Indian Institute of Information Technology, Allahabad

CVBL IIIT Allahabad

Introduction to Machine Learning

Introduction

By

Dr. Shiv Ram Dubey
Associate Professor
Computer Vision and Biometrics Lab
Department of Information Technology
Indian Institute of Information Technology, Allahabad
Web: https://profile.iiita.ac.in/srdubey/

DISCLAIMER

The content (text, image, and graphics) used in this slide are adopted from many sources for academic purposes. Broadly, the sources have been given due credit appropriately. However, there is a chance of missing out some original primary sources. The authors of this material do not claim any copyright of such material.

Course Website

https://profile.iiita.ac.in/srdubey/teaching/iml2026/

Everything you need to know is there:

- Syllabus
- Lecture Slides
- TA Details
- Textbook and other references
- Grading, Course Ethics, etc.

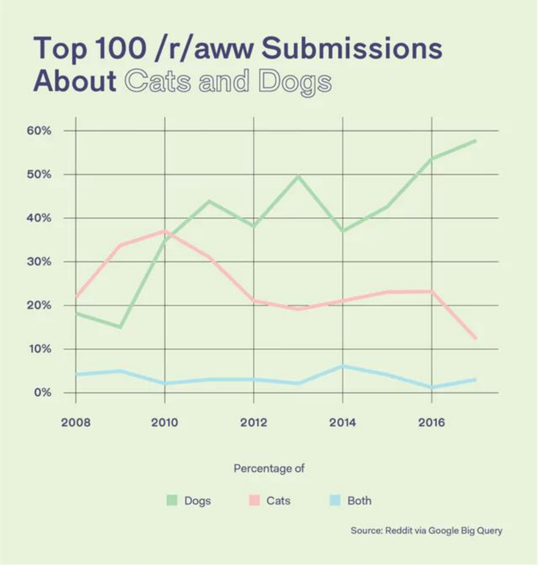
Traditional Algorithms

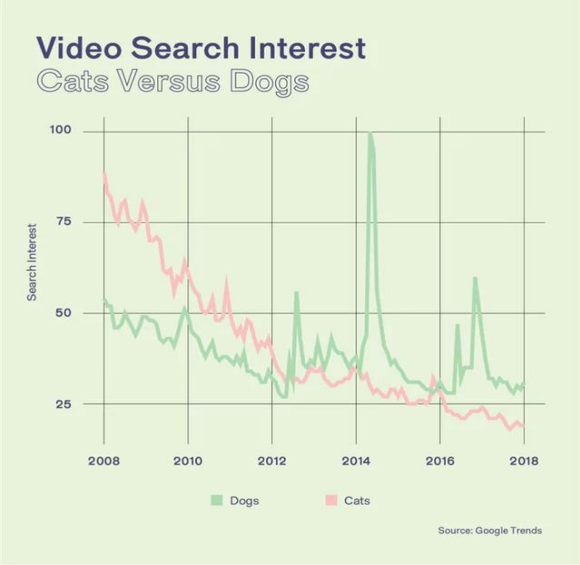
Social media mentions of Cats vs. Dogs

Reddit

Google

Twitter?





Graphics courtesy of https://theoutline.com/post/3128/dogs-cats-internet-popularity?zd=1

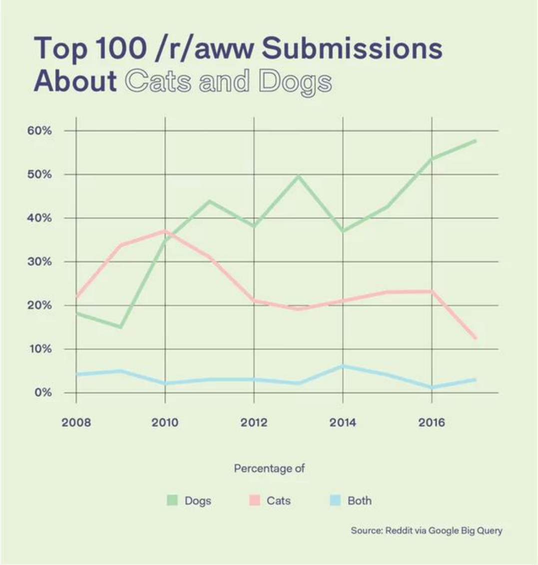
Traditional Algorithms

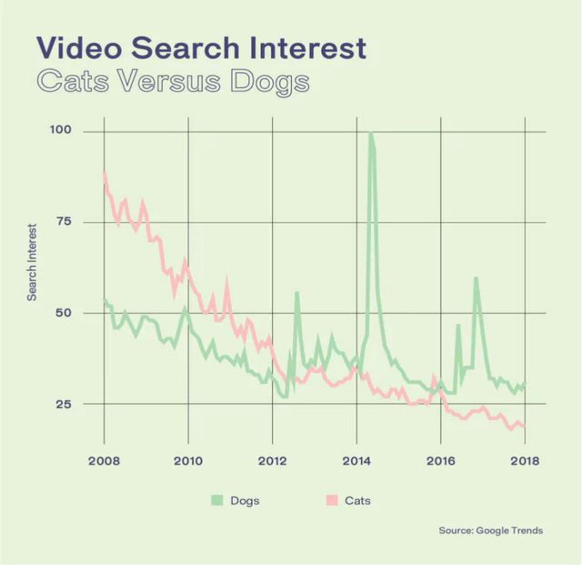
Social media mentions of Cats vs. Dogs

Reddit

Google

Twitter?





