

CSE 446: Machine Learning

Natasha Jaques

Course Website

<https://courses.cs.washington.edu/courses/cse446/26sp/>

Everything you need to know is there:

- Lecture slides
- Links to places to get help (Ed Discussion, staff email)
- Homeworks and due dates
- Past exams to study from
- Textbook and other references
- Sections, office hours
- Grading, FAQ, etc.

Course Staff - Instructor

Natasha Jaques

Assistant Professor in CSE

Staff Research Scientist at Google DeepMind



Course Staff - Teaching Assistants



Donovan Clay
(Head TA)



Alice Gao



Emmanuel Azuh
Mensah



Sankar Vaishnav
Harilal



Leo Maynard-
Zhang



Claudia
Gyonjyan



Min Jang



Hanwen Xu



Christina Hahn

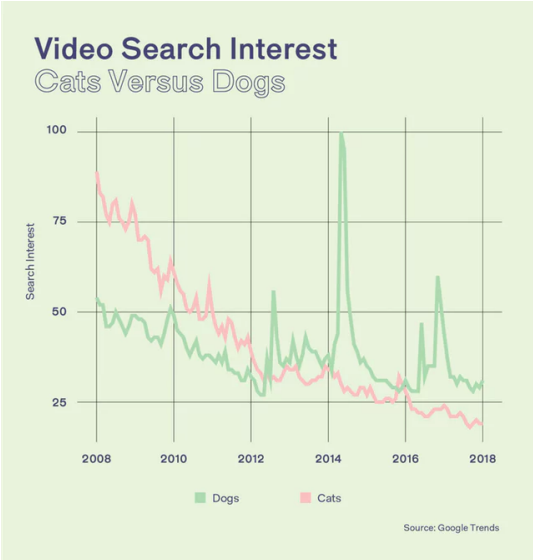
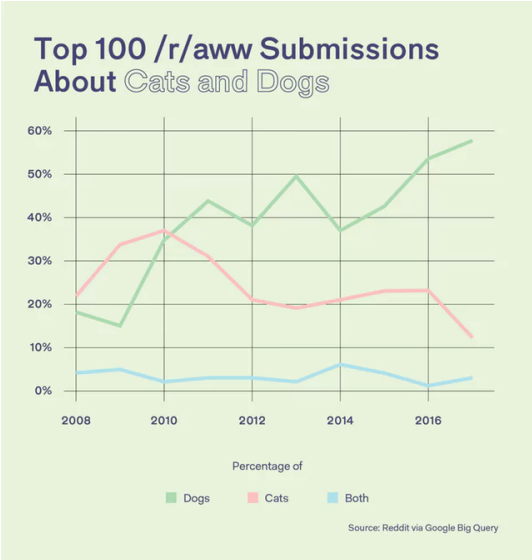
Traditional Algorithms

Social media mentions of Cats vs. Dogs

Reddit

Google

Twitter?



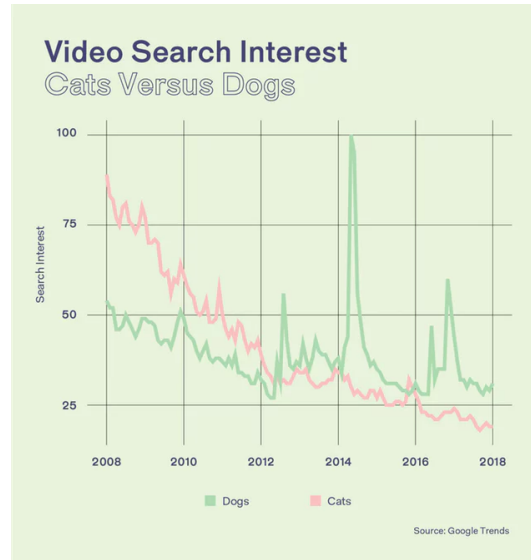
Traditional Algorithms

Social media mentions of Cats vs. Dogs

Reddit

Google

Twitter?



