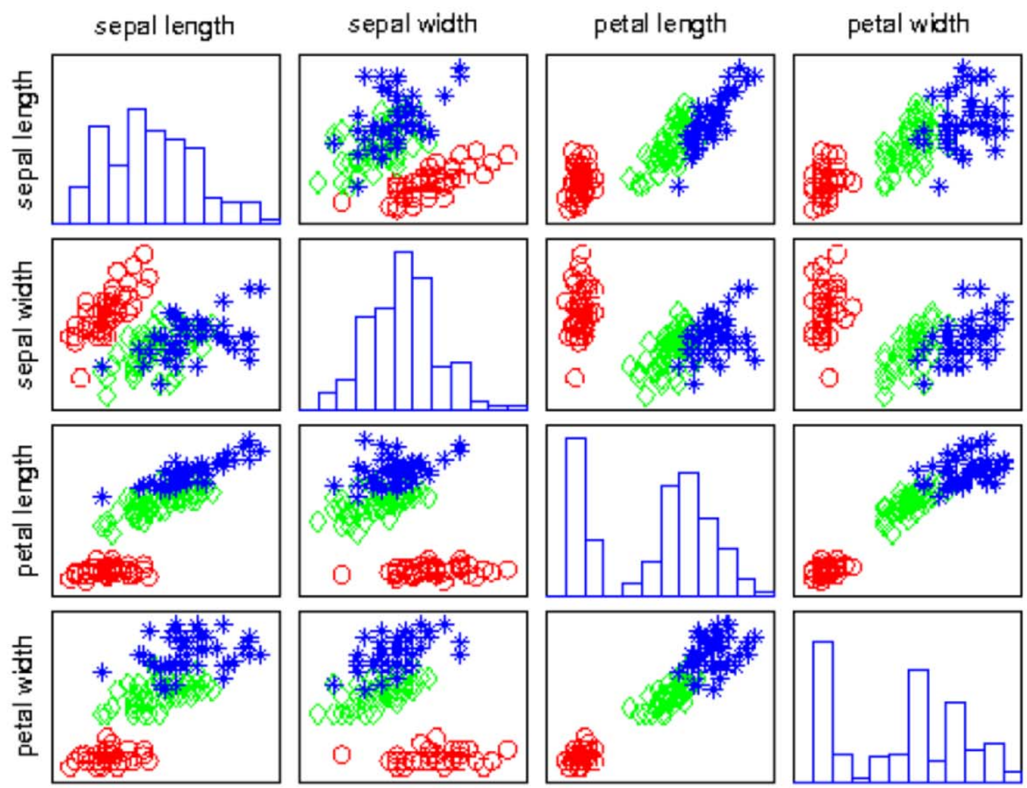
x → Dimensionality Reduction → y Continuous

