Introduction to Machine Learning-CS6140

Dr. Shanu Sushmita

Slides adopted from Jarrar,

Moses Charikar and Chris Ré

Who am I

- Name: Dr. Shanu Sushmita
- Position:
  - Assistant Teaching Professor at Khoury college at Northeastern University
  - Affiliate Assistant Professor at School of Engineering and Technology at University of Washington, Tacoma.
  - Ambassador, Women in AI for Washinghton.
- Education : PhD in Computer Science from University of Glasgow, UK.
- Areas of interests (but not limited to): Information Retrieval, and Machine Learning.
- Mother to a 9 year old (Siya) and 5 years old (Savi) daughters

Research Journey

2004-2007

Artificial Intelligence
- Fuzzy Case Based Reasoning System
- 1 publication

2007-2012

Information Retrieval
- Understanding User Behavior in Aggregated Search (PhD)
- 7 publications

2012-2012

Digital Libraries:
- Author disambiguation for digital libraries

2013 -2014

Social Media
- Personality Prediction
- Named Entity Recognition
- 2 publications

2014-2017

Healthcare:
- Population cost
- Risk of Readmission
- Web search and Health topics.
- Frequent Flyers
- 7 Publications

2018-2022

Music
- User Engagement
- Song Genre Prediction
- Ranking song lyrics
- Content Analysis
- 2 publications

2024

Generative AI Research Group

Information Retrieval (IR); Machine Learning (ML)

Industry Experience

Providence spinout Tegria acquires Seattle healthcare AI startup KenSci

Today in Class

- **Course Logistics**
- **Expectations**
- **Introduction to ML**
- **Types of Learning**
  - Supervised Machine Learning
  - Unsupervised Machine Learning
  - Semi-Supervised Learning
  - Reinforcement Learning
- Evaluation

Course Logistics

Canvas: Go to place for “all” information related to your course

Office Hours: Thursday 9:30 am -10:30 am (Virtual)

Email: s.sushmita@northeastern.edu

TA: No TA

Email Policy: Please give 2-3 working days to hear back, emails sent on Friday will be responded on and after the following Monday. No Teams messaging, please.

No teams chat please

Course Overview

Week	Topic
Week 1	Intro to Machine Learning
Week 2	Optimization in Machine Learning + Regression
Week 3	Classification + Text Classification + Project Proposal (all teams to be present in person)
Week 4	Classification – Decision Trees
Week 5	Instance-Based Learning
Week 6	Bayesian Learning + Project 1 Update Presentation (all teams to be present in person)
Week 7	Support Vector Machines
Week 8	Logistic Regression
Week 9	Neural Network
Week 10	Deep Learning: RNNs
Week 11	Project 2 Update Presentation (all teams to be present in person)
Week 12	Transformers
Week 13	Clustering
Week 14	Final Project (all teams to be present in person)

Course Tips

Work hard

Be on time to class

Do the assigned readings

Do other readings

Be patient and have reasonable expectations

you're not supposed to understand everything we cover in class during class

Seek help sooner rather than later

office hours

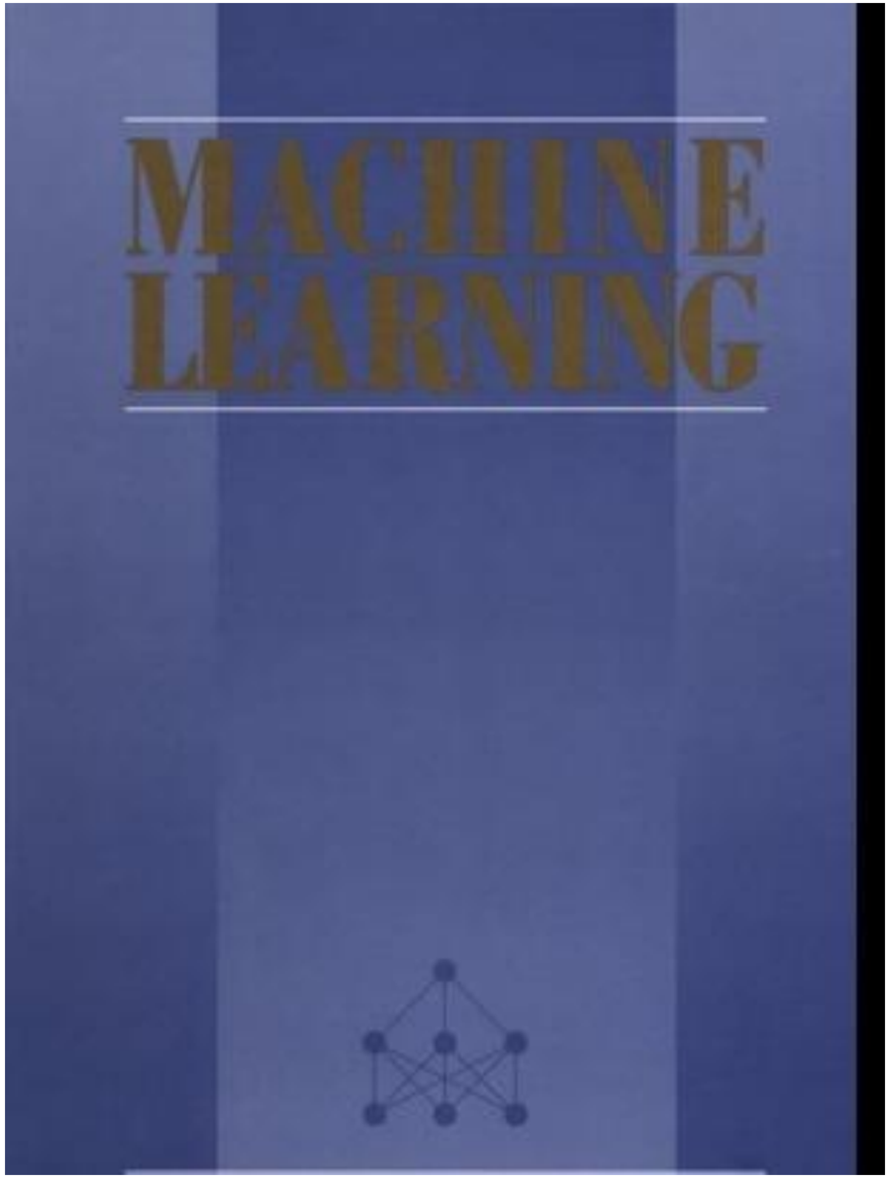
Questions via email

Remember the golden rule:

no pain, no gain

Textbook

- Tom Mitchell, Machine Learning (McGraw-Hill)
- before the lecture:
  skim through the material
- after the lecture:
  read the text carefully
- make sure that you understand all the assigned reading material



Pre-requisite

Probability

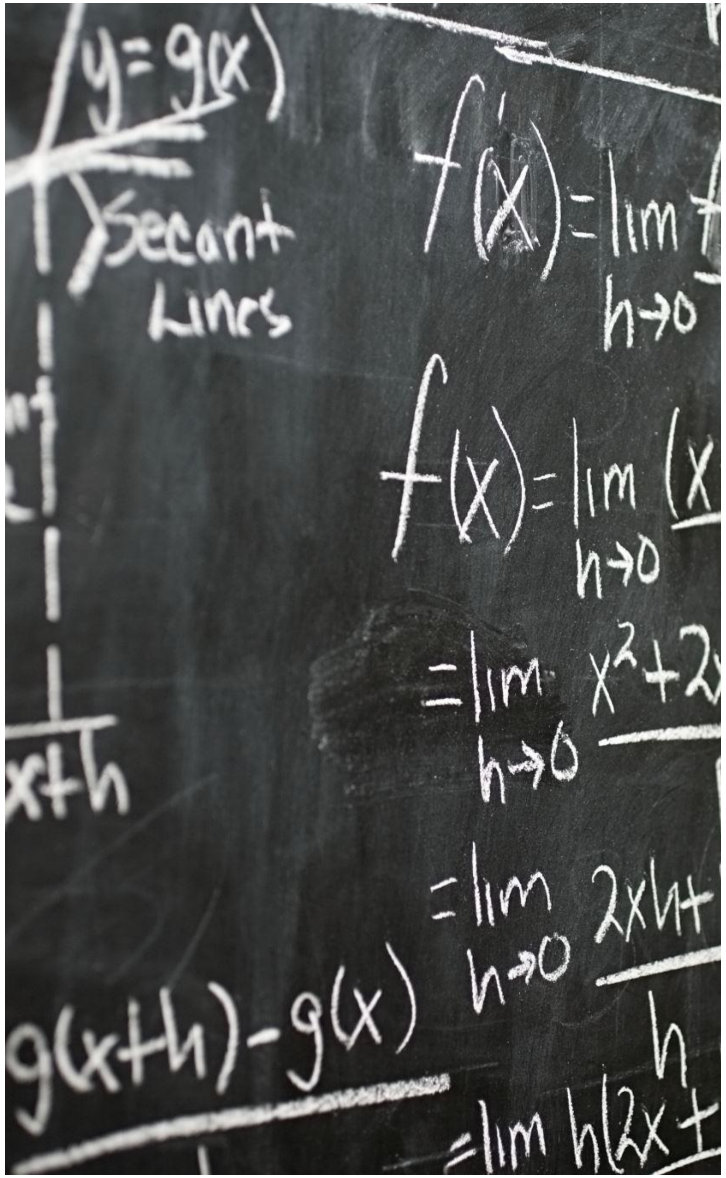
- distribution, random variable, expectation, conditional probability, variance, density

Linear algebra

- matrix multiplication
- eigenvector

Python:

- Basic programming
- You will learn a LOT during this class.



Coursework and Grading

Homework: Individual (40%)

- Four Homework assignments (10 points each)
- Graded:
  - Code Functionality (5 points)
  - Quality of Comments (2 points)
  - Code Organization and Structure (3 points)

Project: Group (60%)

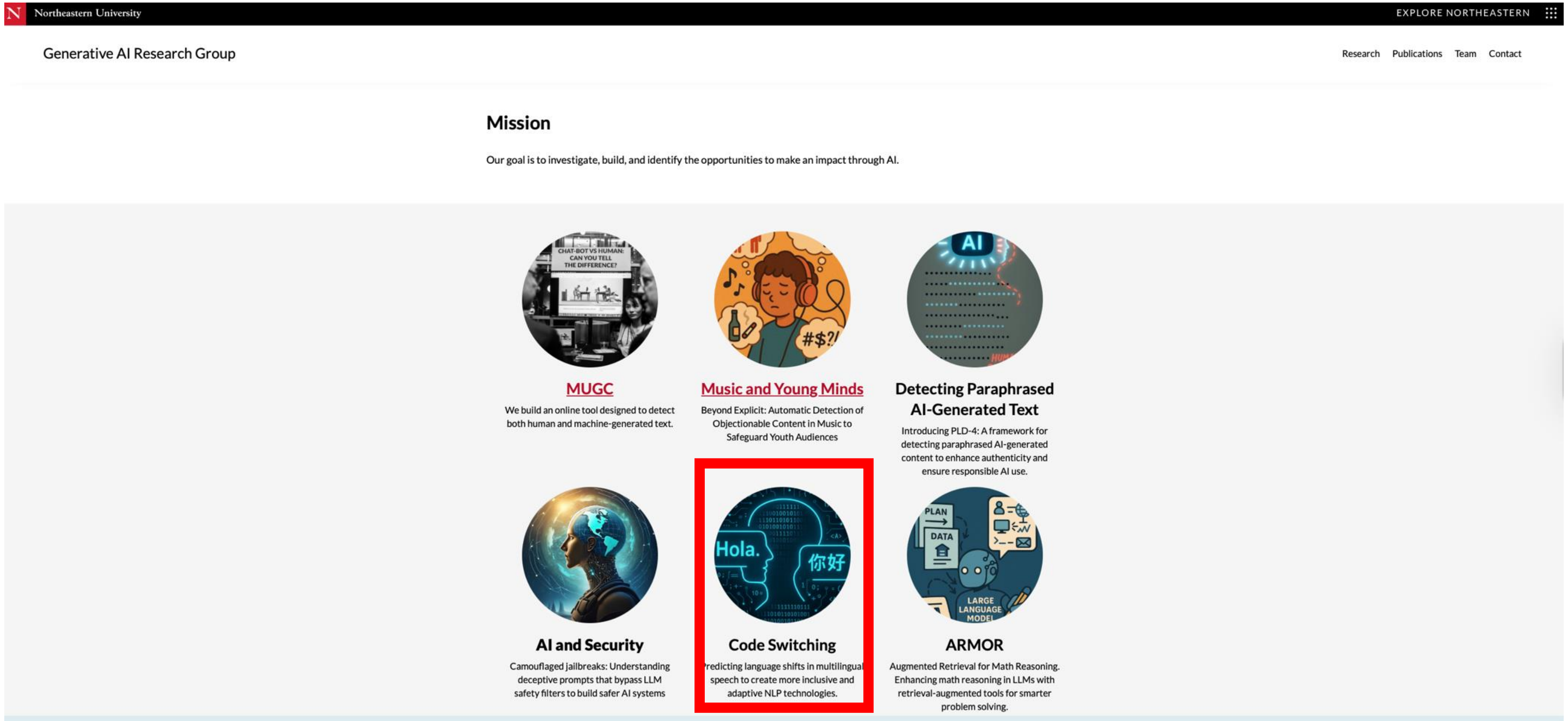
- Project Proposal (week 3)
- Project Update 1 (Week 6)
- Project Update 2 (week 10)
- Final Project (Week 13 &14)
- Peer Review

Canvas Walkthrough

Group Projects

Find your team member today…..

GenAI Research Lab (Seattle)



https://genairesearch.sites.northeastern.edu

Predictive Multitask Code-Switching

- We are building a model that anticipates the future of bilingual conversation.
- Students should choose their teams.
- This project constitutes 60% of your course grade.

What is Code-Switching?

Definition:

The practice of alternating between two or more languages or language varieties in a single conversation.

Not Random:

It's not a sign of language deficiency. It's a complex, rule-governed linguistic phenomenon.

Example (English-Spanish):

"I'm going to the supermercado to get some groceries."

"Can you send me the link, por favor?"

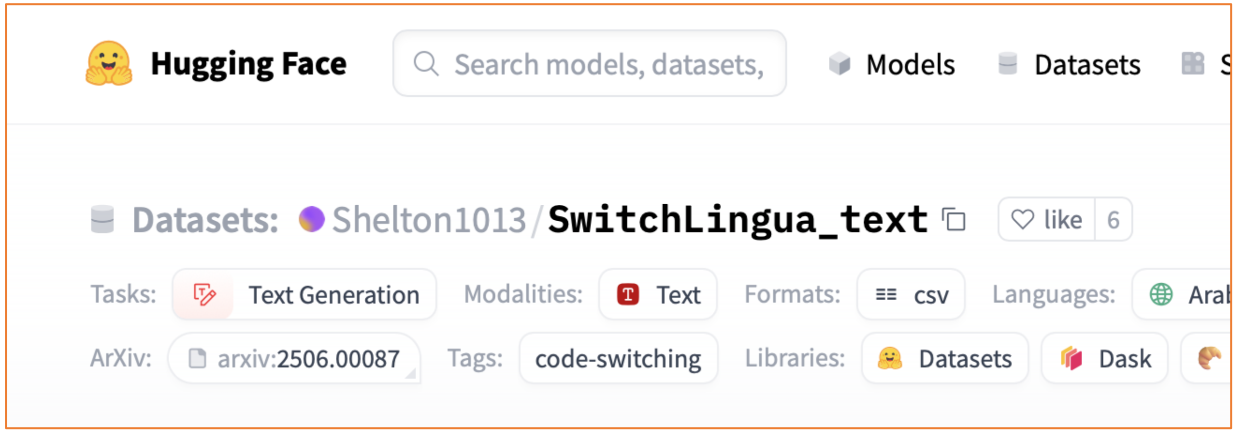
Computational Challenge:

How can a machine understand the contextual cues and predict where these switches might occur? This requires understanding context, syntax, and semantics across two languages.