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

Traditional Algorithms

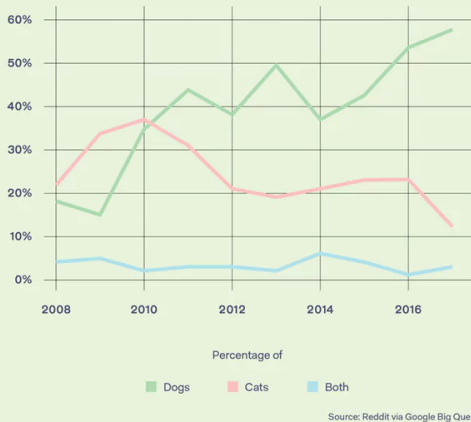
Social media mentions of Cats vs. Dogs

Reddit

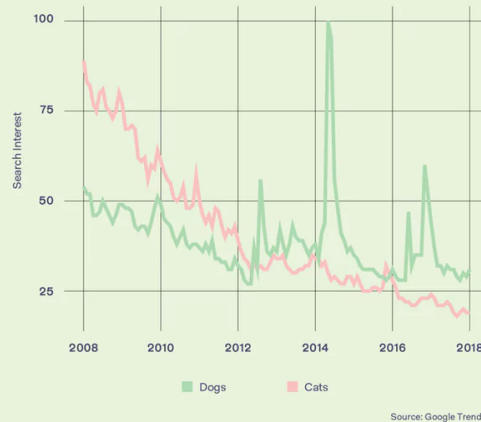
Google

Twitter? 

Top 100 /r/aww Submissions About Cats and Dogs



Video Search Interest Cats Versus Dogs

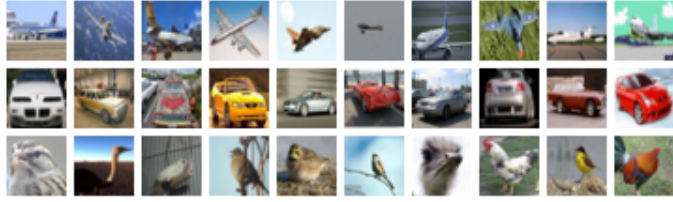


Write a program that sorts tweets into those containing "cat", "dog", or other

```
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
```

Machine Learning Algorithms

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



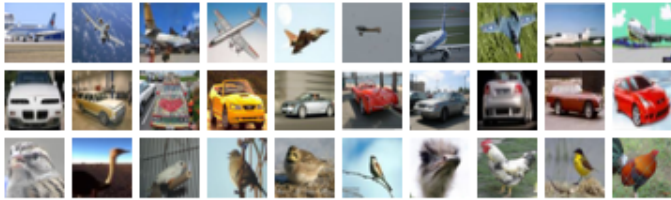
airplane

other

bird

Machine Learning Algorithms

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



airplane

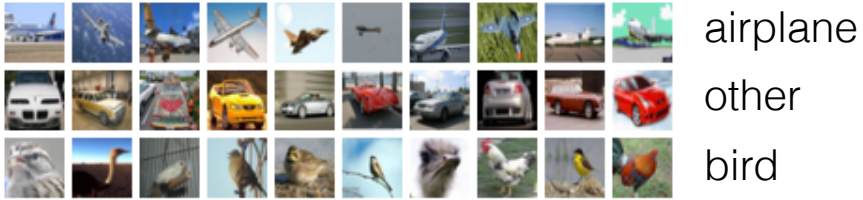
other

bird

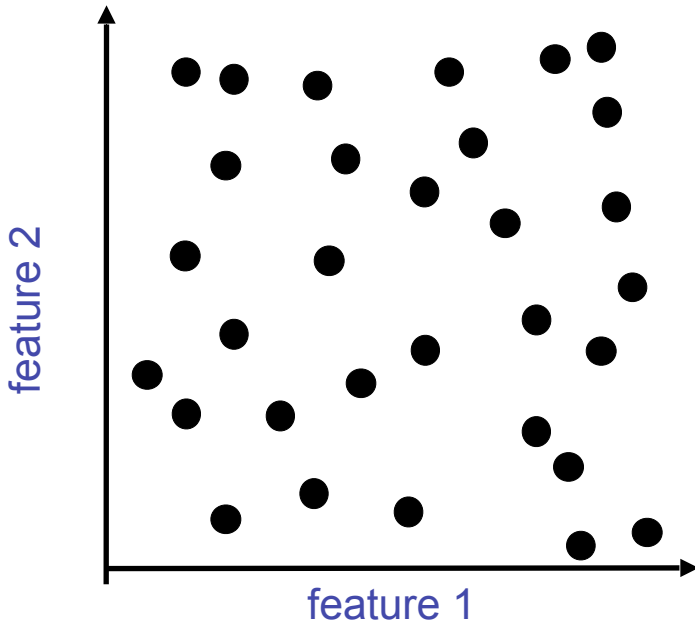
```
birds = []
planes = []
other = []
for image in images:
    if bird in image:
        birds.append(image)
    elif plane in image:
        planes.append(image)
    else:
        other.append(image)
return birds, planes, other
```