Irises: Supervised Classification

setosa

versicolor

virginica

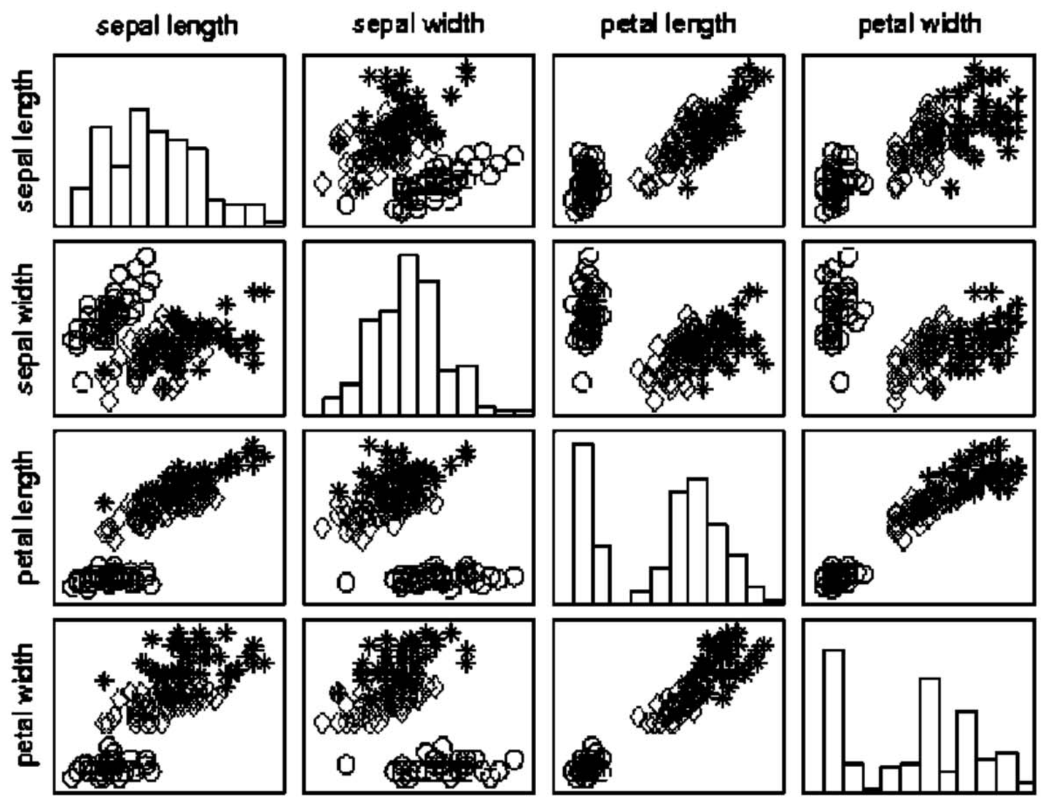


Irises: Unsupervised Classification