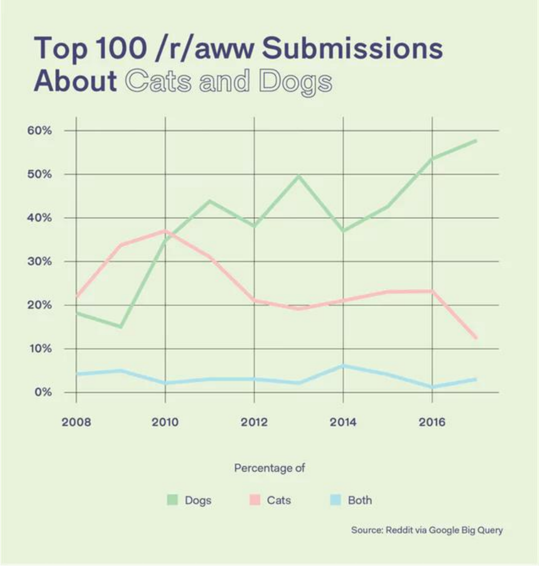
Write a program that sorts tweets into those containing “**cat**”, “**dog**”, or ***other***

Graphics courtesy of https://theoutline.com/post/3128/dogs-cats-internet-popularity?zd=1

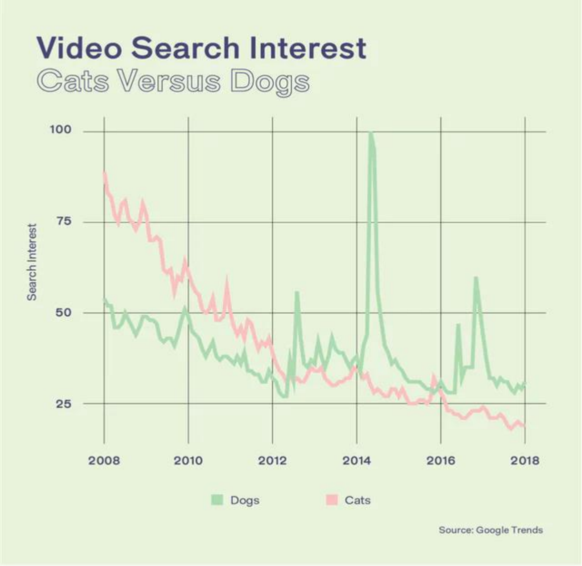
Traditional Algorithms

Social media mentions of Cats vs. Dogs

Reddit



Google



Twitter?

```
cats = []
dogs = []
other = []
for tweet in tweets:
    if "cat" in tweet:
        cats.append(tweet)
    elif "dog" in tweet:
        dogs.append(tweet)
    else:
        other.append(tweet)
return cats, dogs, other
```

Write a program that sorts tweets into those containing “**cat**”, “**dog**”, or ***other***

Graphics courtesy of https://theoutline.com/post/3128/dogs-cats-internet-popularity?zd=1

Machine Learning Algorithms

Write a program that sorts images into those containing “**birds**”, “**airplanes**”, or ***other.***

airplane

other

bird

Machine Learning Algorithms

Write a program that sorts images into those containing "**birds**", "**airplanes**", or ***other.***

airplane

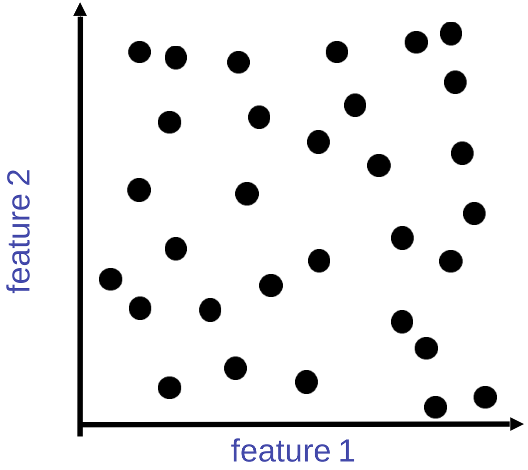
other

bird

```
birds = []
planes = []
other = []
for image in images:
    if bird in image:
        birds.append(image)
    elif plane in image:
        planes.append(image)
    else:
        other.append(image)
return birds, planes, other
```