Machine Learning Algorithms

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

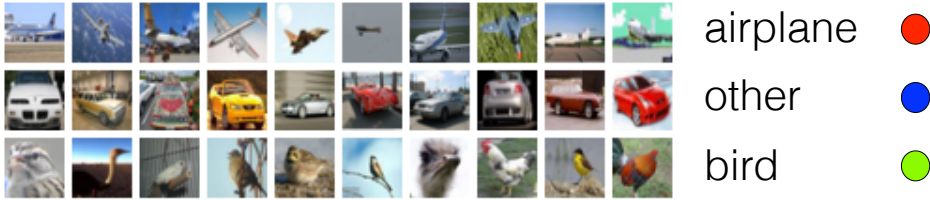


```
birds = []  
planes = []  
other = []  
for image in images:  
    if bird in image:  
        birds.append(image)  
    elif plane in image:  
        planes.append(image)  
    else:  
        other.append(image)  
return birds, planes, other
```

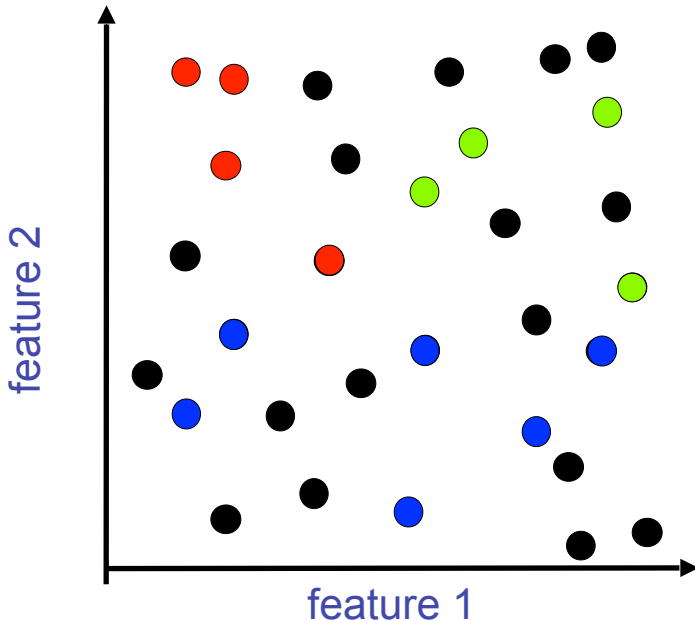


Machine Learning Algorithms

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

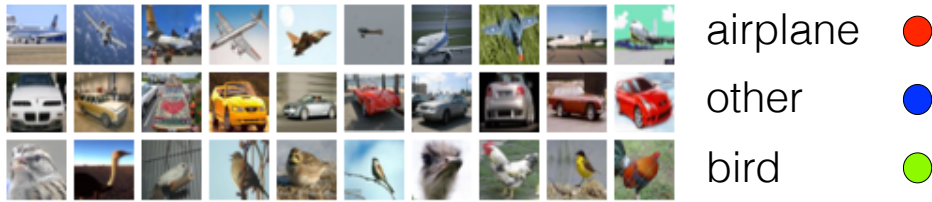


```
birds = []  
planes = []  
other = []  
for image in images:  
    if bird in image:  
        birds.append(image)  
    elif plane in image:  
        planes.append(image)  
    else:  
        other.append(image)  
return birds, planes, other
```