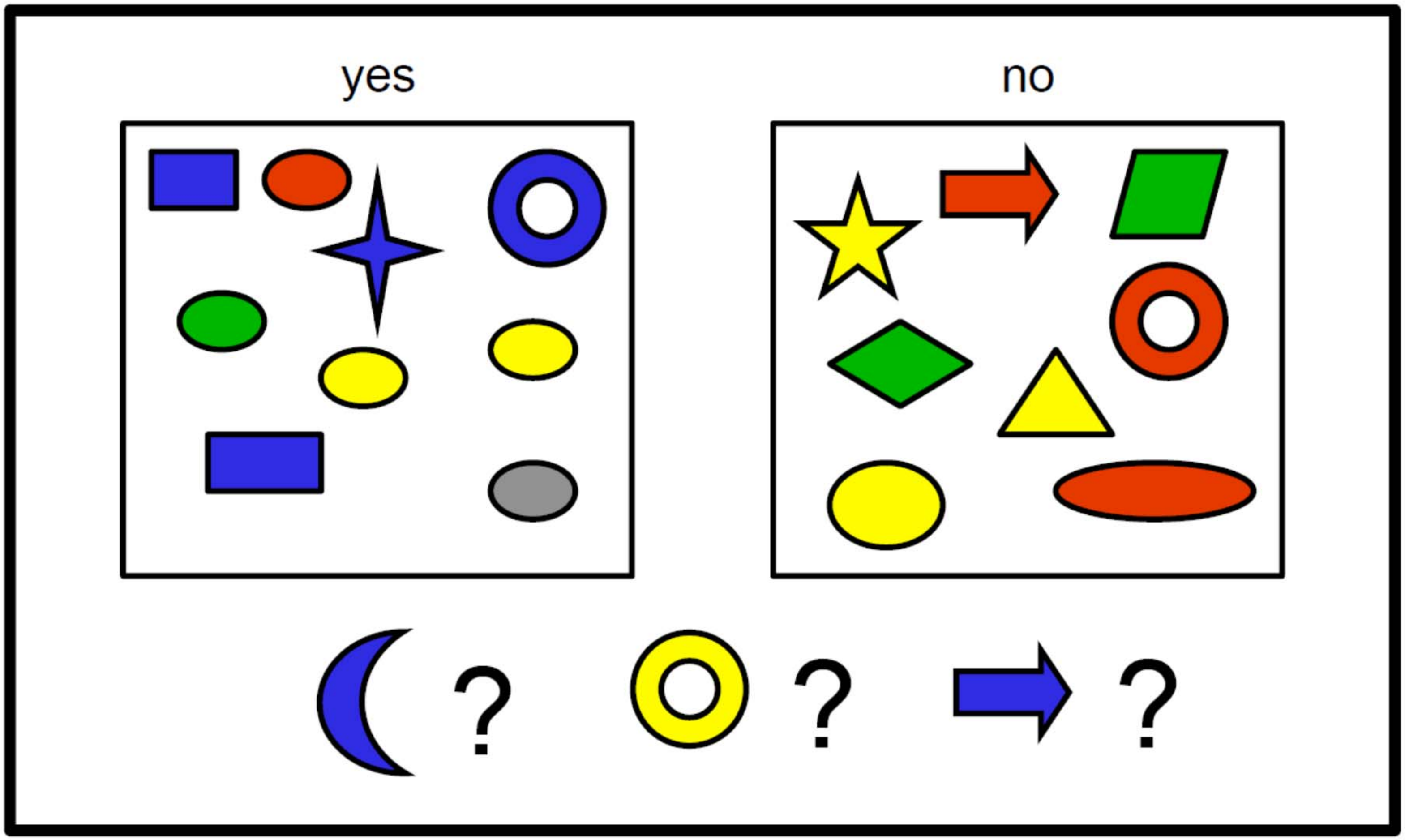
setosa

versicolor

virginica



Classification Example



Feature Encoding

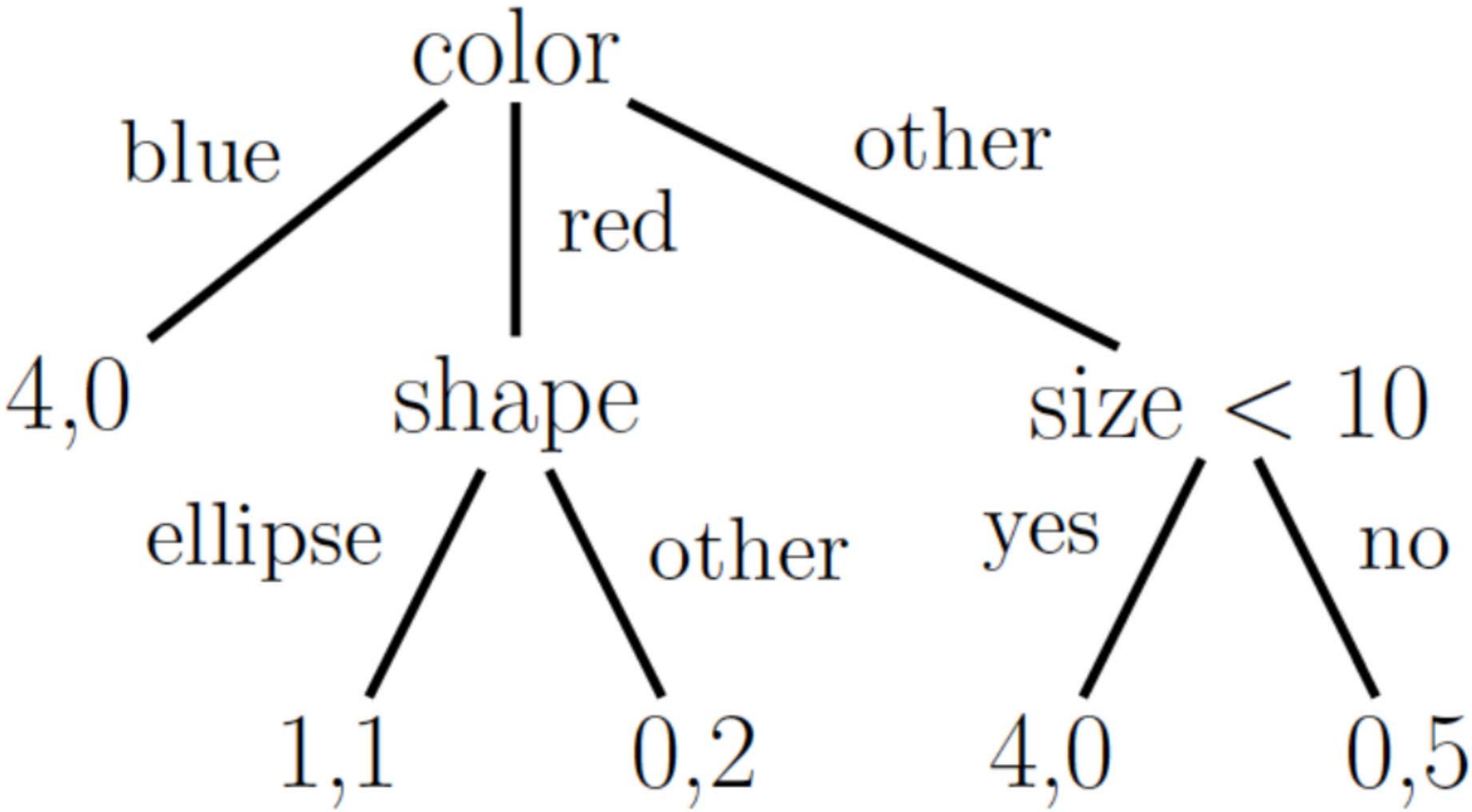
d features (attributes)