The Problem – Detection vs. Prediction

- Real-world applications (chatbots, keyboards) need to predict *before* the switch happens, not after.
- **Imitation of Detection:** Traditional models identify switches in full, static sentences (post-hoc).
- **The Goal:** Predict at token t if token t+1 will be a switch.
- **The Predictive Challenge:** In a live stream, we only have the "prefix" (context so far).

The Dataset – SwitchLingua

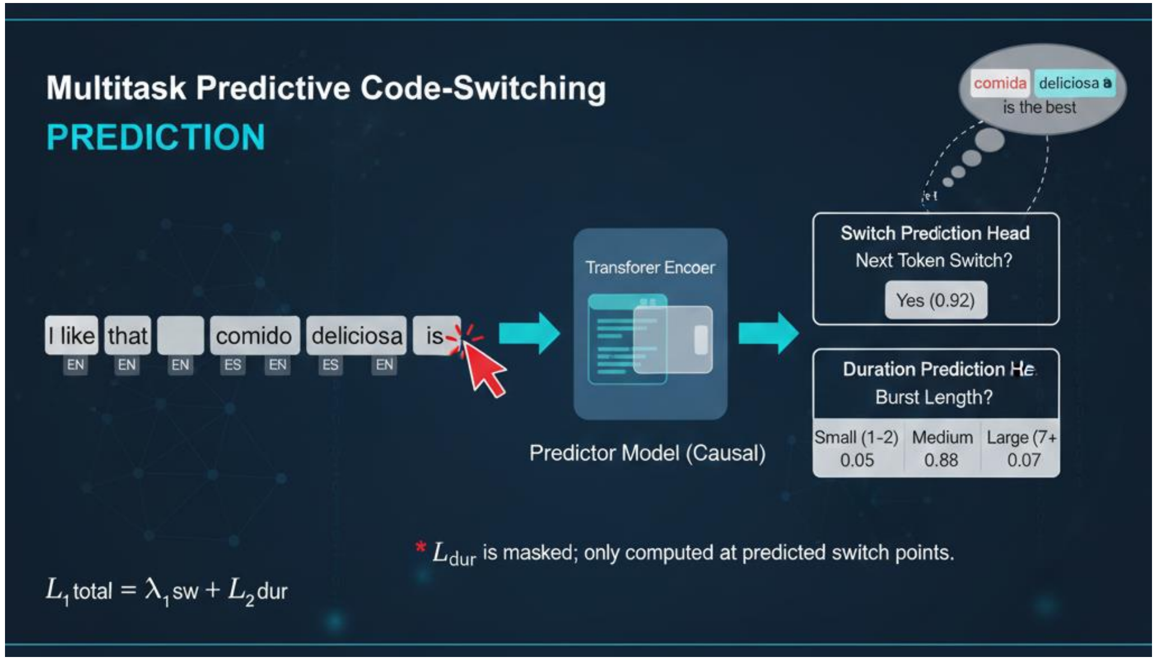


- We are testing on a diverse, high-fidelity corpus of 420K samples.
- **Corpus:** *SwitchLingua_text* (Shelton et al.).
- **Diversity:** 12+ language pairs (e.g., Hindi-English, Spanish-English, Cantonese-English).
- **Naturalness:** Grounded in linguistic constraints (Equivalence & Functional Head constraints).
- Apply today for data access, its free!!

https://huggingface.co/datasets/Shelton1013/SwitchLingua_text

Multitask Task Definition

- We aren't just predicting *if* a switch happens, but *how long* it will last.
- **Task 1:**
  - **Switch Prediction (y_sw):** Binary classification (Next token = Switch?).
- **Task 2:**
  - **Duration Prediction (y_dur):** 3-class classification (Small: 1-2, Medium: 3-6, Large: 7+ tokens).
- **Rationale:** Multi-tasking helps the model learn the "rhythm" of different language pairs.



Methodology – Causal Engineering

We enforce a "Causal Mask" to ensure no future leakage.

Model Backbone: Comparative study of XLM-RoBERTa vs. mBERT/Unicoder.

Architecture: Shared Encoder → Dual Classification Heads.

The Shifted Label: Training the model at index t using the label of t+1.

The "Universality" Metric

- A truly "Universal" model isn't just accurate; it's **consistent**.
  - **Performance:** Measured via F1-score.
  - **The Variance Audit (σ):** We calculate the standard deviation of performance across all 12 language pairs.
  - **Hypothesis:** The "Superior" model is the one with the **lowest** variance (lowest σ).

Expected Results & Analysis

1

We expect different architectures to "specialize" in different linguistic "burst" sizes.

2

Hypothesis: XLM-R will dominate in switch detection due to better cross-lingual embeddings.

3

Duration Analysis: Will the model be better at predicting "Small" lexical switches or "Large" clausal switches?

Project Timeline (14 Weeks)

01

Weeks 1-4: Data & Label Engineering (The "Shifted" Pipeline).

02

Weeks 5-7: Architectural Implementation (Custom Heads & Causal Mask).

03

Weeks 8-10: Training & Hyperparameter Tuning.

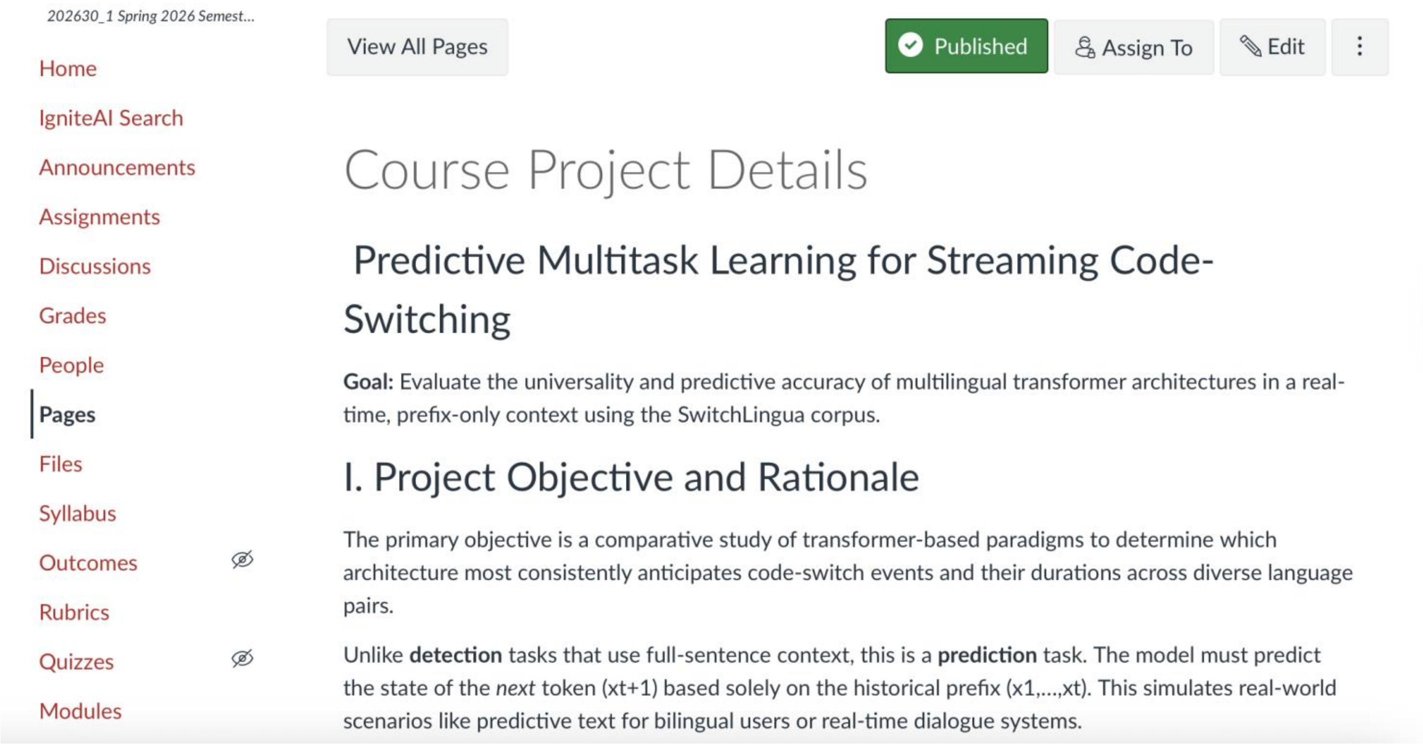
04

Weeks 11-14: Evaluation, Universality Audit, and Final Paper.

Grading

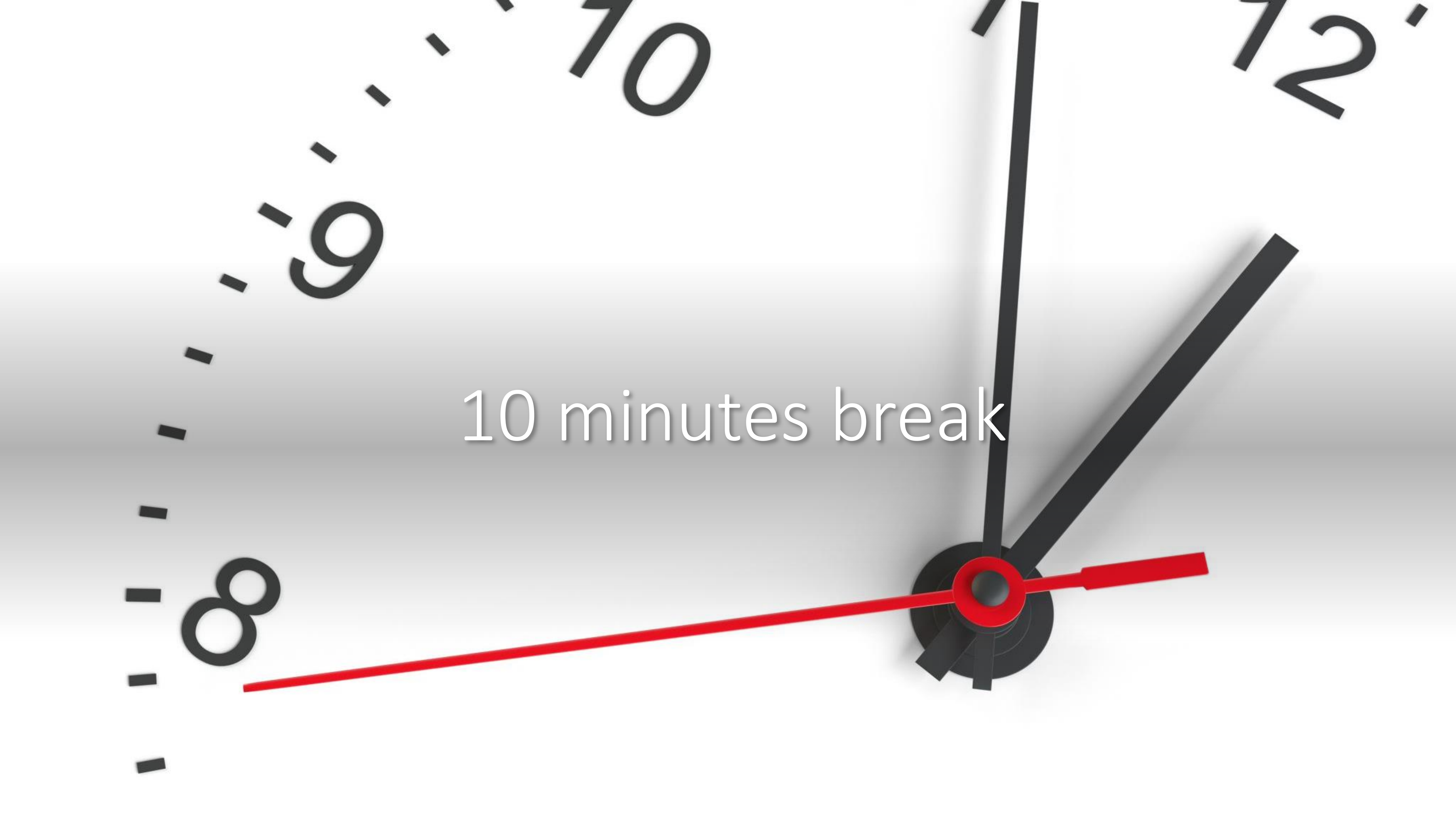
- By default, every member of the team will get same score.
  - *If you feel that one of your team member did not contribute enough, team meeting with the instructor will be required to mitigate the score.*

Complete Project detail on Canvas



Questions?

Before we start the lecture.


10 minutes break

Why learn ML?

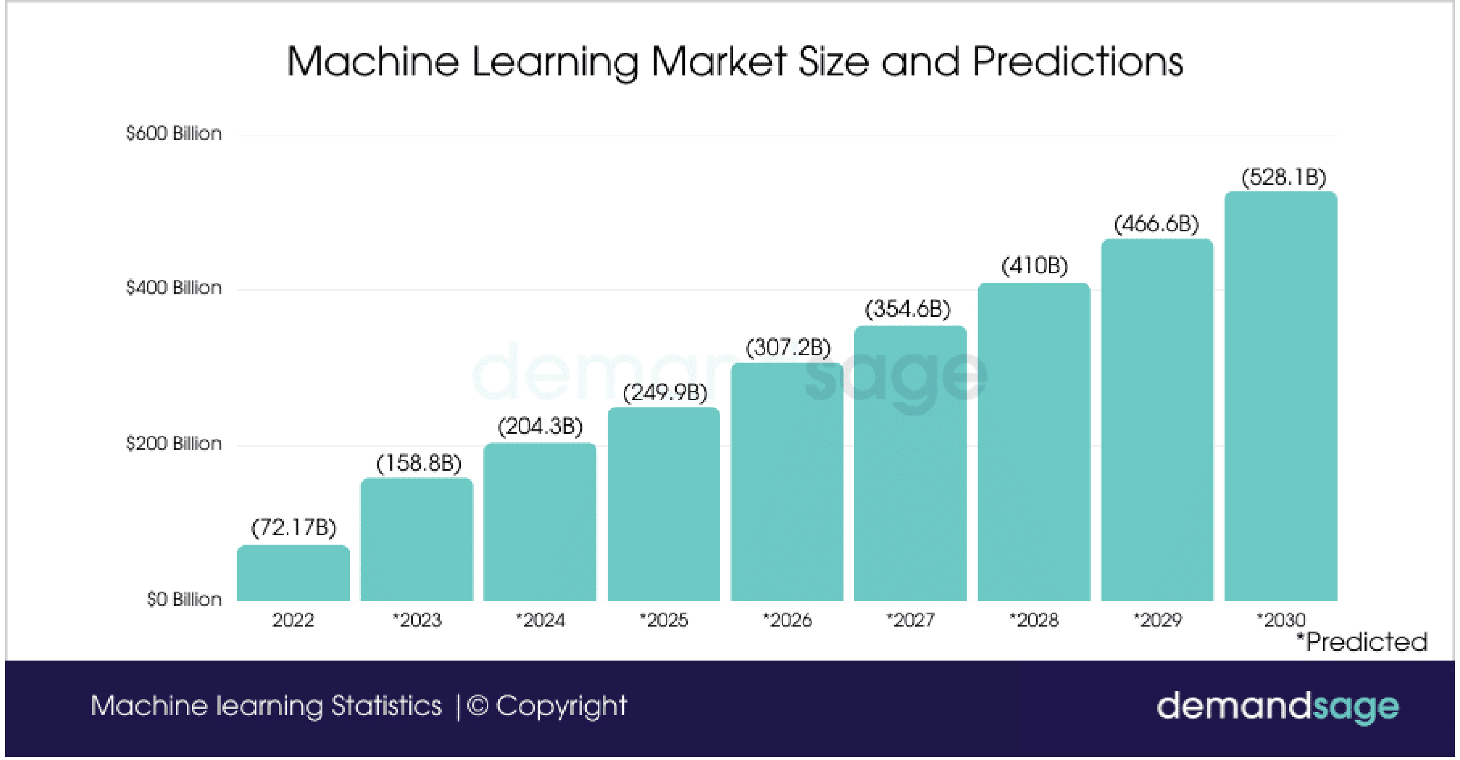
Many application areas

- Web search
- Computational biology
- Finance
- E-commerce
- Space exploration
- Robotics
- Information extraction
- Social networks
- Debugging
- [Your favorite area]

slide from P. Domingos, UW Seattle

ML Job Market

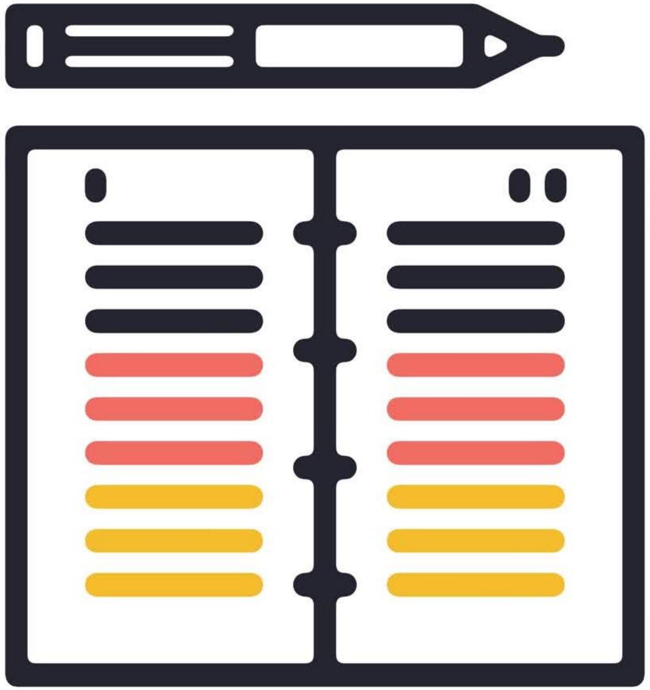
- The demand for AI and ML specialists will grow by 40% from 2023 to 2027.
- The average ML engineer's salary is $133,336/year.
- The most sought-after degree for ML engineer positions is computer science.
- 8% of ML engineer job offers require Python.