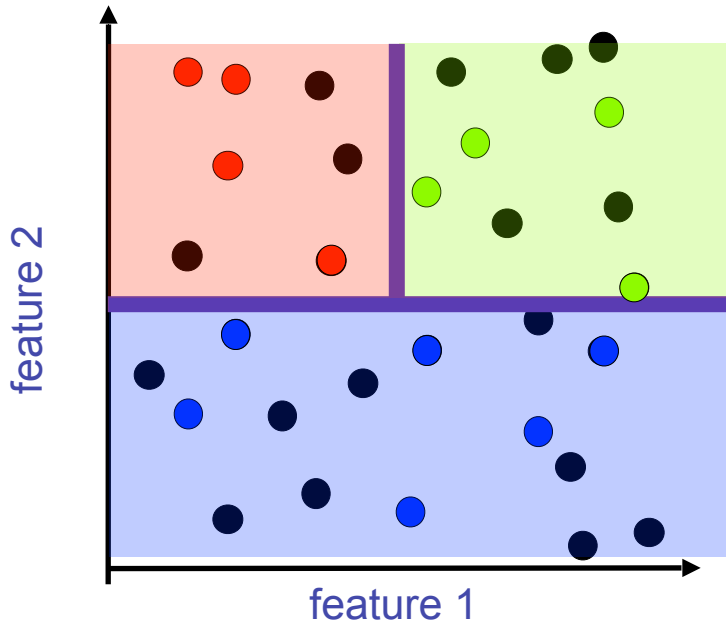


Machine Learning Algorithms

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

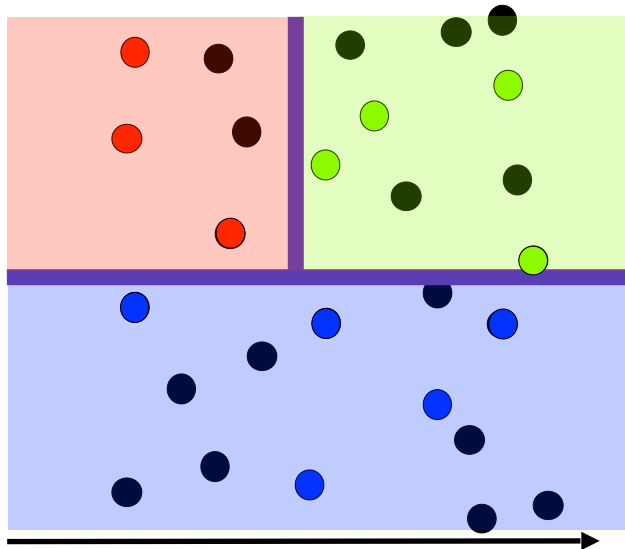
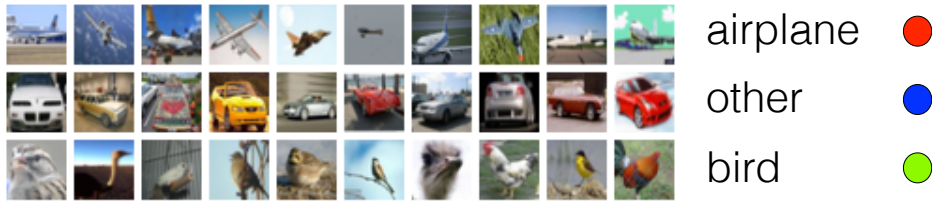


```
birds = []  
planes = []  
other = []  
for image in images:  
    if bird in image:  
        birds.append(image)  
    elif plane in image:  
        planes.append(image)  
    else:  
        other.append(image)  
return birds, planes, other
```



Machine Learning Algorithms

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



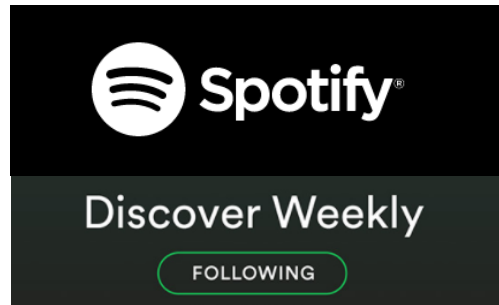
```
birds = []  
planes = []  
other = []  
for image in images:  
    if bird in image:  
        birds.append(image)  
    elif plane in image:  
        planes.append(image)  
    else: ...
```