n cases

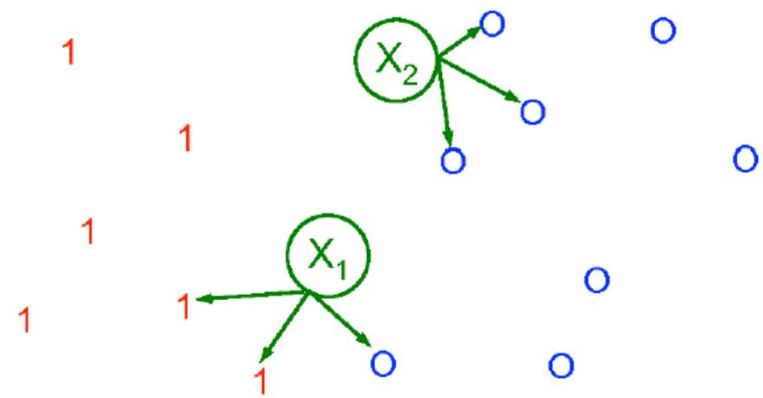
Color	Shape	Size (cm)
Blue	Square	10
Red	Ellipse	2.4
Red	Ellipse	20.7

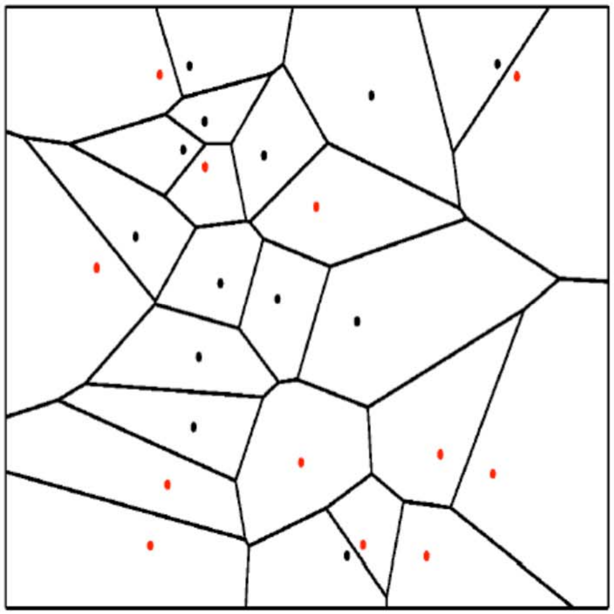
Binary Label
1
1
0

Decision Tree



Nearest Neighbor





- Define some notion of distance among input features
- For test examples, assign label of closest training example
- K-NN: Take majority vote among K closest training examples

Probability Theory Review

The Axioms of Probability

- $0 <= P(A) <= 1$
- $P(True) = 1$
- $P(False) = 0$
- $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

Overview

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

Discrete Random Variables

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

- ***Discrete*** versus ***Continuous*** ***random variable:*** an r.v. x is discrete if it can assume a finite or countably infinite number of values. An r.v. x is continuous if it can assume all values in an interval.

Examples

- Which of the following random variables are discrete and which are continuous?