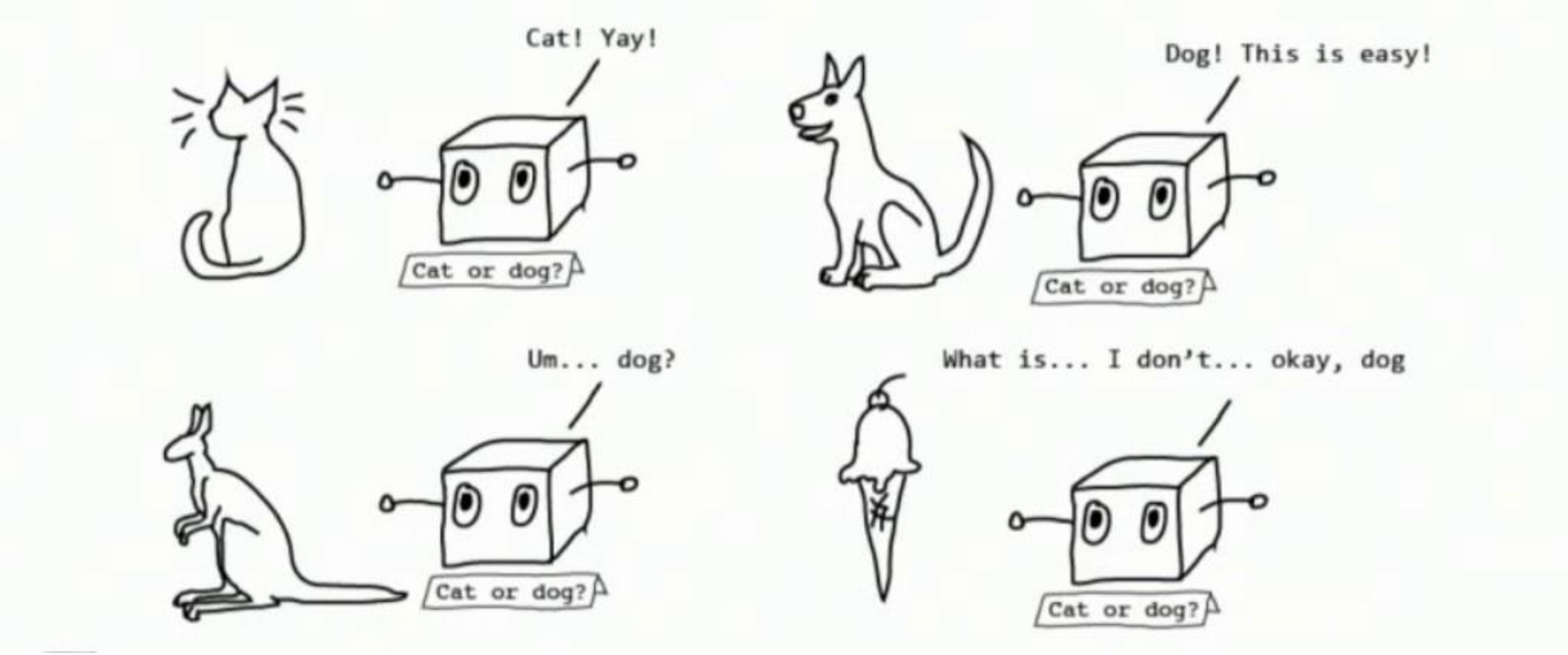
https://www.demandsage.com/machine-learning-statistics/

Today: Remaining Class

- Introduction to Machine Learning
- Types of Learning
  - Supervised,
  - Unsupervised,
  - Semi-Supervised
  - Reinforcement Learning
- Evaluation



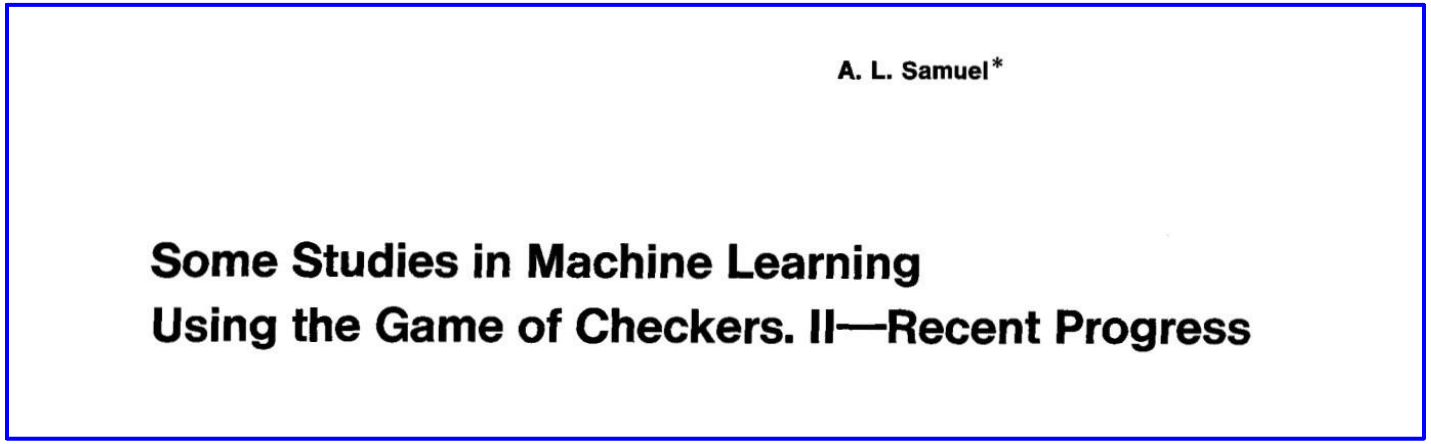
Introduction to ML



Definition of Machine Learning

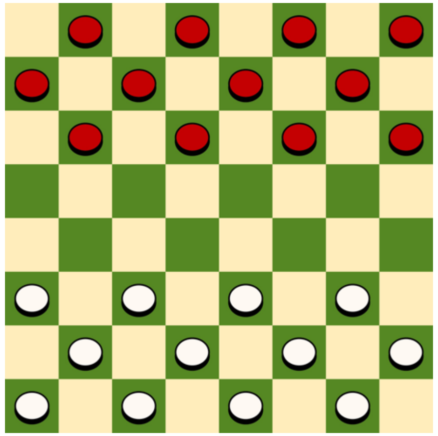
Arthur Samuel (1959):

Machine Learning is the field of study that gives the computer the ability to learn without being explicitly programmed.



A. L. Samuel*

Some Studies in Machine Learning Using the Game of Checkers. II—Recent Progress

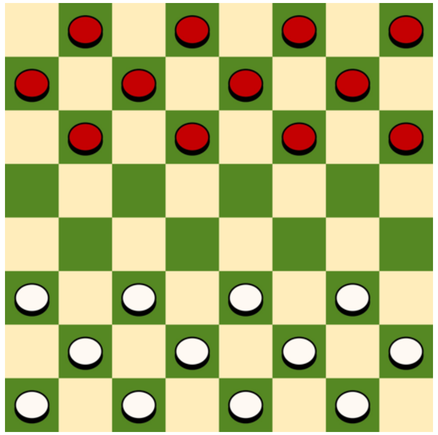


Photos from Wikipedia

Definition of Machine Learning

Tom Mitchell (1998):

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.



Experience (data): games played by the program (with itself)

Performance measure: winning rate

Image from Tom Mitchell's homepage

Building Blocks

Machine Learning always starts from a

“Problem”

- That we want to solve
- Automatically
- Also known as “ML Task”

Data

- That captures information about the problem.
- Often in form of features.

Evaluation

- Some metric
- Mathematical score
- A number

Data: representation: features

example: recognizing triangles

color	size	# slides	equal sides	...	label
red	big	3	no	...	yes
green	big	3	yes	...	yes
blue	small	inf	yes	...	no
blue	small	4	yes	...	no
.	.	.	.	.	.
red	big	3	yes	...	yes

Machine Learning Pipeline

color	size	sides	equal sides	...	label
red	big	3	no	...	yes
green	big	3	yes	...	yes
blue	small	inf	yes	...	no
blue	small	4	yes	...	no
.	.	.	.	.	.
red	big	3	yes	...	yes

labeled examples

training

machine learning algorithm

model

testing

color	size	sides	equal sides	...	label
red	big	3	no	...	???
green	big	3	yes	...	???
blue	small	inf	yes	...	???
blue	small	4	yes	...	???
.	.	.	.	.	???
red	big	3	yes	...	???

new, unlabeled examples

model

color	size	sides	equal sides	...	label
red	big	3	no	...	yes
green	big	3	yes	...	yes
blue	small	inf	yes	...	no
blue	small	4	yes	...	no
.	.	.	.	.	.
red	big	3	yes	...	yes

predictions

Machine Learning

basic ingredients

1. Training data: a set of examples of the concept we want to automatically recognize

2. Representation: a set of features that we believe are useful in recognizing the desired concept

3. Learning algorithm: a computer program that uses the training data to learn a predictive model of the concept

Machine Learning

basic ingredients

4. Model: a (mathematical) function that describes a predictive relationship between the feature values and the presence/absence of the concept

5. Test data: a set of previously unseen examples used to estimate the model's effectiveness

6. Performance metrics: a set of statistics used measure the predictive effectiveness of the model

Evaluating a Model

1. Make predictions on data that is “unlabeled” .
   - we know what the true labels are, but the model doesn’t “see” or use these labels
2. Evaluate based on some metric that compares the model’s predicted labels with the known true labels .

color	size	sides	equal sides	...	label
red	big	3	no	...	???
green	big	3	yes	...	???
blue	small	inf	yes	...	???
blue	small	4	yes	...	???
.	.	.	.	.	???
red	big	3	yes	...	???

new, unlabeled examples

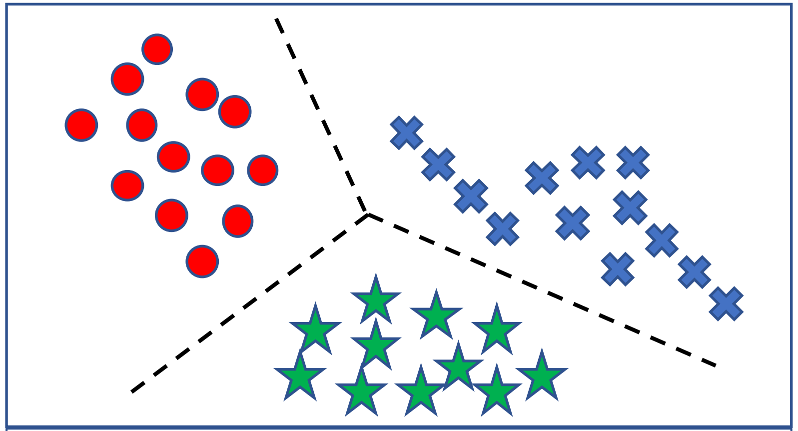
testing

model

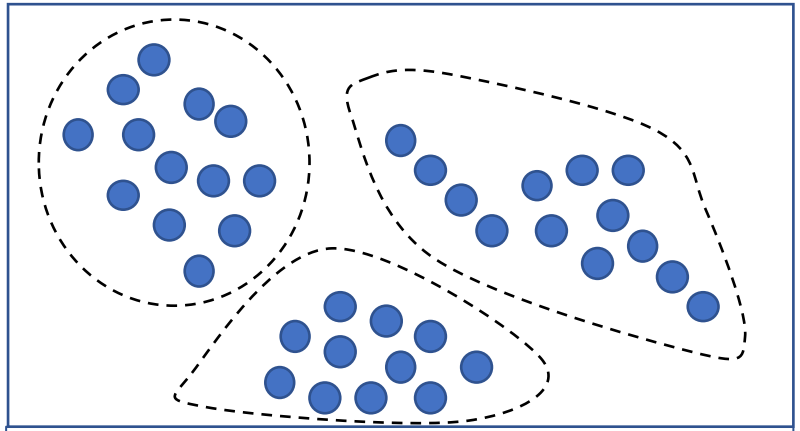
color	size	sides	equal sides	...	label
red	big	3	no	...	yes
green	big	3	yes	...	yes
blue	small	inf	yes	...	no
blue	small	4	yes	...	no
.	.	.	.	.	.
red	big	3	yes	...	yes