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**

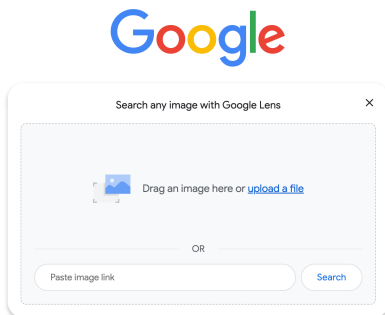
Machine Learning Ingredients

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

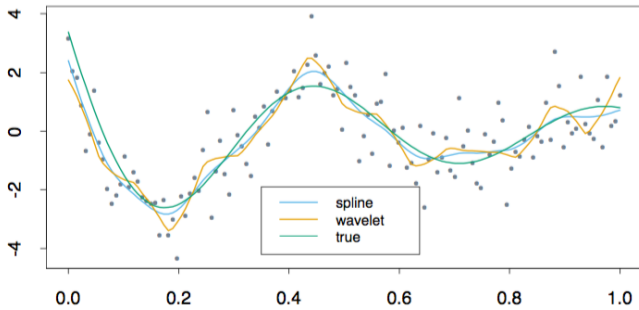


You may also like...

Even before LLMs like Claude, GPT, Gemini, you were already interacting with ML algorithms every day

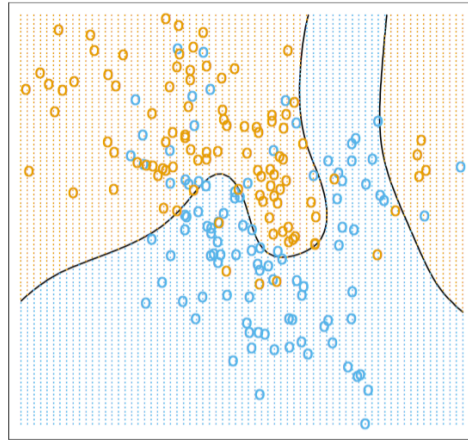


Flavors of ML



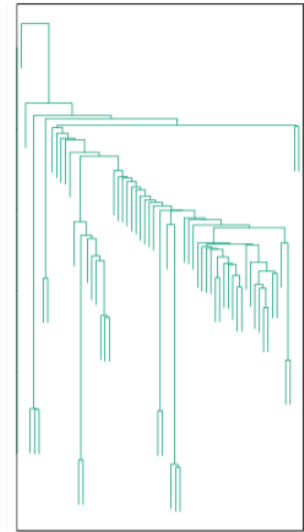
Regression

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?