- X = Number of houses sold by real estate developer per week?
- X = Number of heads in ten tosses of a coin?
- X = Weight of a child at birth?
- X = Time required to run100 yards?

Examples

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

Probability Distribution Example: *X* is the Sum of Two Dice

red	1	2	3	4	5	6

This sequence provides an example of a discrete random variable. Suppose that you have a red die which, when thrown, takes the numbers from 1 to 6 with equal probability.

Probability Distribution Example: X is the Sum of Two Dice

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

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

Probability Distribution Example: *X* is the Sum of Two Dice

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

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

Probability Distribution Example: *X* is the Sum of Two Dice

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

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

Probability Distribution Example: *X* is the Sum of Two Dice

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

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

Probability Distribution Example: X is the Sum of Two Dice

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

The table shows all the possible outcomes.

Probability Distribution Example: *X* is the Sum of Two Dice

green \ red	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

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

Probability Distribution Example: *X* is the Sum of Two Dice

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	

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

Probability Distribution Example: *X* is the Sum of Two Dice

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	

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

Probability Distribution Example: *X* is the Sum of Two Dice

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**

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

Probability Distribution Example: *X* is the Sum of Two Dice

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	

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

Probability Distribution Example: X is the Sum of Two Dice

green \ red	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	

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.

Probability Distribution Example: *X* is the Sum of Two Dice

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

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

Probability Distribution Example: *X* is the Sum of Two Dice

The distribution is shown graphically. in this example it is symmetrical, highest for *X* equal to 7 and declining on either side.

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:

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

- The expected value of a random variable, also known as its population mean, is the 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

$$E(X + Y) = E(X) + E(Y)$$
$$E(bX) = bE(X)$$
$$E(b) = b$$

$$Y = b_1 + b_2X$$
$$\begin{aligned} E(Y) &= E(b_1 + b_2X) \\ &= E(b_1) + E(b_2X) \\ &= b_1 + b_2 E(X) \end{aligned}$$

- Also denoted by μ

Variance

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

Var(X) = σ^2

Var(X) = E[(X- μ)^2] = E[X^2] - (E[X])^2

(Prove it.)

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 $p_{ij}=Pr[x=x_i, y=y_j]$
- We can also define a joint probability mass function P(x,y) which offers a complete characterization of the pair of r.v.

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

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

Marginal distributions

Note that P_x and P_y are different functions

Statistical Independence

Two random variables *x* and *y* are said to be independent, if and only if

$$P(x,y)=P_x(x)\,P_y(y)$$

that is, when knowing the value of *x* does not give us additional information for the value of *y*.

Or, equivalently

$$E[f(x)g(y)] = E[f(x)]\,E[g(y)]$$

for any functions f(x) and g(y).

Conditional Probability

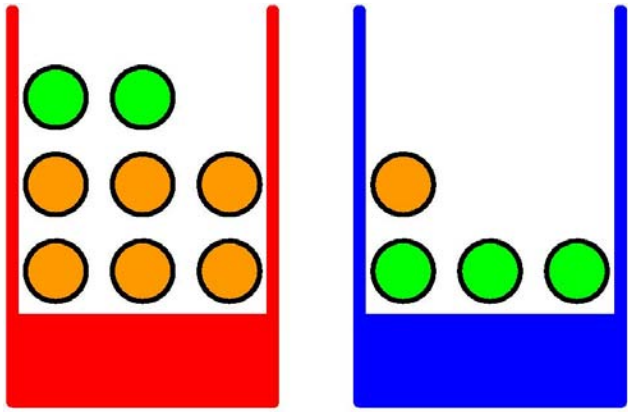
- When two r.v. are not independent, knowing one allows better estimate of the other (e.g. outside temperature, season)

$$\Pr[x = x_i \mid y = y_j] = \frac{\Pr[x = x_i , y = y_j]}{\Pr[y = y_j]}$$

- If independent P(x|y)=P(x)

Sum and Product Rules (1/7)