predictions

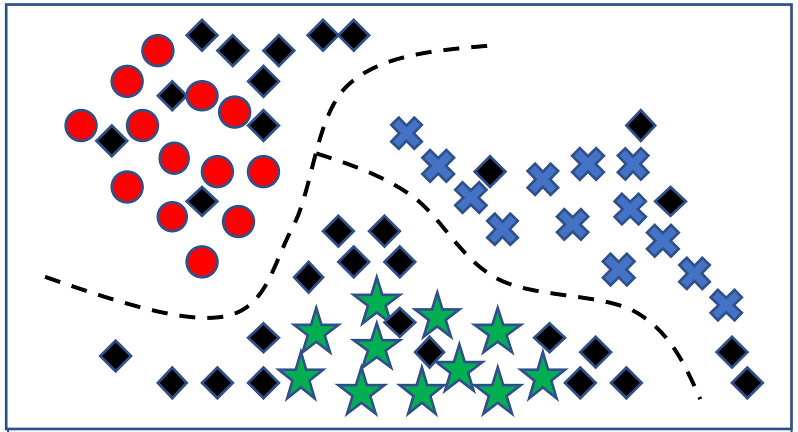
Types of Learning



Supervised learning



Unsupervised learning



Semi-supervised learning

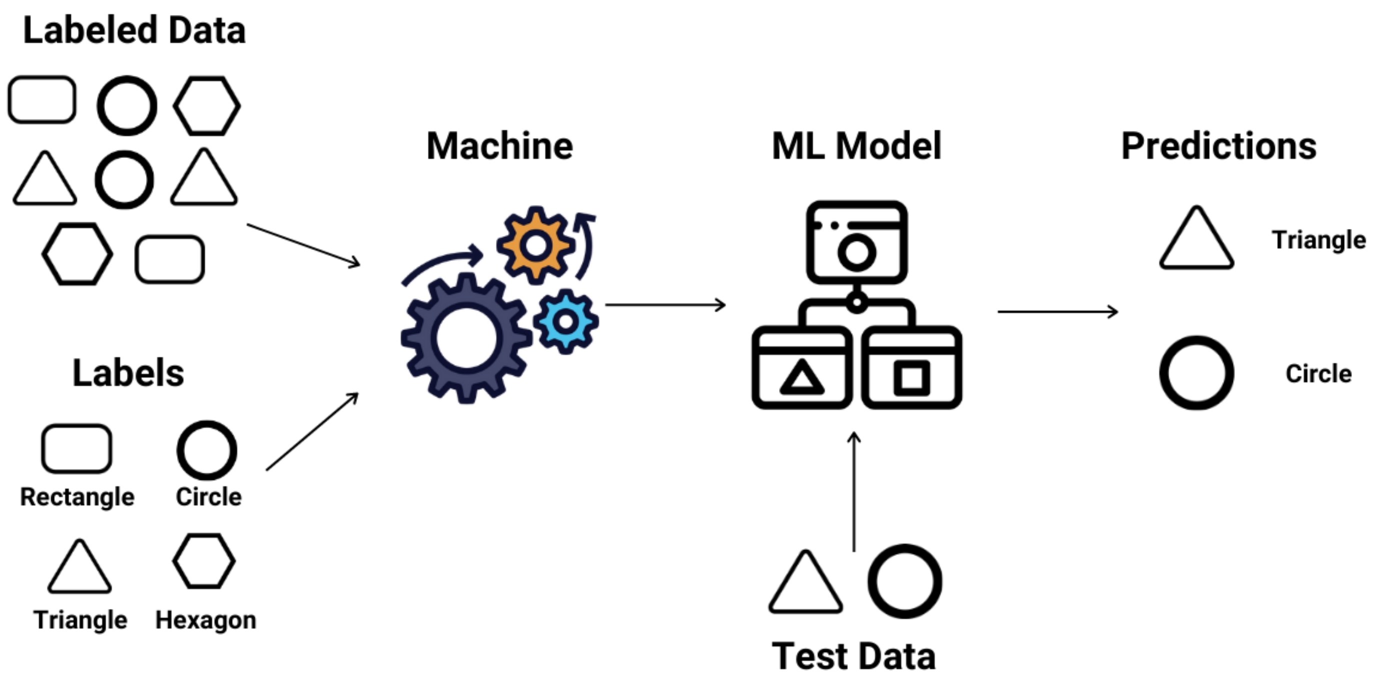
Types of Learning

Supervised Learning

Unsupervised Learning

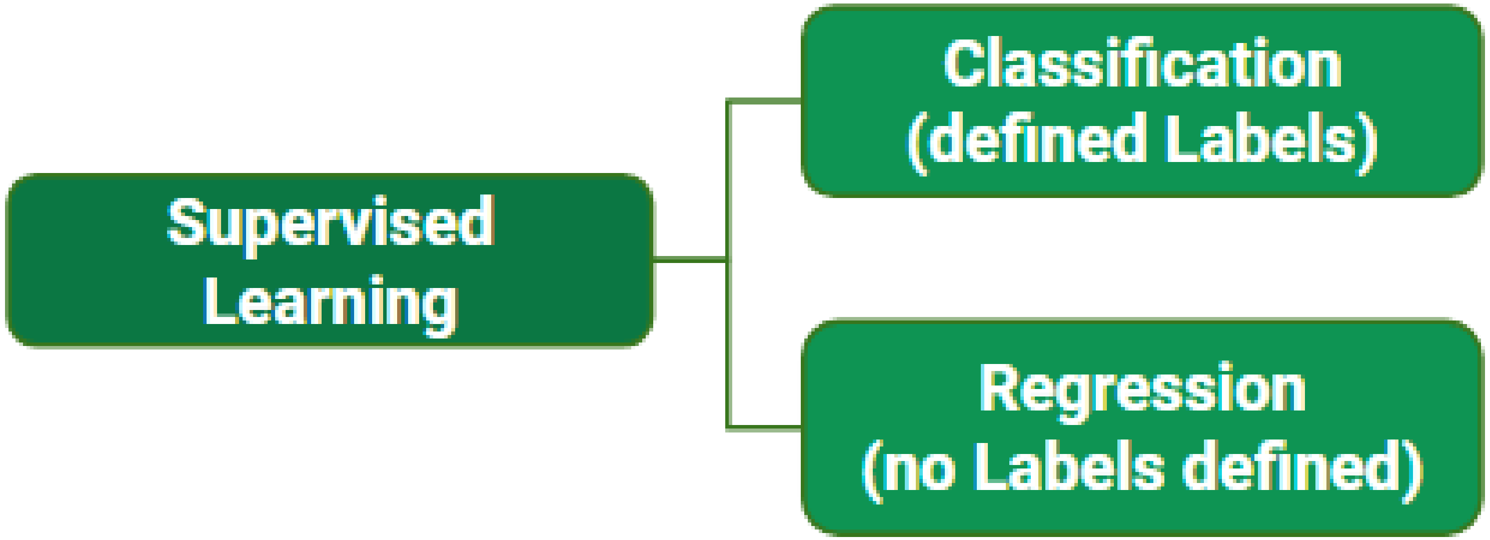
Semi-Supervised Learning

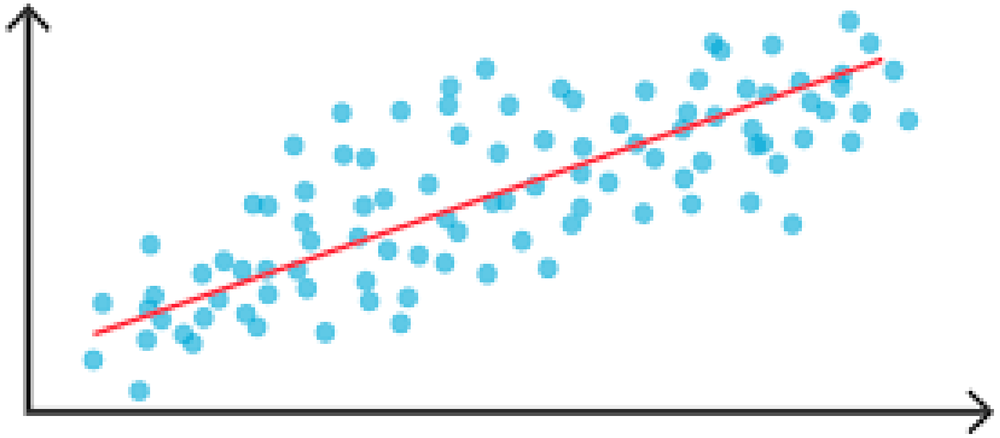
Supervised Learning



Learning by ‘known’ example

Popular Tasks (T)





Regression

Fitting the “best” line

Introduction to Linear Regression

Linear Regression is one of the simplest and most widely used algorithms in machine learning.

It models the relationship between a dependent variable (target) and one or more independent variables (features)

It does so by fitting a linear equation to observed data.

Applications:

Predicting house prices based on features like size, location, etc.

Estimating sales based on advertising expenditure.

Predicting student scores based on study hours.

Etc.

Types of Linear Regression

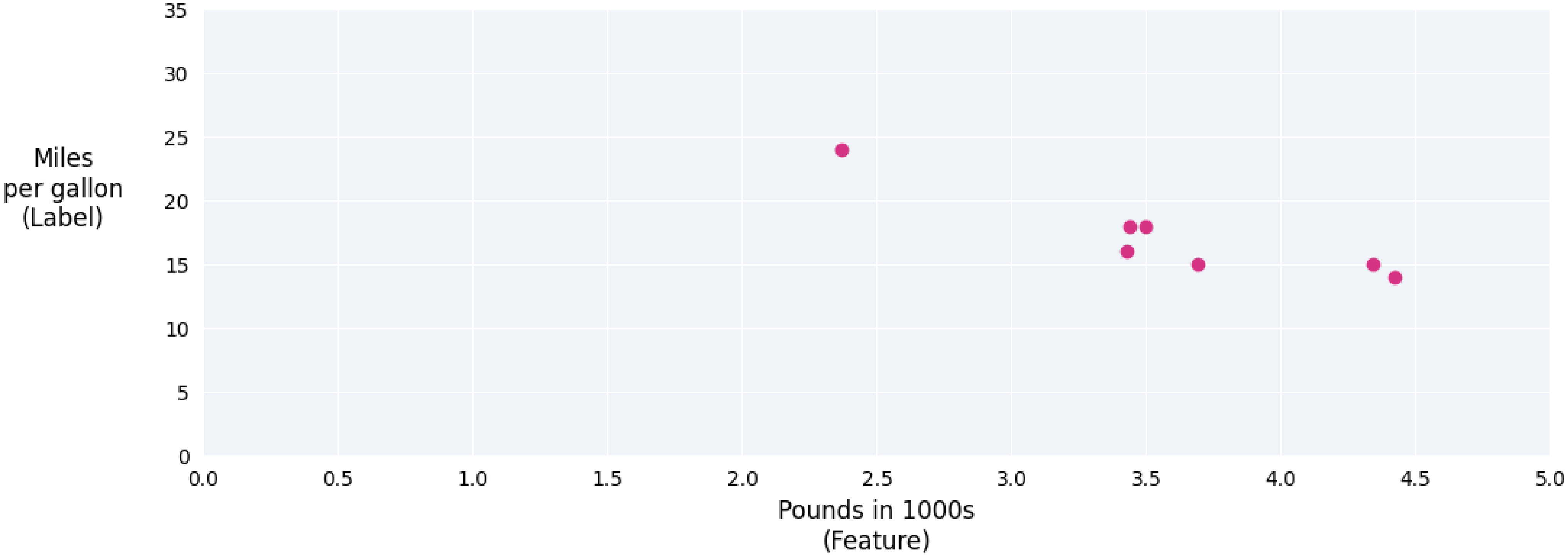
- **Simple Linear Regression:**
  - Involves a single independent variable.
  - The model assumes a linear relationship between the input feature X and the output Y.
  - The model is represented by the equation:
    - $Y = \beta_0 + \beta_1 X + \epsilon$
  - where:
    - β_0 is the intercept (the value of $Y\ when\ X = 0$).
    - β1 is the slope of the line (change in Y for a unit change in X).
    - ϵ is the error term (captures the noise in the data).
- **Multiple Linear Regression:**
  - Involves multiple independent variables.
  - The model assumes a linear relationship between several input features $X_1, X_2, \ldots, X_n$ and the output Y.
  - The model is represented by the equation:
    - $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon$

Relationship between Features and Variables

- Linear regression is a statistical technique used to find the relationship between variables.
- In an ML context, linear regression finds the relationship between features and a label.
- For example:
  - Predict a car's fuel efficiency in miles per gallon based on how heavy the car is

Pounds in 1000s (feature)	Miles per gallon (label)
3.5	18
3.69	15
3.44	18
3.43	16
4.34	15
4.42	14
2.37	24

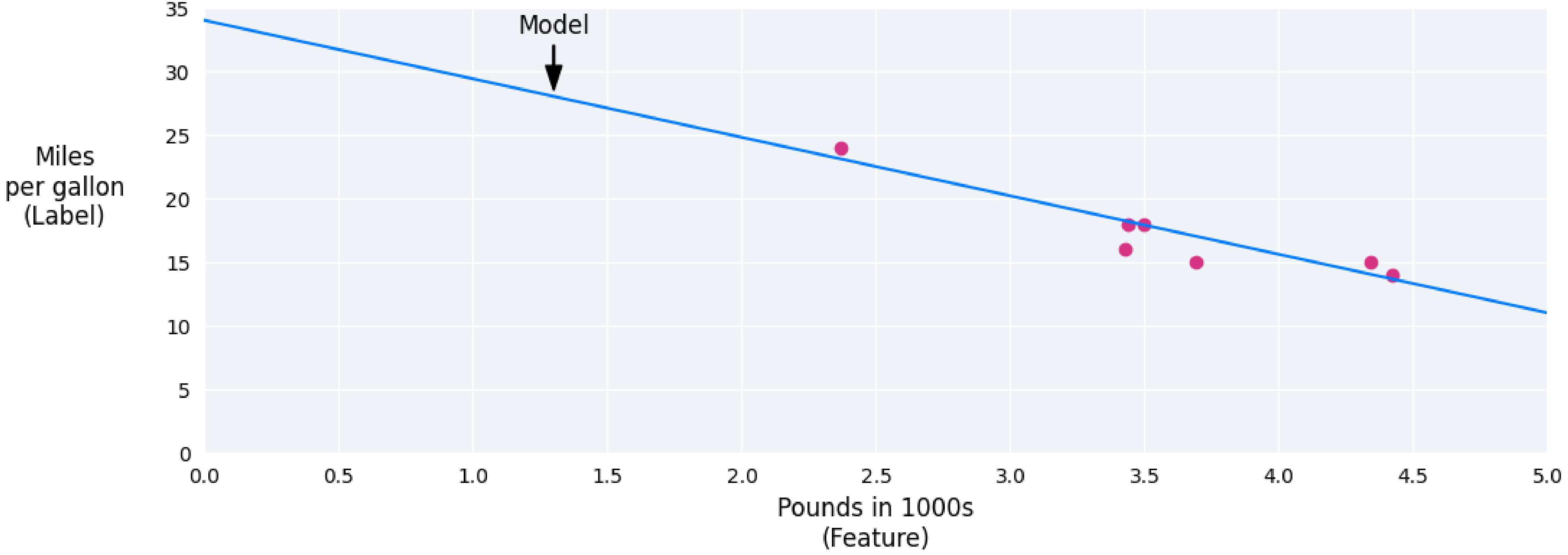
Relationship between Features and Variables



Car heaviness (in pounds) versus miles per gallon rating. As a car gets heavier, its miles per gallon rating generally decreases.

Regression Model

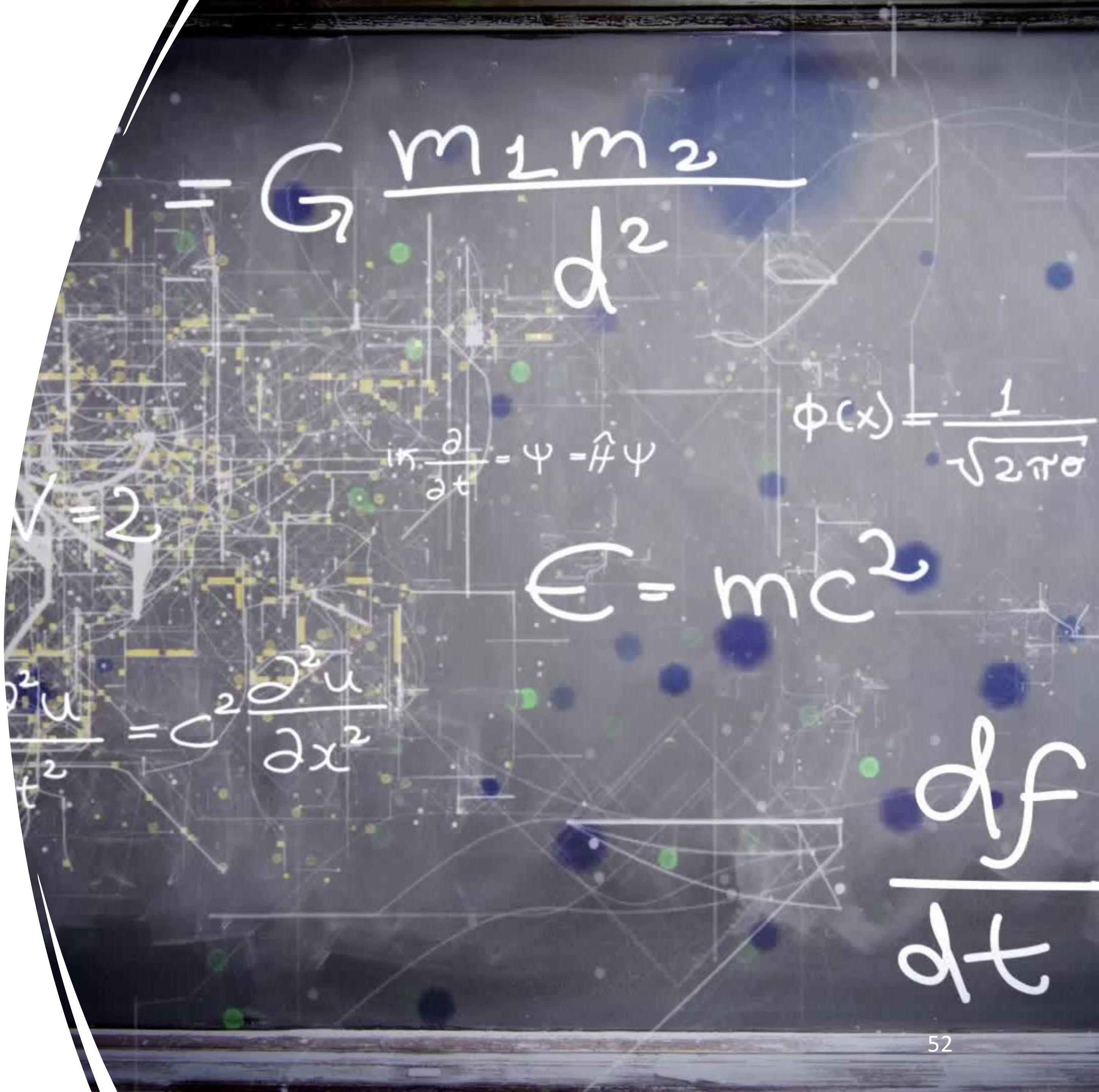
- We could create our own model by drawing a best fit line through the points.