Machine Learning Algorithms

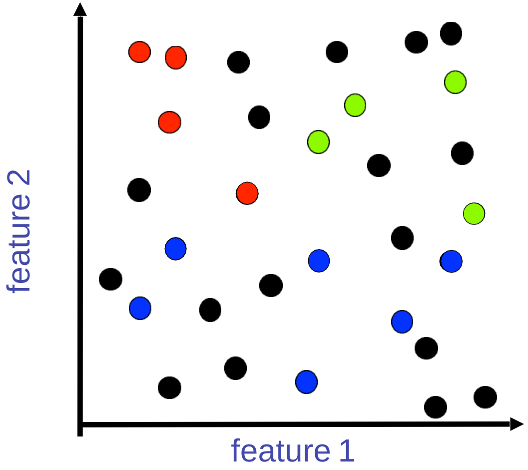
Write a program that sorts images into those containing "**birds**", "**airplanes**", or ***other.***




```
birds = []
planes = []
other = []
for image in images:
    if bird in image:
        birds.append(image)
    elif plane in image:
        planes.append(image)
    else:
        other.append(image)
return birds, planes, other
```

Machine Learning Algorithms

Write a program that sorts images into those containing "**birds**", "**airplanes**", or ***other.***

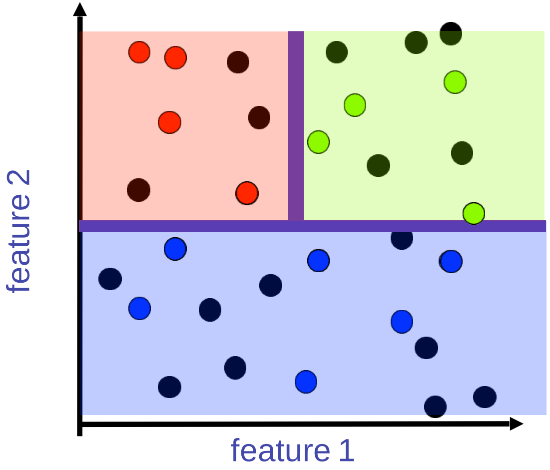



```
birds = []
planes = []
other = []
for image in images:
    if bird in image:
        birds.append(image)
    elif plane in image:
        planes.append(image)
    else:
        other.append(image)
return birds, planes, other
```

Machine Learning Algorithms

Write a program that sorts images into those containing "**birds**", "**airplanes**", or ***other.***


```
birds = []
planes = []
other = []
for image in images:
    if bird in image:
        birds.append(image)
    elif plane in image:
        planes.append(image)
    else:
        other.append(image)
return birds, planes, other
```



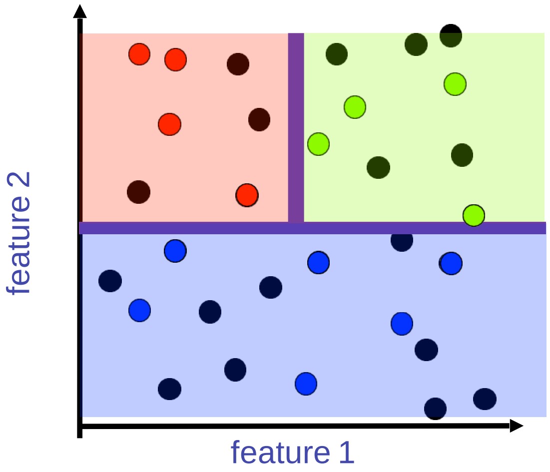
Machine Learning Algorithms

Write a program that sorts images into those containing "**birds**", "**airplanes**", or ***other.***

airplane

other

bird




```
birds = []
planes = []
other = []
for image in images:
    if bird in image:
        birds.append(image)
    elif plane in image:
        planes.append(image)
    else: …
```

The decision rule of

```
if "cat" in tweet:
```

is **hard coded by expert.**

The decision rule of

```
if bird in image:
```

is **LEARNED using DATA**

Machine Learning

- Grew out of work in AI
- New capability for computers

Machine Learning

- Grew out of work in AI
- New capability for computers

Examples:

- Database mining

  Large datasets from growth of automation/web.

  E.g., Web click data, medical records, biology, engineering
- Applications can’t program by hand.

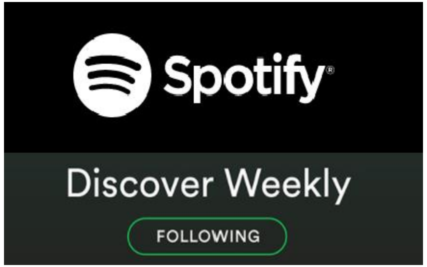
  E.g., Autonomous helicopter, handwriting recognition, most of Natural Language Processing (NLP), Computer Vision.