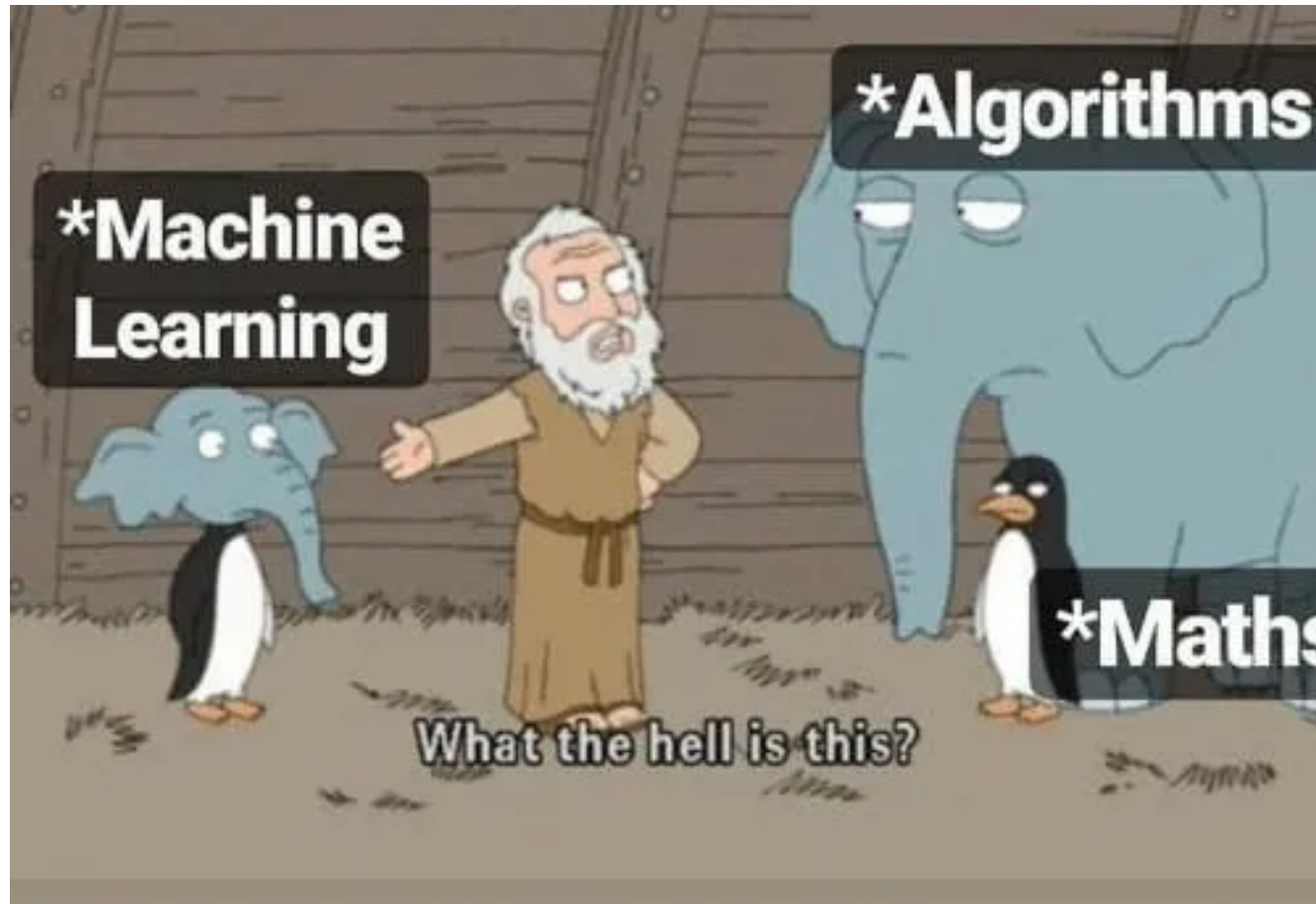
Unsupervised Learning

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)

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

What this class is not:

- **Survey course:** laundry list of algorithms, how to win Kaggle
- **An easy course:** familiarity with intro linear algebra and probability are assumed, homework will be time-consuming

Prerequisites

- Familiarity with:
 - Linear algebra
 - linear dependence, rank, linear equations
 - Multivariate calculus
 - Probability and statistics
 - Distributions, densities, marginalization, moments
 - Algorithms
 - Basic data structures, complexity
- Use HW0 as a refresher on the skills you will need
- **See assigned reading and website for additional review materials!**

Course Registration

- As of today, all enrollment restrictions have been dropped.
- If there are open spaces, any UW student may register for those spaces.
- All CSE course registration processes are managed centrally by CSE. **Do not email instructors, we cannot help.**

Resources:

- <https://www.cs.washington.edu/academics/ugrad/advising/>
- <https://www.cs.washington.edu/academics/ugrad/courses/petition>
- <https://www.cs.washington.edu/students/ugrad/non-major-registration>

Lectures

- Will be broadcast on Panopto/Canvas.
- Will be recorded and posted shortly after class.
 - Find Panopto recordings and videos in Canvas.
- **In-person attendance is strongly encouraged.**
 - Material can be tough, and I can teach better if I know where you get lost
 - **Please ask lots of questions!**
 - We can afford to slow down
- **For exams, in-person attendance is mandatory.**
- **New this quarter: due to new accessibility laws, I would like to get rid of hand-writing equations**

Grading

- **5 homeworks (40%)**
 - Each contain both theoretical questions and will have programming.
 - Collaboration okay. You must write, submit, and understand your answers and code (which we may run).
 - **READ COLLABORATION POLICY ON WEBSITE**
 - If you get someone (or Claude) to do your homework, did you actually learn it?
 - This stuff is foundational, if you want to work in this area need to actually understand
 - HW0 (4%), HW1 (9%), HW2 (9%), HW3 (9%), HW4 (9%)
- **Midterm (24%) and Final (33%)**
- **Section participation (3%)**

Homeworks

- **HW 0** is out (**Due next Wednesday 4/8 @ 11:59pm**)
 - Should be review (but being rusty is expected)
 - Work individually, treat as guide for what to brush up on
- **HW 1,2,3,4.** They are not easy or short. Start early.
- **Submit** to Gradescope.
- **Regrade requests** on Gradescope.
- **Late days:** 5 days total over the quarter; no more than 2 per assignment. If an assignment is submitted late and this exceeds your 5 late days, that assignment will receive 0 credit.
 - **Do not need to email us about using a late day**
- **Assignments due at 11:59pm**, submit early and often (do not email us at 12:05).