Linear regression equation

In algebraic terms, the model would be defined as $y = mx + b$, where

- y is miles per gallon—the value we want to predict.
- m is the slope of the line.
- x is pounds—our input value.
- b is the y-intercept.



Machine Learning Model Equation

In ML, we write the equation for a linear regression model as follows:

$$y' = b + w_1 x_1$$

where:

- y' is the predicted label—the output.
- b is the **bias** of the model. Bias is the same concept as the y-intercept in the algebraic equation for a line. In ML, bias is sometimes referred to as w_0. Bias is a **parameter** of the model and is calculated during training.
- w_1 is the **weight** of the feature. Weight is the same concept as the slope m in the algebraic equation for a line. Weight is a **parameter** of the model and is calculated during training.
- x_1 is a **feature**—the input.

Mathematical representation of a linear model.

$$y' = b + w_1 x_1$$

Prediction Bias Weight Feature value

Calculated from training

Models with multiple features

For example, a model that relies on five features would be written as follows:

$$y' = b + w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + w_5x_5$$

For example, a model that predicts gas mileage could additionally use features such as the following:

- Engine displacement
- Acceleration
- Number of cylinders
- Horsepower

A model with five features to predict a car's miles per gallon rating.

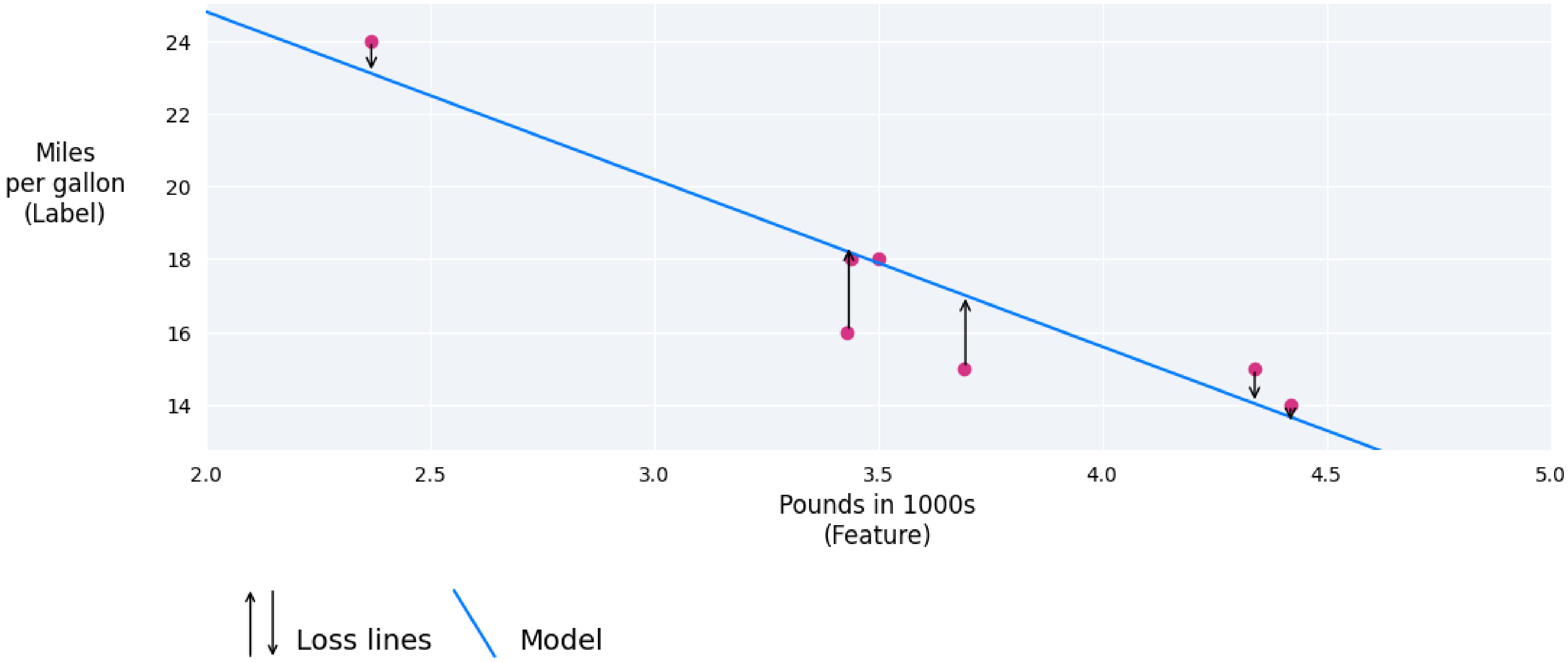
$$y' = b + w_1\boxed{x_1} + w_2\boxed{x_2} + w_3\boxed{x_3} + w_4\boxed{x_4} + w_5\boxed{x_5}$$

Pounds Displacement Acceleration Number of cylinders Horsepower

Linear regression: Loss

Loss is a numerical metric that describes how wrong a model's predictions are.

Loss is measured from the actual value to the predicted value.

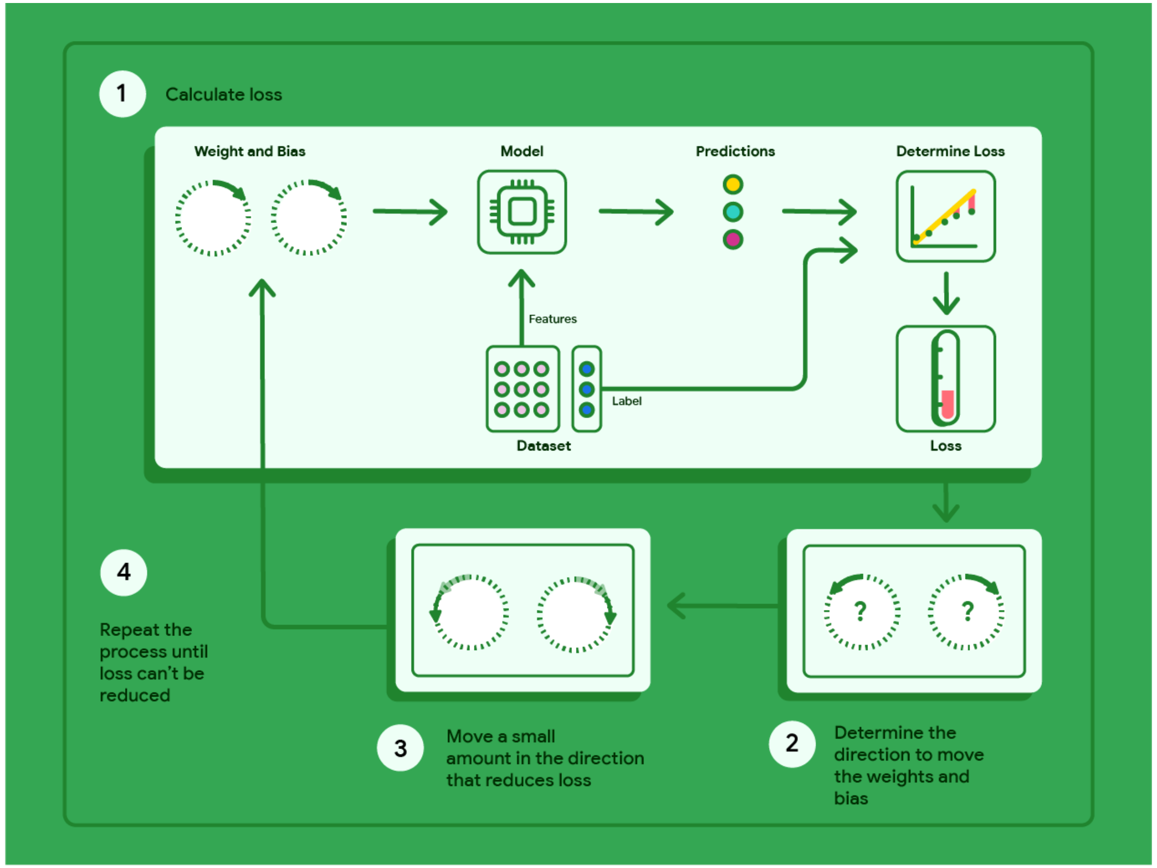


Types of loss

Loss type	Definition	Equation
L_1 loss	The sum of the absolute values of the difference between the predicted values and the actual values.	$\sum \lvert actual\ value - predicted\ value \rvert$
Mean absolute error (MAE)	The average of L_1 losses across a set of *N* examples.	$\frac{1}{N} \sum \lvert actual\ value - predicted\ value \rvert$
L_2 loss	The sum of the squared difference between the predicted values and the actual values.	$\sum (actual\ value - predicted\ value)^2$
Mean squared error (MSE)	The average of L_2 losses across a set of *N* examples.	$\frac{1}{N} \sum (actual\ value - predicted\ value)^2$
Root mean squared error (RMSE)	The square root of the mean squared error (MSE).	$\sqrt{\frac{1}{N} \sum (actual\ value - predicted\ value)^2}$

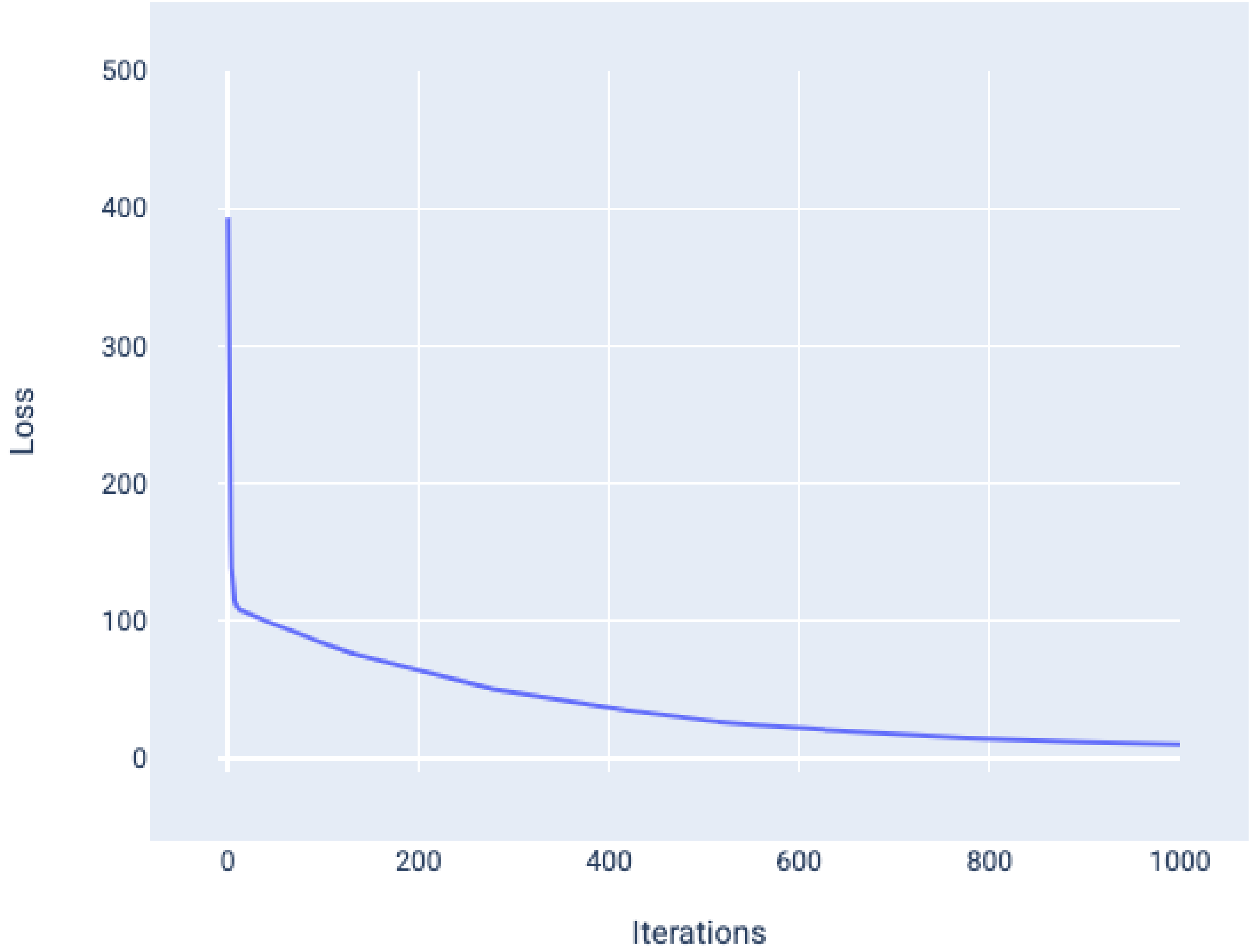
Linear regression: Gradient descent

- Gradient descent is a mathematical technique that iteratively finds the weights and bias that produce the model with the lowest loss.
- The model begins training with randomized weights and biases near zero, and then repeats the following steps:
    1. Calculate the loss with the current weight and bias.
    2. Determine the direction to move the weights and bias that reduce loss.
    3. Move the weight and bias values a small amount in the direction that reduces loss.
    4. Return to step one and repeat the process until the model can't reduce the loss any further.

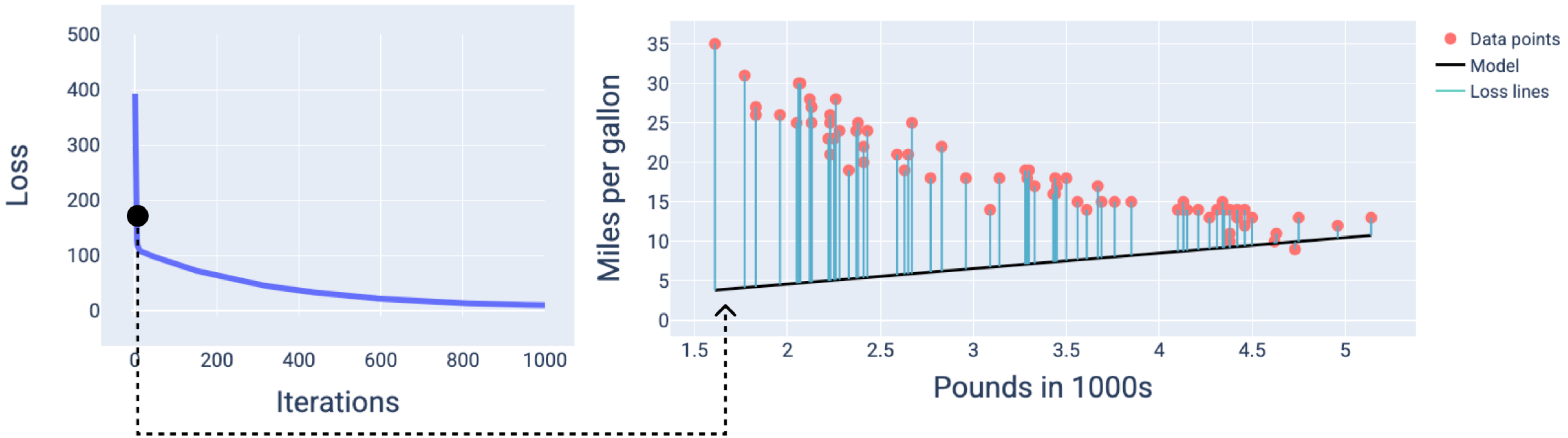


Model convergence and loss curves

- When training a model, you'll often look at a **loss curve** to determine if the model has **converged**.