Machine Learning

- Grew out of work in AI
- New capability for computers

Examples:

- Database mining
  - Large datasets from growth of automation/web.
  - E.g., Web click data, medical records, biology, engineering
- Applications can't program by hand.
  - E.g., Autonomous helicopter, handwriting recognition, most of Natural Language Processing (NLP), Computer Vision.
- Self-customizing programs
  - E.g., Amazon, Netflix product recommendations
- Understanding human learning (brain, real AI).



amazon prime

Even Before LLMs Like Claude, GPT, Gemini, You Were Already Interacting With ML Algorithms Every Day

YouTube

SPAM

Google

Machine Learning definition

- **Arthur Samuel (1959):** Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.

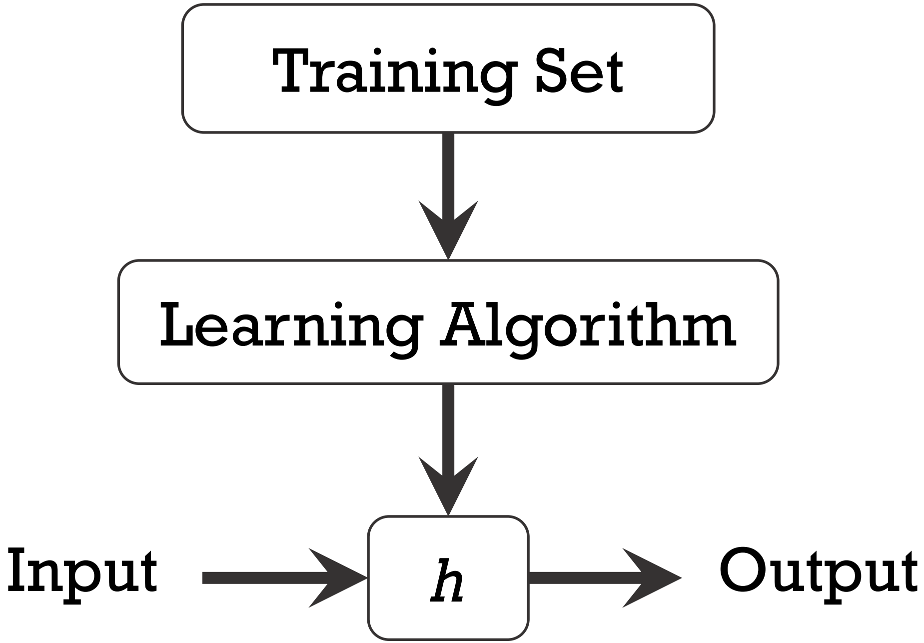
Machine Learning definition

- **Arthur Samuel (1959):** Machine Learning: Field of study that gives computers the ability to learn without being explicitly programmed.

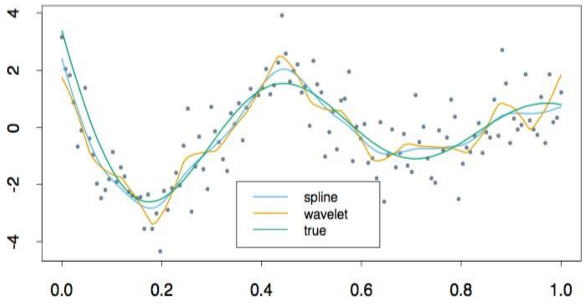
- **Tom Mitchell (1998):** Well-posed Learning Problem: A computer program is said to *learn* from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E.

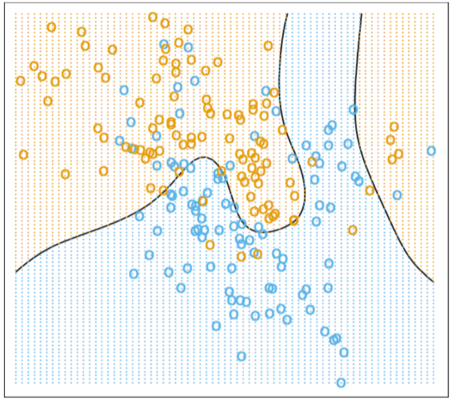
Machine Learning Ingredients

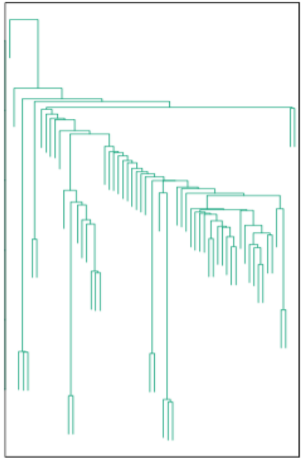
- **Data**: past observations
- **Hypotheses/Models**: devised to capture the patterns in data
- **Prediction**: apply model to forecast future observations



Flavors of ML







Regression

Predict continuous value:
ex: stock market, credit score, temperature, Netflix rating

Classification

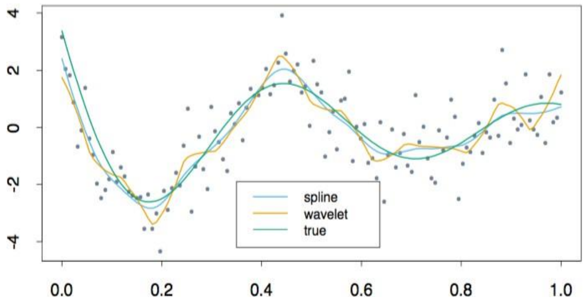
Predict categorical value:
loan or not? spam or not? what disease is this?