- 1. All code must be written in Python**
- 2. All written work must be typeset (e.g., LaTeX)**

See course website for tutorials and references.

Communication Channels

- **Announcements, questions about class, homework help** -> [EdStem](#)
 - “I think there is a typo in the homework?”
 - “What does this notation mean?”
 - “Is this an accurate description of how this works?”
- **Personal concerns** -> cse446-staff@cs.washington.edu
 - “Was in hospital...”, “Laptop was stolen...”
 - Do not email to ask if you can use one of your allowed late days
 - Do not email professor directly
- **TA office hours**
 - “How do I get started on problem 2?”
 - “Am I on the right track?”
 - We will not be able to give detailed code debugging help
- **Professor office hours**
 - Conceptual questions; i.e. “I didn’t get the connection you were making between nearest neighbors and kernel methods”
 - “I have this problem at work—can you point me in the right direction?”
- **Regrade requests**
 - Directly submit on Gradescope
- **Anonymous feedback** (<https://feedback.cs.washington.edu/>)
 - “Your real-world example X lacked nuance. I would like you to...”

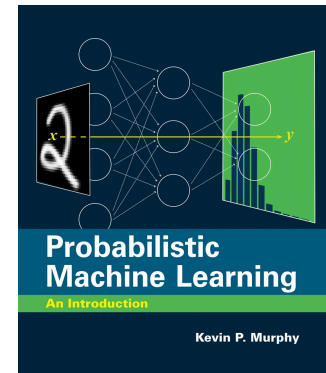
Textbooks

- Free PDF Textbook we will assign most reading from:

Probabilistic Machine Learning: An Introduction

Kevin Murphy

[PDF](#), also in print



- So many more resources on the website!

The textbook for the course is:

- [Murphy] [Probabilistic Machine Learning: An Introduction](#), Kevin Murphy, 2022 (Note, this is not the 2012 edition used in past years). Follow the link to obtain a free, PDF pre-print as well as options to purchase a hard copy.

For a gentler introduction to machine learning the following text is available for free online:

- [CIML] [A Course in Machine Learning](#) by Hal Daume III.

The following three texts are also excellent general machine learning texts and their PDFs are available for free online.

- [B] [Pattern Recognition and Machine Learning](#), Christopher Bishop.
- [HTF] [The Elements of Statistical Learning: Data Mining, Inference, and Prediction](#), Trevor Hastie, Robert Tibshirani, Jerome Friedman.
- [EH] [Computer Age Statistical Inference: Algorithms, Evidence and Data Science](#), Bradley Efron, Trevor Hastie.
- [ZLLS] [Dive into Deep Learning](#), Aston Zhang, Zach Lipton, Mu Li, Alex Smola.

You may also find these reference materials useful throughout the quarter.

- **Machine Learning (and related topics)**

- [Understanding Machine Learning: From Theory to Algorithms](#), Shai Shalev-Shwartz, Shai Ben-David. An introduction to theoretical machine learning.
- [Fairness and Machine Learning](#), Solon Barocas, Moritz Hardt, Arvind Narayanan
- [Foundations of Data Science](#), by Avrim Blum, John Hopcroft and Ravi Kannan. This freely available pdf has nice chapters on machine learning (chapter 5), clustering (chapter 7) and SVD (chapter 3).
- [Crib sheet of math for ML](#) by Iain Murray

- **Linear Algebra and Matrix Analysis**

- [These wonderful videos](#) by 3blue1brown provide a gentle and highly intuitive overview of linear algebra. (The same person created most of the videos on multivariable calculus on Khan Academy -- also excellent).
- [Linear Algebra Review and Reference](#) by Zico Kolter and Chuong Do (free). Light refresher for linear algebra and matrix calculus if you're a bit rusty.
- [Linear Algebra](#), David Cherney, Tom Denton, Rohit Thomas and Andrew Waldron (free). Introductory linear algebra text.
- [Matrix Analysis](#), Horn and Johnson. A great reference from elementary to advanced material.
- [The Matrix Cookbook](#), Pederson and Pederson. Contains many useful matrix manipulation identities, including derivatives