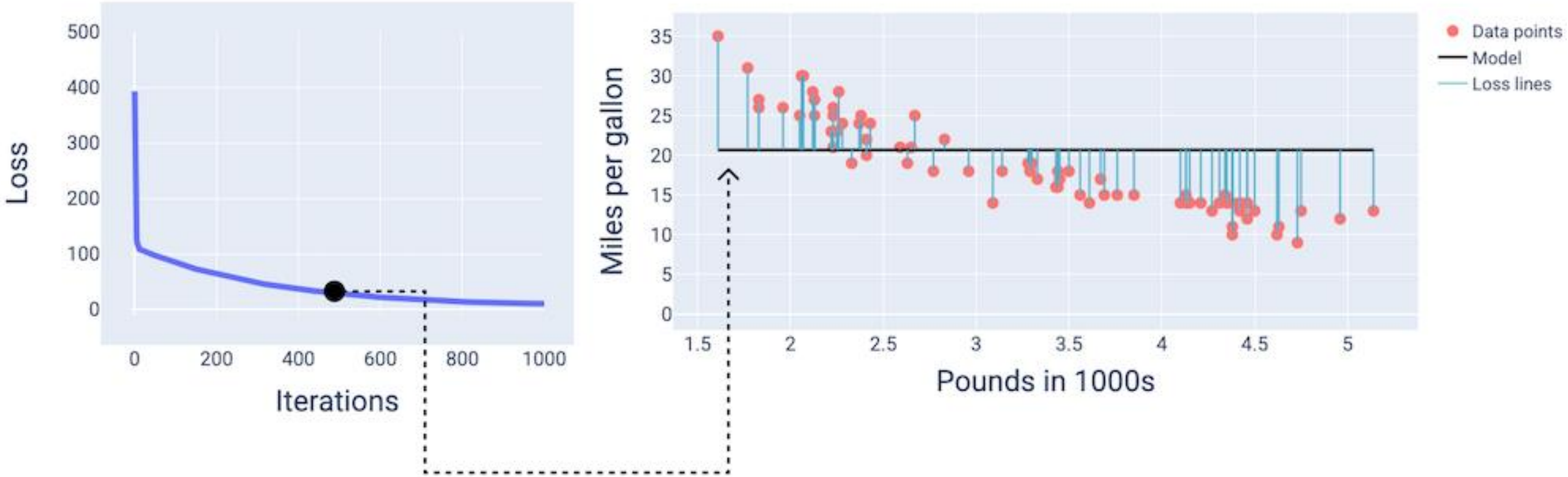


Loss curve and snapshot of the model at the beginning of the training process.



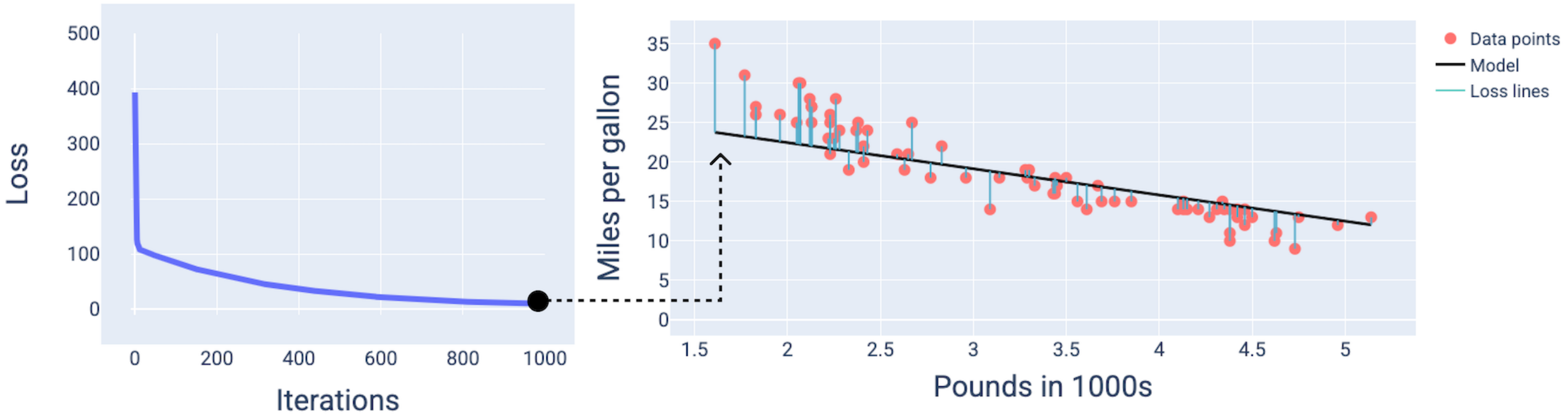
In the following figure, we can see that around the second iteration the model would not be good at making predictions because of the high amount of loss.

After 400 iterations



At around the 400th-iteration, we can see that gradient descent has found the weight and bias that produce a better model.

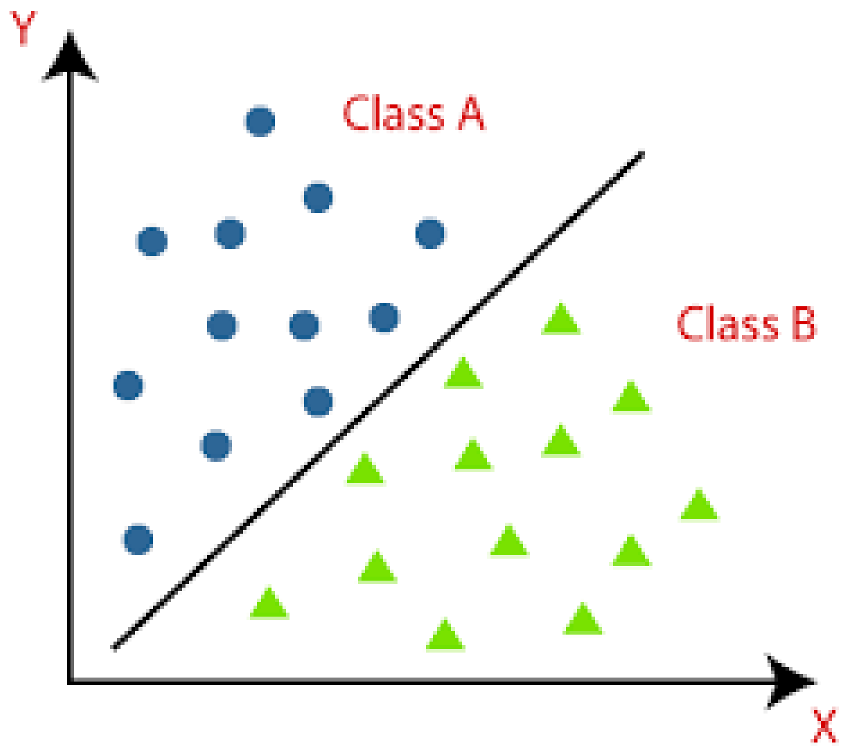
After 1,000th iteration



And at around the 1,000th-iteration, we can see that the model has converged, producing a model with the lowest possible loss.

Classification

Another Supervised Task



Introduction to Classification

- **Classification** is a supervised learning task where the goal is to predict the categorical label of input data.
  - Unlike regression, where the output is a continuous value,
  - Classification outputs discrete values (categories or classes).

Types of Classification

Binary Classification:

- Only two possible classes.
- Examples: Spam detection, cancer diagnosis (malignant/benign).

Multiclass Classification:

- More than two classes.
- Examples: Handwritten digit recognition (0-9), animal species classification.

Multilabel Classification:

- Each instance can belong to multiple classes simultaneously.
- Examples: Tagging a news article with multiple topics, identifying all objects in an image.

Classification Algorithms

- Logistic Regression
- k-Nearest Neighbors (k-NN)
- Decision Trees
- Random Forest
- Support Vector Machines (SVM)
- Neural Networks (NN)
- etc

Common Challenges in Classification

Class Imbalance:

- When one class significantly outnumbers the other(s), leading to biased models.
- Solutions:
  - Resampling techniques (oversampling minority class, undersampling majority class),
  - using appropriate evaluation metrics (e.g., F1-score).

Choosing the Right Algorithm:

- No one-size-fits-all; choice depends on the dataset size, feature types, and problem complexity.
- Solution:
  - Experimentation and cross-validation are key.

Interpretability vs. Accuracy:

- Simpler models like logistic regression or decision trees are more interpretable.
- Complex models like neural networks might offer higher accuracy but are less interpretable.

Examples

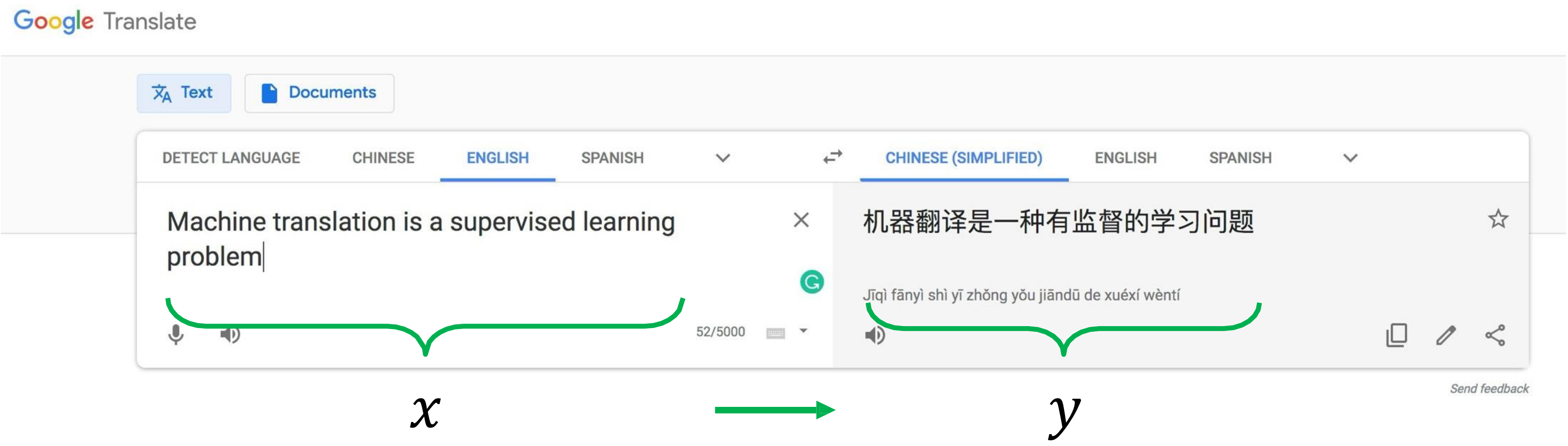
Supervised Learning in Computer Vision

- Image Classification
  - $x =$ raw pixels of the image, $y =$ the main object

ImageNet Large Scale Visual Recognition Challenge. Russakovsky et al.'2015

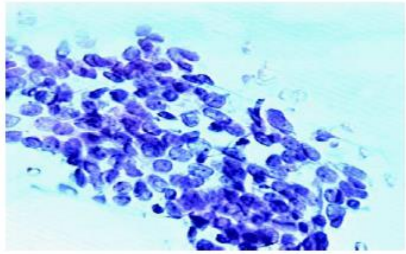
Supervised Learning in Natural Language Processing

- Machine translation

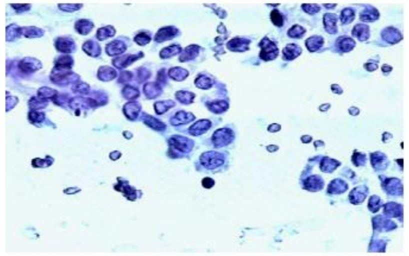


▶ Note: this course only covers the basic and fundamental techniques of supervised learning (which are not enough for solving hard vision or NLP problems.)

Example from biomedical domain



Benign cells



Malignant cells

Breast cancer prognosis from a microscopic image of a needle biopsy

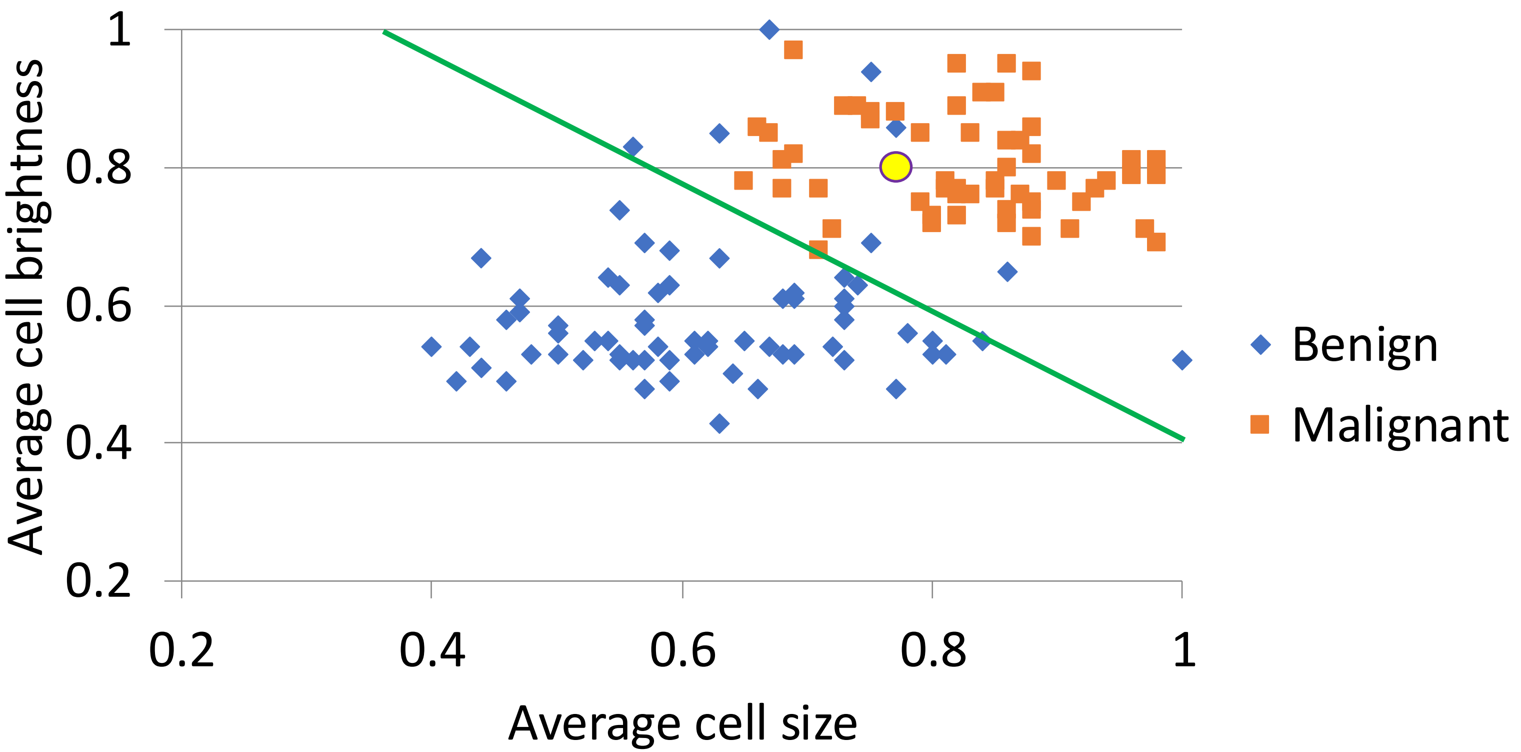
Goal: predict automatically from a tissue sample whether a patient has cancer or not