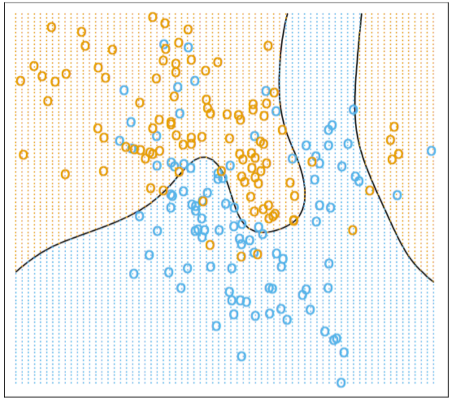
Unsupervised Learning

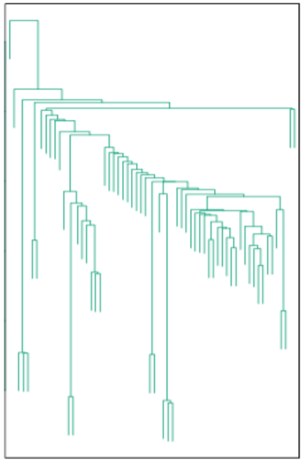
Predict structure:
tree of life from DNA, find similar images, community detection

Self-supervised:
Model distribution of large-scale data, like text or images (large vision and language models)

Flavors of ML







Regression

Predict continuous value:

ex: stock market, credit score, temperature, Netflix rating

Classification

Predict categorical value:

loan or not? spam or not? what disease is this?

Unsupervised Learning

Predict structure:

tree of life from DNA, find similar images, community detection

Self-supervised:

Model distribution of large-scale data, like text or images (large vision and language models)

Mix of -

- **statistics (conceptual) and**
- **algorithms (programming)**

What this class is:

- **Fundamentals of ML:** bias/variance tradeoff, overfitting, optimization and computational tradeoffs, supervised learning (e.g., linear, boosting, deep learning), unsupervised models (e.g. k-means, PCA)

- **Preparation for further learning:** the field is fast-moving, you will learn the foundations of ML to understand the latest results

- **ML is:**
  - Becoming ubiquitous in science, engineering and beyond
  - Transforming the world

- This class should give you a basic foundation for understanding and applying ML

Prerequisites

- Familiarity with:
  - Linear algebra
    - linear dependence, rank, linear equations
  - Multivariate calculus
  - Probability and statistics
    - Distributions, densities, marginalization, moments
  - Algorithms
    - Basic data structures, complexity

Grading

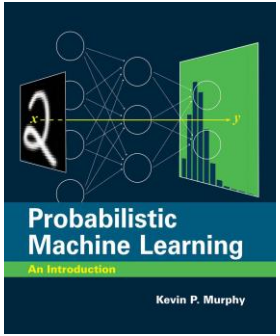
- Class Participation (5 Marks)
- Assignments (15 Marks)
- Project (15 Marks)
- Midterm Exam (25 Marks)
- Final Exam (40 Marks)

Textbooks

- **Free PDF** Textbook:

Probabilistic Machine Learning: An Introduction

Kevin Murphy PDF, also in print



- Tom M. Mitchell, Machine Learning, 1st Edition, McGraw-Hill Education, 1997.
- Christopher M. Bishop, Pattern Recognition and Machine Learning, 1st Edition, Springer, 2006.
- Hal Daume III, A Course in Machine Learning, 2017.

Reference Books

- Trevor Hastie, Robert Tibshirani and Jerome Friedman, The Elements of Statistical Learning, 2nd Edition, Springer, 2009.
- Ethem Alpaydin, Introduction to Machine Learning, 4th Edition, MIT Press, 2020.
- Ian Goodfellow, Yoshua Bengio and Aaron Courville, Deep Learning, 1st Edition, MIT Press, 2016.

Probability Review

Definitions

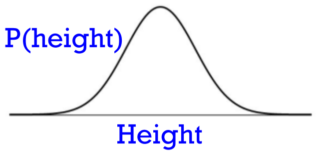
- **Random Variable**: A variable that takes on different values determined randomly.
  - Example: The height of a person from the India $P(\quad)$

Definitions

- **Random Variable:** A variable that takes on different values determined randomly.
  - Example: The height of a person from the India

P( )

- **Distribution**: The different values a random variable can take on along with the probability of that value.