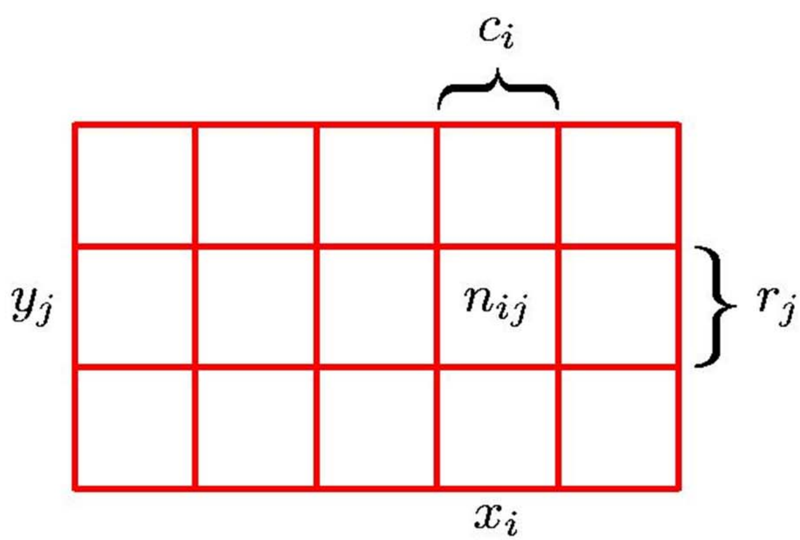
- Example:
  - We have two boxes: one red and one blue
  - Red box: 2 apples and 6 oranges
  - Blue box: 3 apples and 1 orange
  - Pick red box 40% of the time and blue box 60% of the time, then pick one item of fruit



Sum and Product Rules (2/7)

- Define:
  - B random variable for box picked (r or b)
  - F identity of fruit (a or o)
- $p(B=r)=4/10$ and $p(B=b)=6/10$
  - Events are mutually exclusive and include all possible outcomes => their probabilities must sum to 1

Sum and Product Rules (3/7)



Marginal Probability

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

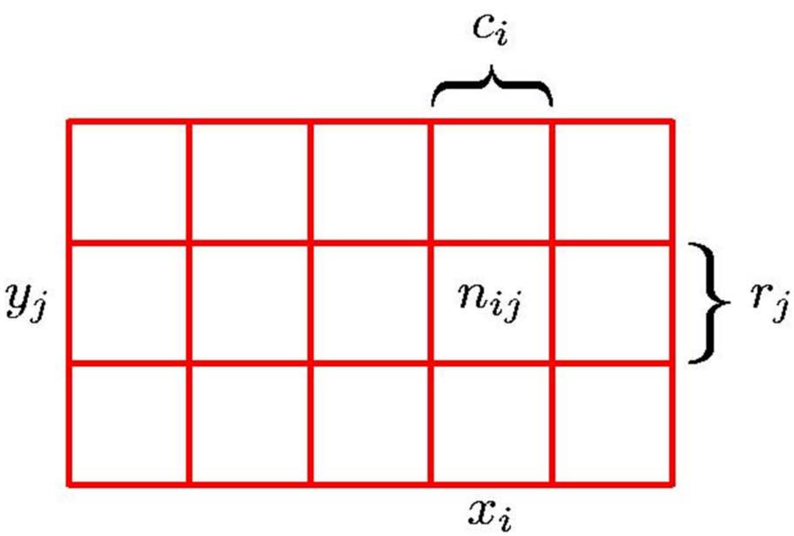
Joint Probability

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

Conditional Probability

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

Sum and Product Rules (4/7)



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

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

Sum and Product Rules (5/7)

- Sum Rule $$p(X) = \sum_Y p(X,Y)$$
- Product Rule $$p(X,Y) = p(Y|X)p(X)$$

Law of Total Probability

- If an event A can occur in *m* different ways and if these *m* different ways are mutually exclusive, then the probability of A occurring is the sum of the probabilities of the sub-events

$$P(X = x_i) = \sum_j P(X = x_i \mid Y = y_j) P(Y = y_j)$$

Sum and Product Rules (6/7)

- Back to the fruit baskets
  - $p(B=r)=4/10$ and $p(B=b)=6/10$
  - $p(B=r) + p(B=b) = 1$
- Conditional probabilities
  - $p(F=a \mid B = r) = 1/4$
  - $p(F=o \mid B = r) = 3/4$
  - $p(F=a \mid B = b) = 3/4$
  - $p(F=o \mid B = b) = 1/4$