In order to “generalize”, we need some **training examples** of benign as well as malignant tissues

- Data collected from patients forms the **dataset**. Each patient constitutes one learning example <image,label>

example from Y. Saeys, UGent

Approach



a linear decision boundary

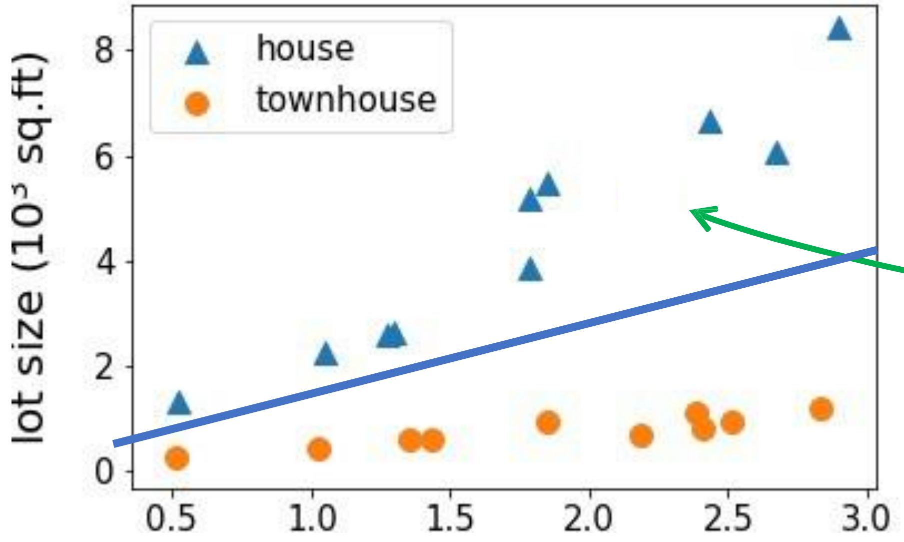
Regression vs Classification

regression: if $y \in \mathbb{R}$ is a continuous variable

- e.g., price prediction

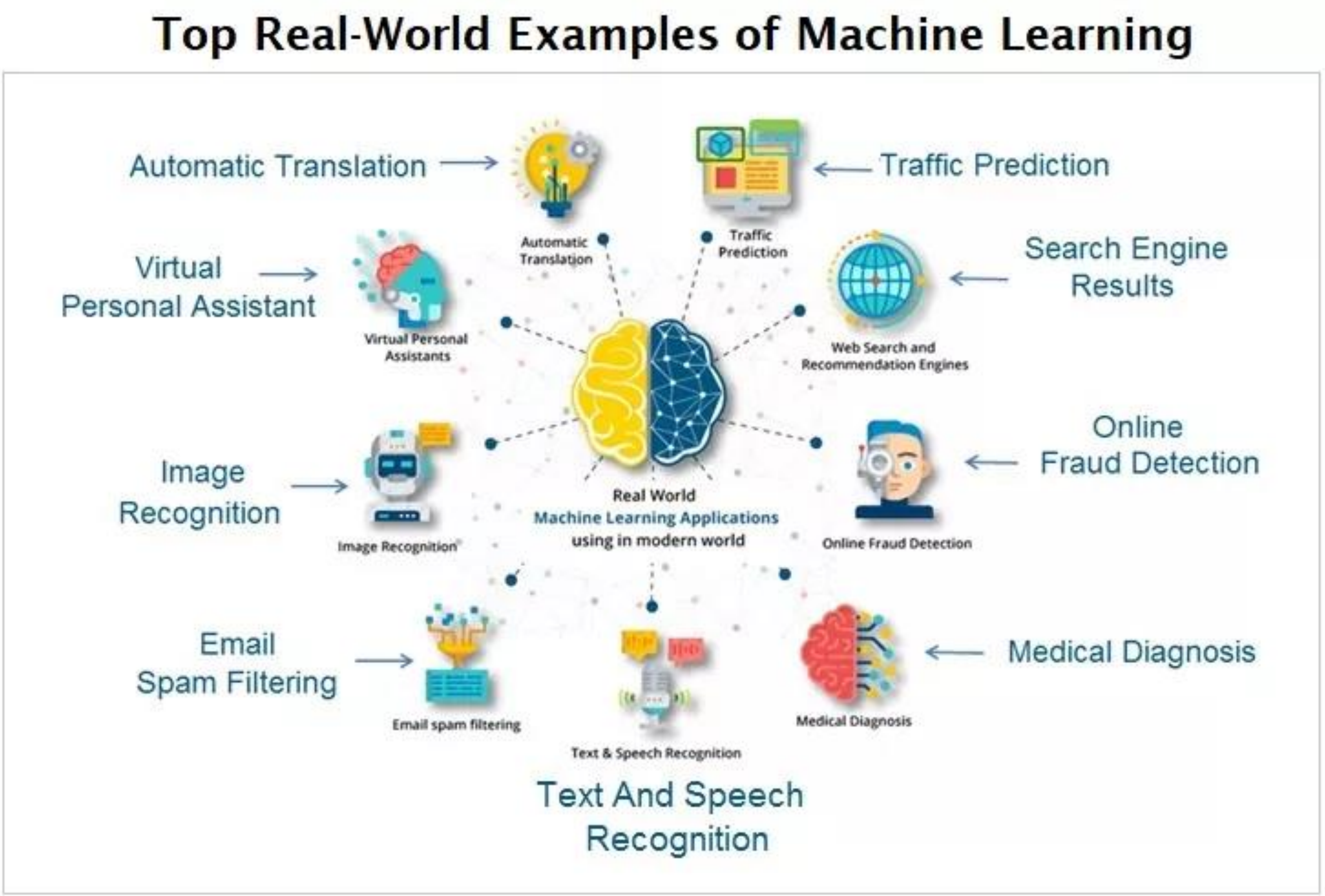
classification: the label is a discrete variable

- e.g., the task of predicting the types of residence
  - (size, lot size) → house or townhouse?



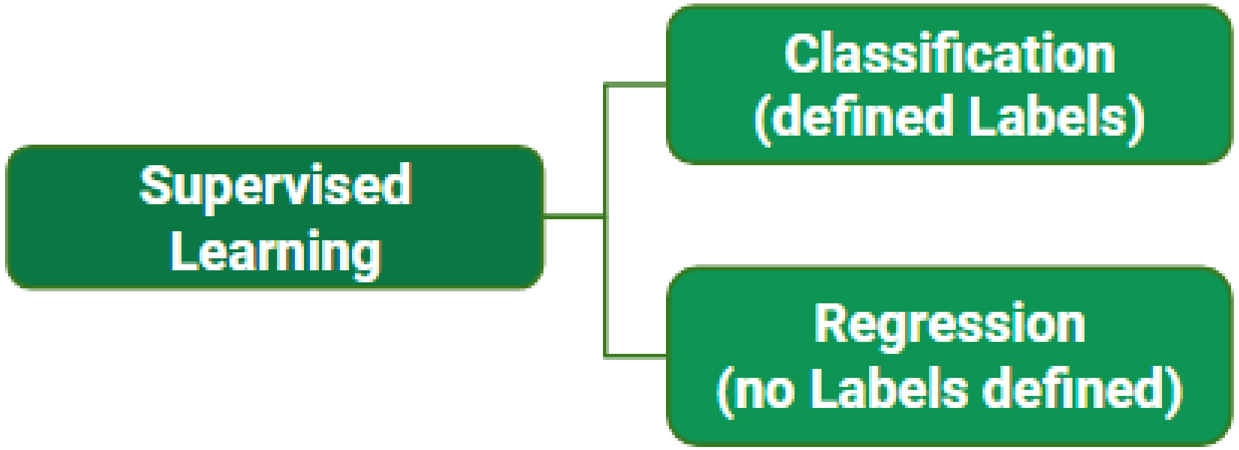
$y =$ house or townhouse?

Supervised Learning -- Many Application areas



https://www.aiiottalk.com/machine-learning-examples-in-real-life/

Supervised Learning: Summary



- You always have labeled data
- Most common tasks:
  - Regression
    - Fid the “best” line to fit the data.
  - Classification
    - Find the “best” line to separate the data.
- Many application areas

Challenges in Supervised Learning

Data Quality: The model's performance heavily depends on the quality and quantity of labeled data.

Labeling Costs: Acquiring labeled data can be time-consuming and expensive, especially for large datasets.

Overfitting: There's a risk that the model may perform well on training data but fail to generalize to new data.

Taxonomy of Machine Learning

Supervised Learning

Unsupervised Learning

Semi-Supervised Learning

Introduction to Unsupervised Learning

- **Unsupervised Learning** is a type of machine learning where the algorithm is trained on unlabeled data.
- Unsupervised learning involves identifying patterns, relationships, or structures within the data without the need for explicit labels.

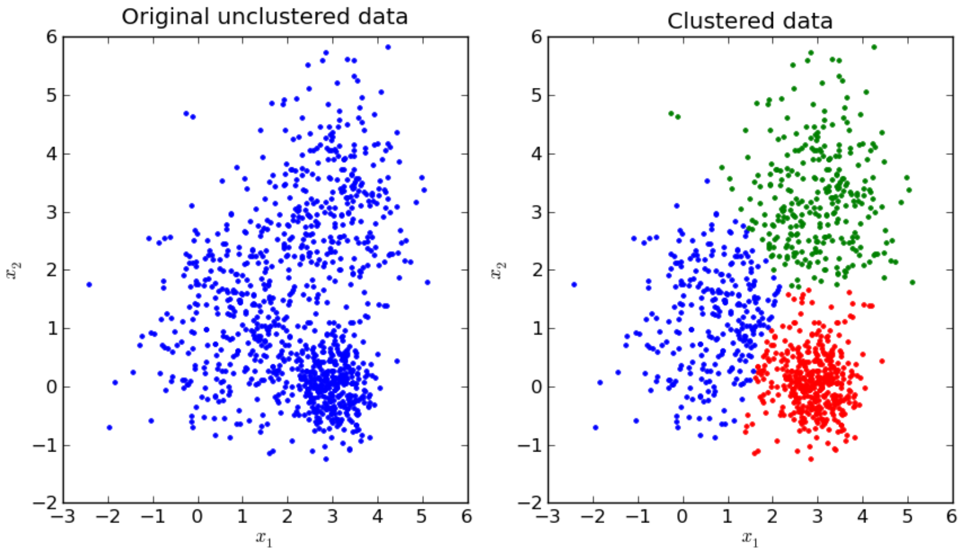
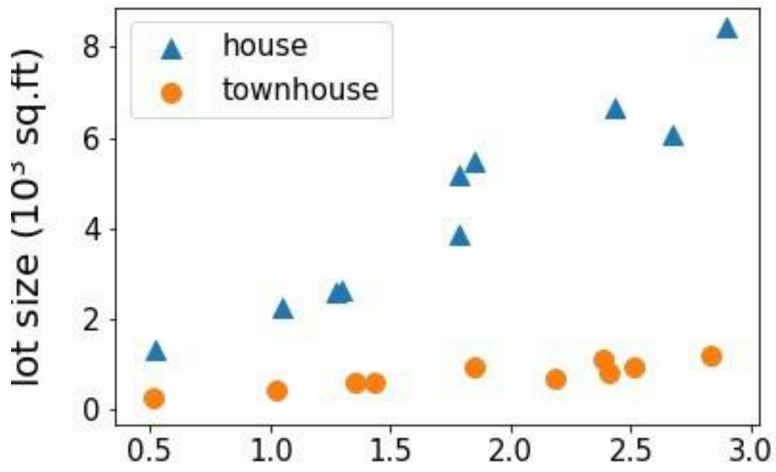


Image source: https://medium.com/ascentic-technology/unsupervised-machine-learning-kmeans-clustering-with-scikit-learn-bc8895cd66a8

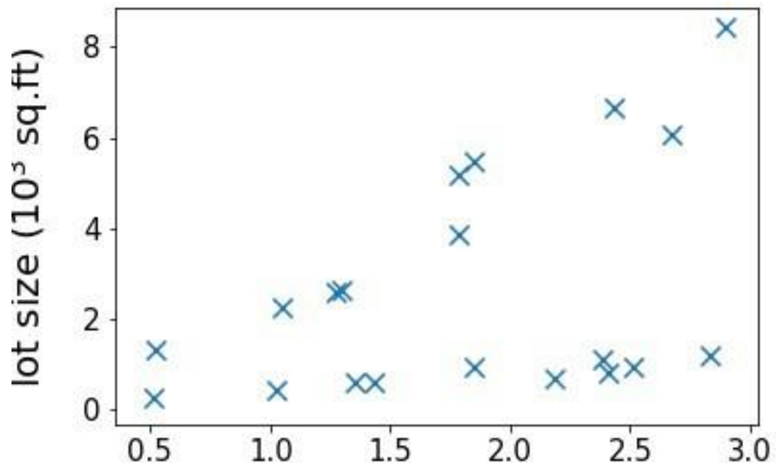
Unsupervised Learning

- Dataset contains no labels: $x^{(1)}, \ldots x^{(n)}$
- Goal (vaguely-posed): to find interesting structures in the data

supervised



unsupervised



Applications

Computer vision:

Unsupervised learning algorithms are used for visual perception tasks, such as object recognition.

Medical imaging:

Unsupervised machine learning provides essential features to medical imaging devices to diagnose patients quickly and accurately.

Anomaly detection:

Unsupervised learning models can comb through large amounts of data and discover atypical data points within a dataset. These anomalies can raise awareness around faulty equipment, human error, or breaches in security.

Customer personas:

Defining customer personas makes it easier to understand common traits and business clients' purchasing habits.

Recommendation Engines:

Using past purchase behavior data, unsupervised learning can help to discover data trends that can be used to develop more effective cross-selling strategies.

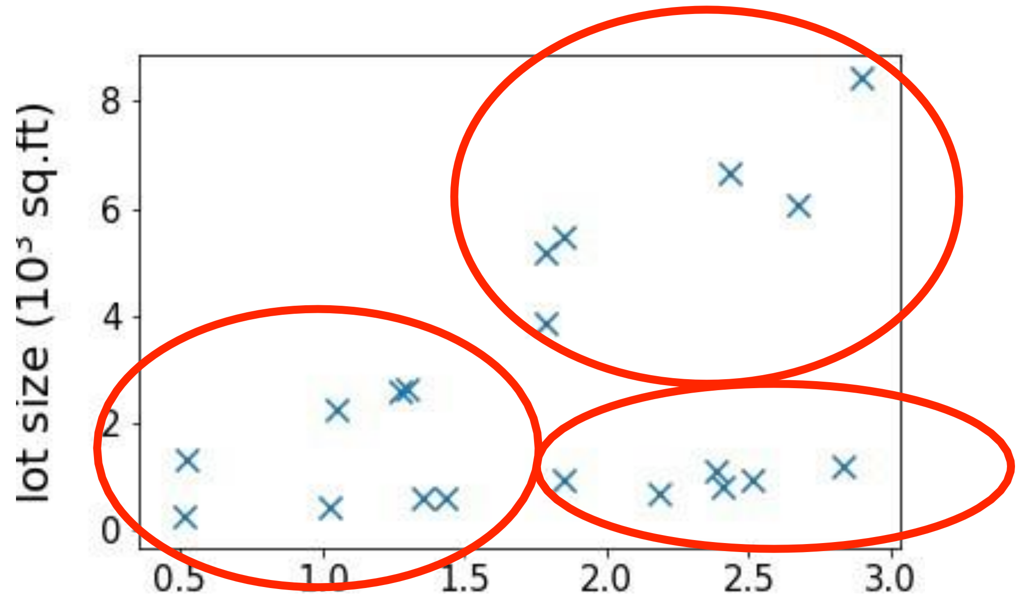
News Sections:

Google News uses unsupervised learning to categorize articles on the same story from various online news outlets.