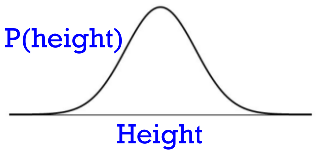


Definitions

- **Random Variable**: A variable that takes on different values determined randomly.
  - Example: The height of a person from the India

P( )

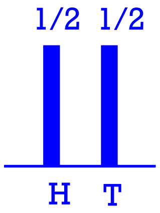
- **Distribution**: The different values a random variable can take on along with the probability of that value.



For a coin toss

P(X=H) = 1/2

P(X=T) = 1/2

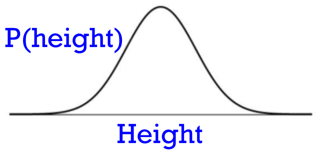


Definitions

- **Random Variable**: A variable that takes on different values determined randomly.
  - Example: The height of a person from the India

$P(\ \ )$

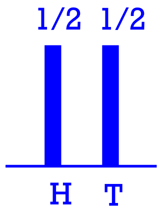
- **Distribution**: The different values a random variable can take on along with the probability of that value.



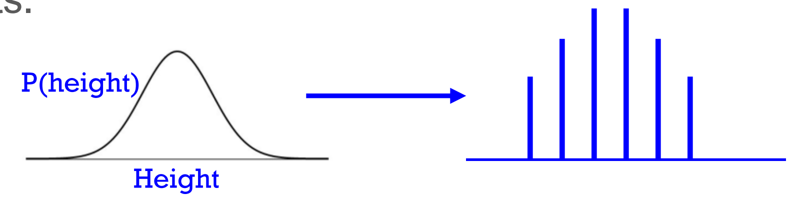
For a coin toss

$P(X=H) = 1/2$

$P(X=T) = 1/2$



- We talk about **sampling** from a distribution:
  - "Consider a sample of 100 different heights of people from the India drawn randomly from the distribution of all heights."



Independence

Let X and Y be **random variables**

Ex. X is the outcome of the first roll of a 6-sided dice,

Y is the outcome of the second roll of the dice

(X and Y take values in {1,2,3,4,5,6} each with equal probability)

Independence

Let X and Y be **random variables**

Ex. X is the outcome of the first roll of a 6-sided dice,

Y is the outcome of the second roll of the dice

(X and Y take values in {1,2,3,4,5,6} each with equal probability)

An **event** is statement about the world that holds or not:

Define events $A = \{X \in \{3,4\}\}$,

$B = \{X = 1\}$,

$C = \{Y \in \{3,4\}\}$

Independence

Let X and Y be **random variables**

Ex. X is the outcome of the first roll of a 6-sided dice,

Y is the outcome of the second roll of the dice

(X and Y take values in {1,2,3,4,5,6} each with equal probability)

An **event** is statement about the world that holds or not:

Define events $A = \{X \in \{3,4\}\}$,

$B = \{X = 1\}$,

$C = \{Y \in \{3,4\}\}$

Every event is assigned a **probability**:

$P(A) = P(X \in \{3,4\}) =$ 2/6 = 1/3

Independence

Let X and Y be **random variables**

Ex. X is the outcome of the first roll of a 6-sided dice,

Y is the outcome of the second roll of the dice

(X and Y take values in {1,2,3,4,5,6} each with equal probability)

An **event** is statement about the world that holds or not:

Define events $A = \{X \in \{3,4\}\}$,

$B = \{X = 1\}$,

$C = \{Y \in \{3,4\}\}$

Every event is assigned a **probability**:

$P(A) = P(X \in \{3,4\}) =$ 1/3

For any events U, V we have $P(U \cup V) = P(U) + P(V) - P(U \cap V)$

Independence

Let X and Y be **random variables**

Ex. X is the outcome of the first roll of a 6-sided dice,

Y is the outcome of the second roll of the dice

(X and Y take values in {1,2,3,4,5,6} each with equal probability)

An **event** is statement about the world that holds or not:

Define events $A = \{X \in \{3,4\}\}$,

$B = \{X = 1\}$,

$C = \{Y \in \{3,4\}\}$

$$P(U \cup V) = P(U) + P(V) - P(U \cap V)$$

Any events U, V are **independent** if $P(U \cap V) = P(U)P(V)$

Independence

Let X and Y be **random variables**

Ex. X is the outcome of the first roll of a 6-sided dice,

Y is the outcome of the second roll of the dice

(X and Y take values in {1,2,3,4,5,6} each with equal probability)

An **event** is statement about the world that holds or not:

Define events $A = \{X \in \{3,4\}\},$

$B = \{X = 1\},$

$C = \{Y \in \{3,4\}\}$

$P(U \cup V) = P(U) + P(V) - P(U \cap V)$

Any events U, V are **independent** if $P(U \cap V) = P(U)P(V)$

Are A, B independent?