- **Probability and Statistics**

- [20su offering of CSE312 materials and video lectures](#) by Alex Tsun. Slides, full notes, and short video lecture snippets on basic probability and statistics. This course was recently redesigned in part to better prepare students for CSE 446, making it an excellent resource for review.
- [Probability Review](#) by Arian Maleki and Tom Do. (From Andrew Ng's machine learning class.)
- Section notes from Anna Karlin's 18au offering of 312: [Counting, Combinatorics + intro probability, Conditional probability, Random variables & linearity of expectation, Variance and discrete r.v.s, Conditional expectation, Joint distributions, Continuous random variables, CLT, tail bounds and MLE.](#)
- [All of Statistics](#), Larry Wasserman. Chapters 1-5 are a great probability refresher and the book is a good reference for statistics.
- [A First Course in Probability](#), Sheldon Ross. Elementary concepts (previous editions are a couple bucks on Amazon)

- **Optimization**

- [Numerical Optimization](#), Nocedal, Wright. Practical algorithms and advice for general optimization problems.
- [Convex Optimization: Algorithms and Complexity](#), Sébastien Bubeck. Elegant proofs for the most popular optimization procedures used in machine learning.

- **Python**

- www.learnpython.org "Whether you are an experienced programmer or not, this website is intended for everyone who wishes to learn the Python programming language."
- [NumPy for Matlab users](#)

- **Latex**

- [Learn Latex in 30 minutes](#)
- [Overleaf](#). An online Latex editor.
- [Standalone Latex editor](#) on your local machine
- [Latex Math symbols](#)
- [Detexify](#) LaTeX handwritten symbol recognition

Enjoy!

- 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

Probability review



Definitions

- **Random Variable:** A variable that takes on different values determined randomly.

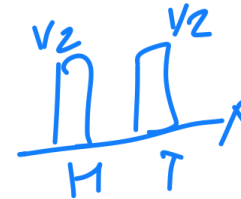
- Example: The height of a person from the US $P(\)$

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



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

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



- We talk about **sampling** from a distribution:

- “Consider a sample of 100 different heights of people from the US 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)

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

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

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

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

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

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

A, C ? 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 $A = \{X \in \{3,4\}\}$,

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

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

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

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

What is $P(X \leq 4 | X \geq 3)$?

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

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

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

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

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

What is $P(X \leq 4 | X \geq 3)$

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

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

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

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

$$P(U | V) = \frac{P(U \cap V)}{P(V)} \quad \text{If independent then: } = \frac{P(U)P(V)}{P(V)} = P(U)$$

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

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

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.

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

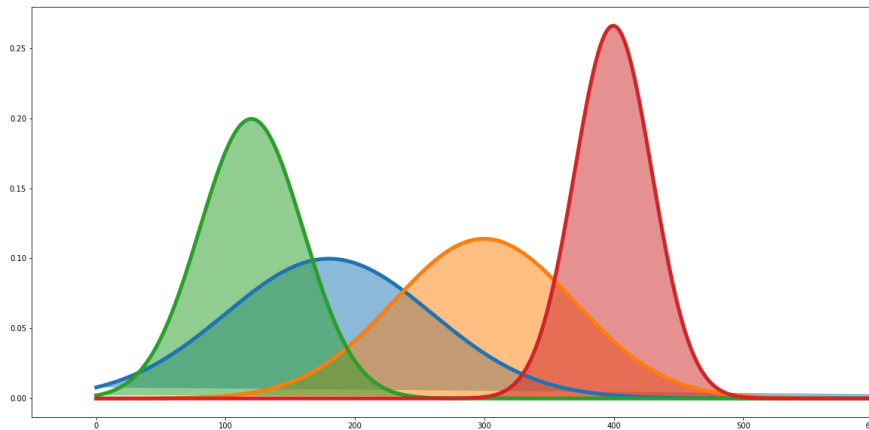
Median M

The value of X that is separating the higher half of its range from the lower half.

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

Mean, variance

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?"



Which distribution has a:

- *Large mean, small variance?*

Red

- *Small mean, large variance?*

Blue

The variance captures the spread in your data. Also captures the error in the prediction using the mean. "How much do people's heights deviate?"

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