Sum and Product Rules (7/7)

Note: $p(F=a \mid B = r) + p(F=o \mid B = r) = 1$

$$p(F=a) = p(F=a \mid B = r)\, p(B=r) + p(F=a \mid B = b)\, p(B=b)$$
$$= 1/4 * 4/10 + 3/4 * 6/10 = 11/20$$

Sum rule: $p(F=o) = ?$

Conditional Probability Example

- A jar contains black and white marbles.
- Two marbles are chosen without replacement.
- The probability of selecting a black marble and then a white marble is 0.34.
- The probability of selecting a black marble on the first draw is 0.47.

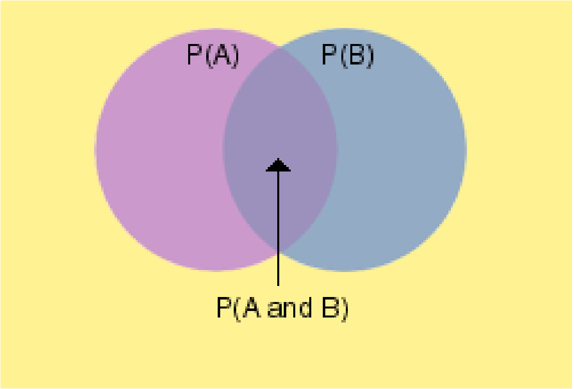
➢ What is the probability of selecting a white marble on the second draw, given that the first marble drawn was black?

Conditional Probability Example

- A jar contains black and white marbles.
- Two marbles are chosen without replacement.
- The probability of selecting a black marble and then a white marble is 0.34.
- The probability of selecting a black marble on the first draw is 0.47.

➢ What is the probability of selecting a white marble on the second draw, given that the first marble drawn was black?

$$P(White \mid Black) = \frac{P(Black \wedge White)}{P(Black)} = \frac{0.34}{0.47} = 0.72$$



A is black in first draw, B is white in second draw

Law of Total Probability

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

$$P(x \mid y) = \frac{P(x, y)}{P(y)}$$

Bayes Rule

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

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

- *x* is the unknown cause
- y is the observed evidence
- Bayes rule shows how probability of *x* changes after we have observed *y*

Bayes Rule on the Fruit Example

- Suppose we have selected an orange. Which box did it come from?

$$p(B=r \mid F=o) = \frac{p(F=o \mid B=r)\,p(B=r)}{p(F=o)} = \frac{\frac{3}{4}\times\frac{4}{10}}{\frac{9}{20}} = \frac{2}{3}$$

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, weight of child at birth…
- Cannot talk about probability of that x has a particular value
- Instead, probability that x falls in an interval => probability density function

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

$$p(x) \geq 0 \text{ and } \int_{-\infty}^{\infty} p(x)dx = 1$$

Expected Value

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

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

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

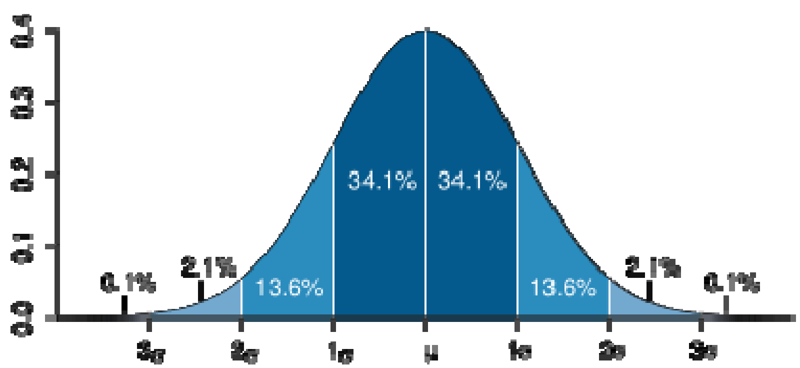
- Bayes rule $p(x \mid y) = \dfrac{p(y \mid x) p(x)}{\int_{-\infty}^{\infty} p(y \mid x) p(x) dx}$