Independence

Let X and Y be **random variables**

Ex. X is the outcome of the first roll of a 6-sided dice,

Y is the outcome of the second roll of the dice

(X and Y take values in {1,2,3,4,5,6} each with equal probability)

An **event** is statement about the world that holds or not:

Define events $A = \{X \in \{3,4\}\}$,

$B = \{X = 1\}$,

$C = \{Y \in \{3,4\}\}$

$P(U \cup V) = P(U) + P(V) - P(U \cap V)$

Any events U, V are **independent** if $P(U \cap V) = P(U)P(V)$

Are A, B independent? $P(A) = 1/3$ $P(B) = 1/6$ $P(A \cap B) = 0$ No

Independence

Let X and Y be **random variables**

Ex. X is the outcome of the first roll of a 6-sided dice,

Y is the outcome of the second roll of the dice

(X and Y take values in {1,2,3,4,5,6} each with equal probability)

An **event** is statement about the world that holds or not:

Define events $A = \{X \in \{3,4\}\},$

$B = \{X = 1\},$

$C = \{Y \in \{3,4\}\}$

$P(U \cup V) = P(U) + P(V) - P(U \cap V)$

Any events U, V are **independent** if $P(U \cap V) = P(U)P(V)$

Are A, B independent? $P(A) = 1/3$ $P(B) = 1/6$ $P(A \cap B) = 0$ No

B, C?

Independence

Let X and Y be **random variables**

Ex. X is the outcome of the first roll of a 6-sided dice,

Y is the outcome of the second roll of the dice

(X and Y take values in {1,2,3,4,5,6} each with equal probability)

An **event** is statement about the world that holds or not:

Define events $A = \{X \in \{3,4\}\}$,

$B = \{X = 1\}$,

$C = \{Y \in \{3,4\}\}$

$P(U \cup V) = P(U) + P(V) - P(U \cap V)$

Any events U, V are **independent** if $P(U \cap V) = P(U)P(V)$

Are A, B independent? $P(A) = 1/3$ $P(B) = 1/6$ $P(A \cap B) = 0$ No

B, C? $P(B) = 1/6$ $P(C) = 1/3$ $P(B \cap C) = 1/3 \times 1/6$ Yes

Independence

Let X and Y be **random variables**

Ex. X is the outcome of the first roll of a 6-sided dice,

Y is the outcome of the second roll of the dice

(X and Y take values in {1,2,3,4,5,6} each with equal probability)

An **event** is statement about the world that holds or not:

Define events $A = \{X \in \{3,4\}\}$,

$B = \{X = 1\}$,

$C = \{Y \in \{3,4\}\}$

$P(U \cup V) = P(U) + P(V) - P(U \cap V)$

Any events U, V are **independent** if $P(U \cap V) = P(U)P(V)$

Are A, B independent? $P(A) = 1/3$ $P(B) = 1/6$ $P(A \cap B) = 0$ *No*

B, C? $P(B) = 1/6$ $P(C) = 1/3$ $P(B \cap C) = 1/3 \times 1/6$ *Yes*

A, C?

Independence

Let X and Y be **random variables**

Ex. X is the outcome of the first roll of a 6-sided dice,

Y is the outcome of the second roll of the dice

(X and Y take values in {1,2,3,4,5,6} each with equal probability)

An **event** is statement about the world that holds or not:

Define events $A = \{X \in \{3,4\}\}$,

$B = \{X = 1\}$,

$C = \{Y \in \{3,4\}\}$

$P(U \cup V) = P(U) + P(V) - P(U \cap V)$

Any events U, V are **independent** if $P(U \cap V) = P(U)P(V)$

Are A, B independent?	$P(A) = 1/3$	$P(B) = 1/6$	$P(A \cap B) = 0$	No
B, C?	$P(B) = 1/6$	$P(C) = 1/3$	$P(B \cap C) = 1/3 \times 1/6$	Yes
A, C?	$P(A) = 1/3$	$P(C) = 1/3$	$P(A \cap C) = 1/3 \times 1/3$	Yes

Independence

Let X and Y be **random variables**

Ex. X is the outcome of the first roll of a 6-sided dice,

Y is the outcome of the second roll of the dice

(X and Y take values in {1,2,3,4,5,6} each with equal probability)

An **event** is statement about the world that holds or not:

Define events $A = \{X \in \{3,4\}\},$

$B = \{X = 1\},$

$C = \{Y \in \{3,4\}\}$

$P(U \cup V) = P(U) + P(V) - P(U \cap V)$

Any events U, V are **independent** if $P(U \cap V) = P(U)P(V)$

We define the **conditional probability** of event U given V as

$$P(U \mid V) = \frac{P(U \cap V)}{P(V)}$$

If independent then: $= \frac{P(U)\cancel{P(V)}}{\cancel{P(V)}} = P(U)$

Observe: if U, V are independent then $P(U \mid V) = P(U)$.