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

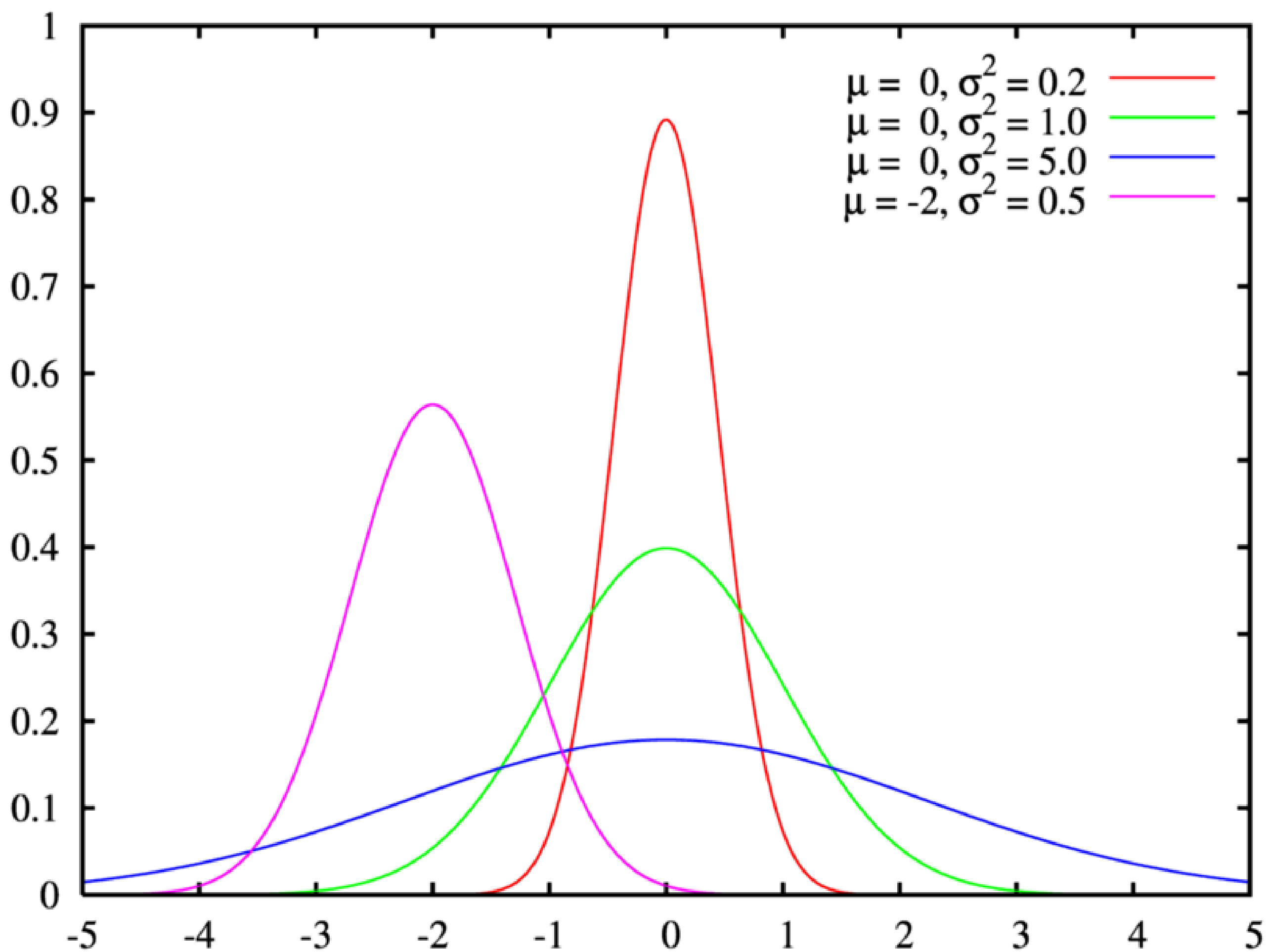
Normal (Gaussian) Distribution

- Central Limit Theorem: under various conditions, the distribution of the sum of *d* independent random variables approaches a limiting form known as the normal distribution

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



Normal (Gaussian) Distribution



Uniform Distribution

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