In words: if independent, V tells you nothing about U (and vice versa)

Mean

Mean $\mathbb{E}[X], \mu$ The expected value of X, each value is weighted by the probability of seeing it.

$$\mathbb{E}[X] = \sum_x P(X = x)x$$

Mean, Variance

Mean $\mathbb{E}[X], \mu$ The expected value of X, each value is weighted by the probability of seeing it.

$$\mathbb{E}[X] = \sum_x P(X = x)x$$

Variance $\text{Var}(X), \sigma^2$ The expected squared deviation of X from its mean.

$$\mathbb{E}[(X - \mathbb{E}[X])^2]$$

Mean, Variance, Median

Mean $\mathbb{E}[X], \mu$ The expected value of X, each value is weighted by the probability of seeing it.

$$\mathbb{E}[X] = \sum_x P(X = x)x$$

Variance $\mathrm{Var}(X), \sigma^2$ The expected squared deviation of X from its mean.

$$\mathbb{E}[(X - \mathbb{E}[X])^2]$$

Median M The value of X that is separating the higher half of it's range from the lower half.

$$P(X \leq M) = .5$$

Mean, Variance

Mean: The mean is a prediction of the value of the random variable.

Answers the question "What do I expect the height of a random person to be?"

Variance: The variance captures the spread in your data.

$$\mathbb{E}[(X - \mathbb{E}[X])^2]$$

Also captures the error in the prediction using the mean. “How much do people’s heights deviate?”

Mean, Variance

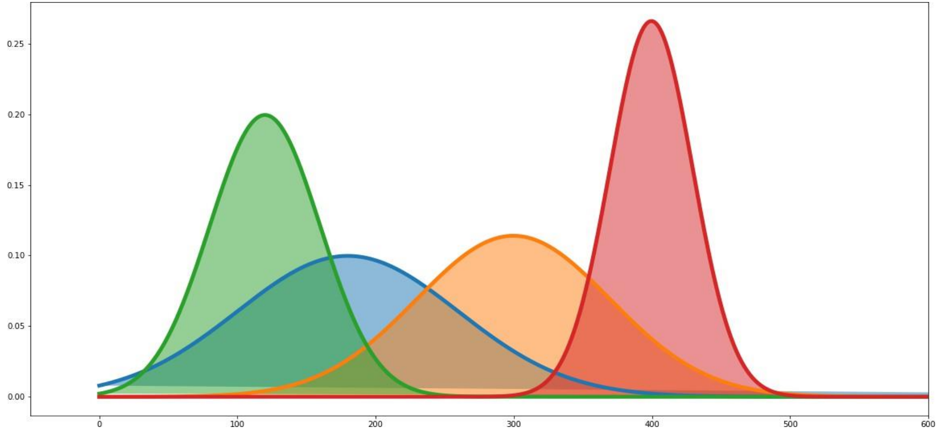
Mean: The mean is a prediction of the value of the random variable.

Answers the question "What do I expect the height of a random person to be?"

Variance: The variance captures the spread in your data.

$$\mathbb{E}[(X - \mathbb{E}[X])^2]$$

Also captures the error in the prediction using the mean. “How much do people’s heights deviate?”



Which distribution has a:

- *Large mean, small variance?*
- *Small mean, large variance?*

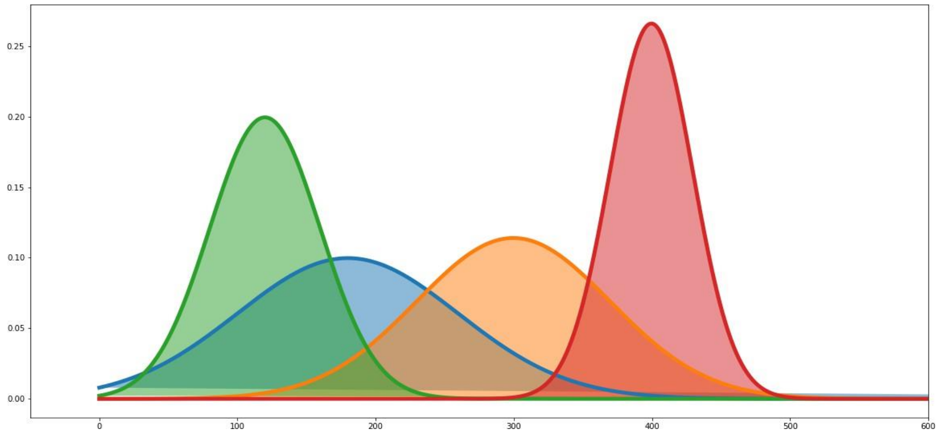
Mean, Variance

Mean: The mean is a prediction of the value of the random variable.

Answers the question "What do I expect the height of a random person to be?"

Variance: The variance captures the spread in your data. $\mathbb{E}[(X - \mathbb{E}[X])^2]$

Also captures the error in the prediction using the mean. “How much do people’s heights deviate?”



Which distribution has a:

- *Large mean, small variance?*

  Red

- *Small mean, large variance?*

  Blue

Acknowledgement

Most of the slides are based on the ML courses of:

- Natasha Jaques, University of Washington
- Andrew Ng, Stanford University

THANK YOU