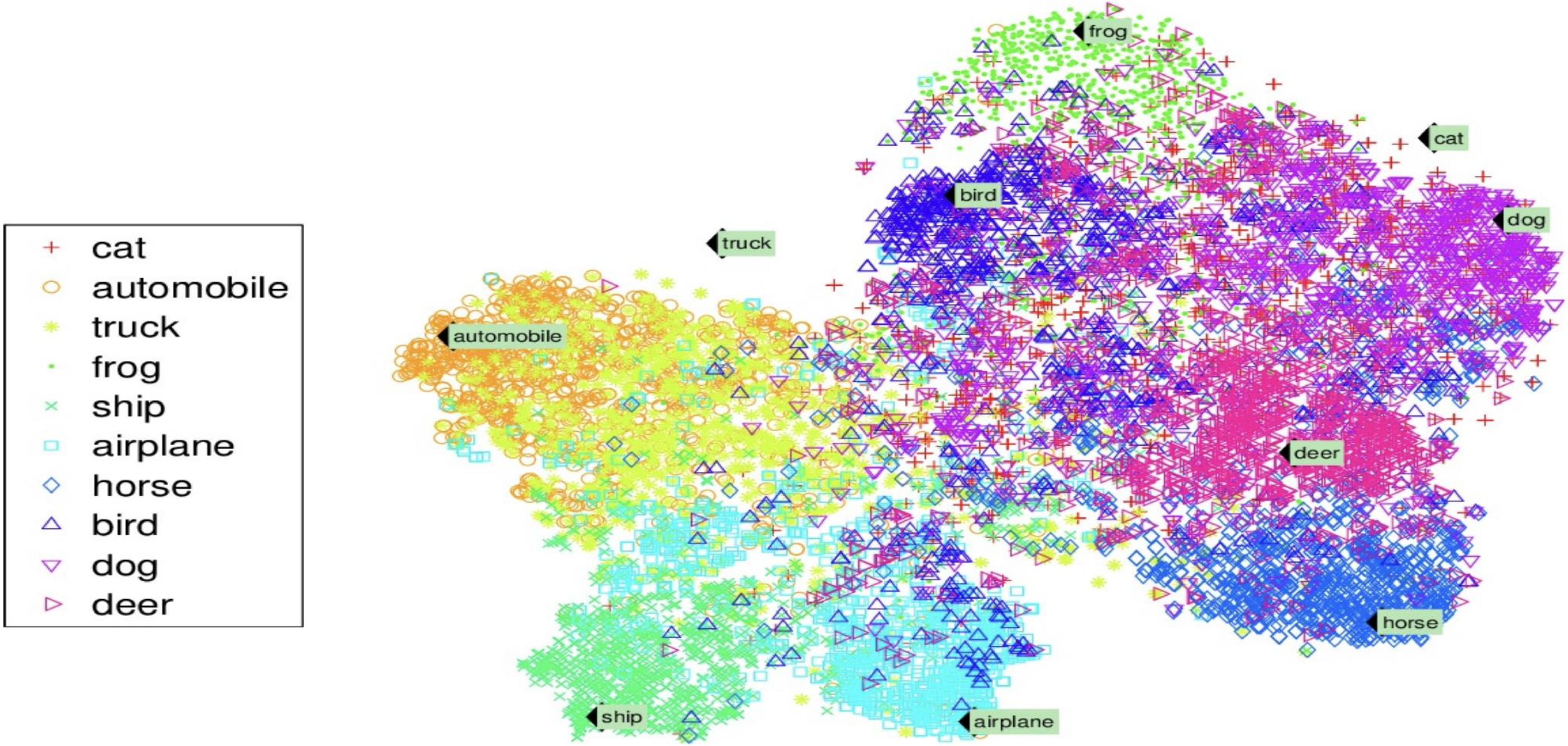
Types of Unsupervised Learning

- Clustering (k-mean)
- Dimensionality Reduction (PCA)
- Anomaly Detection
- Association Rule Learning

Clustering

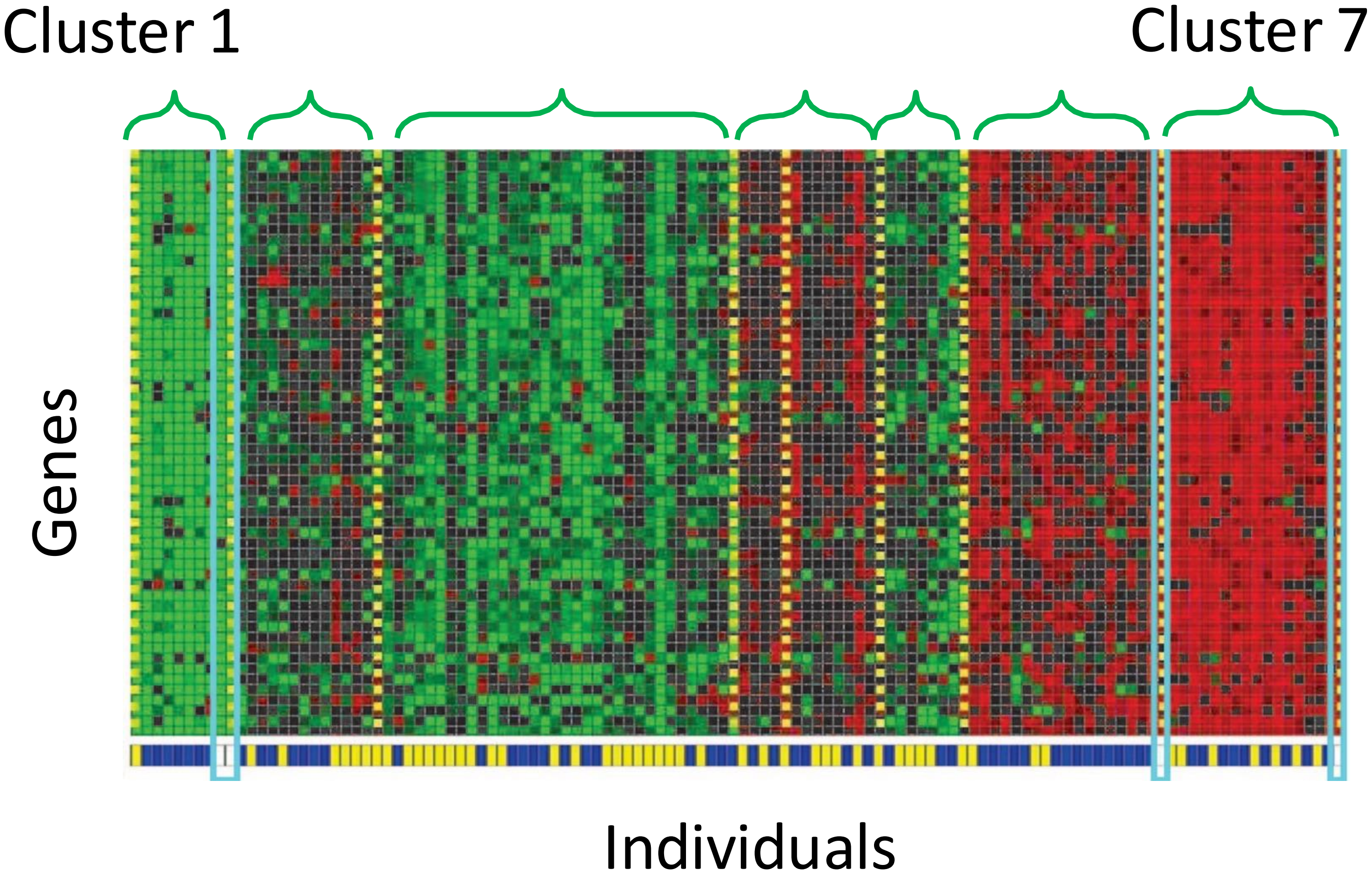


Unsupervised learning



A t-SNE visualization highlighting semantic clusters

Clustering Genes

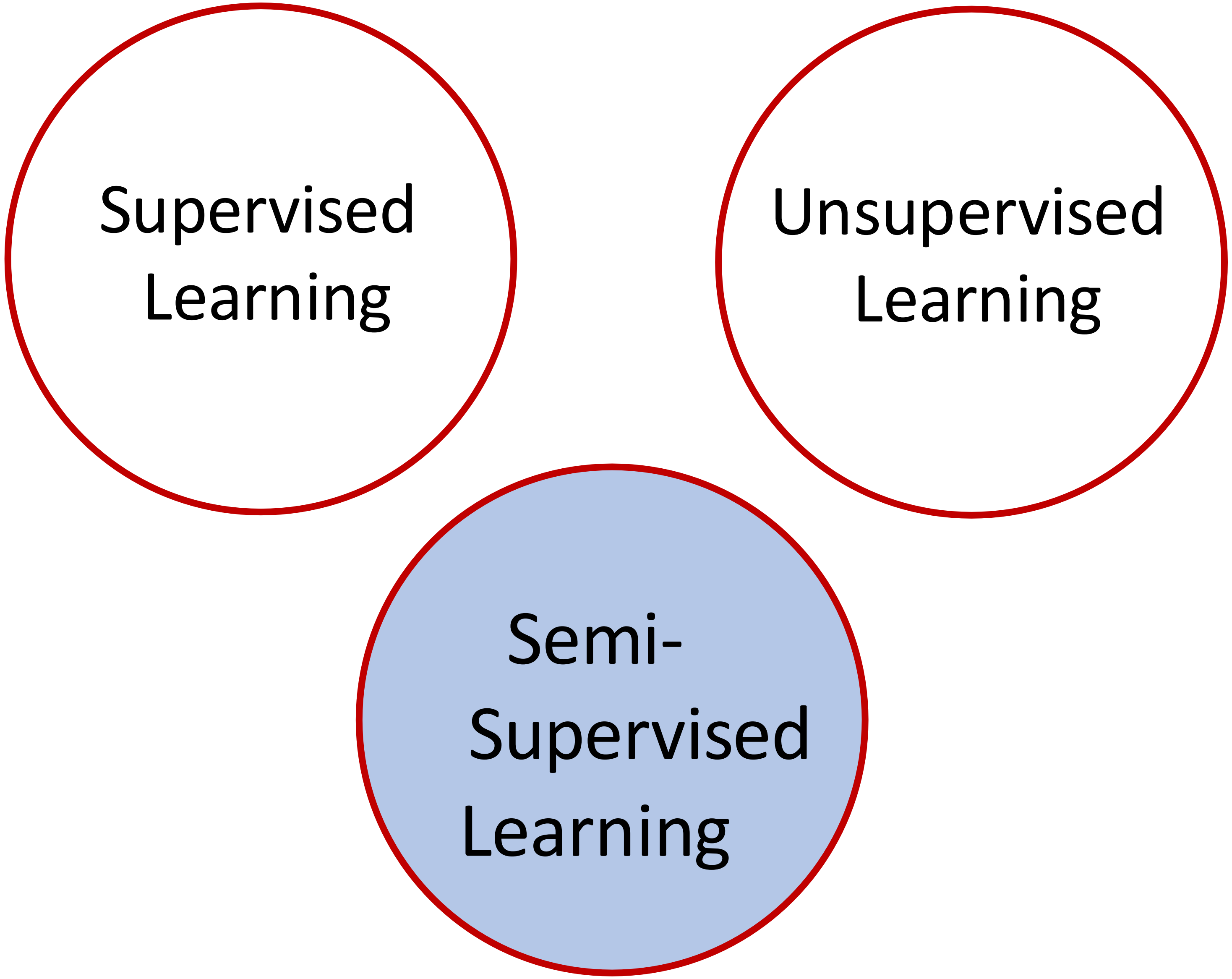


Identifying Regulatory Mechanisms using Individual Variation Reveals Key Role for Chromatin Modification. [Su-In Lee, Dana Pe'er, Aimee M. Dudley, George M. Church and Daphne Koller. '06]

Challenges of unsupervised learning

- Computational complexity due to a high volume of training data
- Longer training times
- Higher risk of inaccurate results
- Human intervention to validate output variables
- Lack of transparency into the basis on which data was clustered

Taxonomy of Machine Learning



Introduction to Semi-Supervised Learning

Semi-supervised learning is a machine learning paradigm that lies between supervised and unsupervised learning.

It uses a small amount of labeled data combined with a large amount of unlabeled data to build models.

This approach is particularly useful when acquiring labeled data is expensive or time-consuming, but unlabeled data is plentiful.

Why Semi-Supervised Learning?

Labeled Data is Expensive:

Labeling data often requires expert knowledge (e.g., medical diagnosis, legal text classification), making it costly and time-consuming.

Unlabeled Data is Plentiful:

In many domains, vast amounts of unlabeled data are readily available (e.g., text documents, and images on the internet), which can be leveraged to improve model performance.

Improved Performance:

By incorporating unlabeled data, models can generalize better, especially when labeled data is scarce, leading to improved performance on real-world tasks.

Techniques Used in Semi-Supervised Learning: Pseudo-labeling

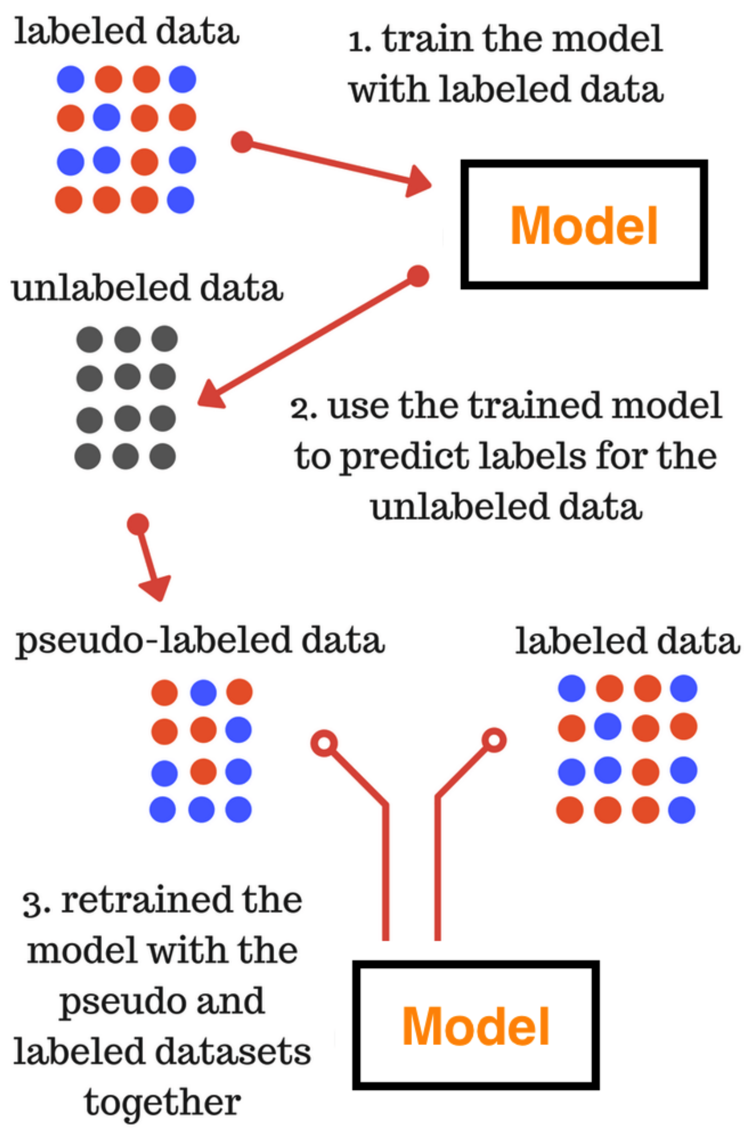


Image source: https://blog.roboflow.com/what-is-semi-supervised-learning/

One more type of learning….

Reinforcement Learning

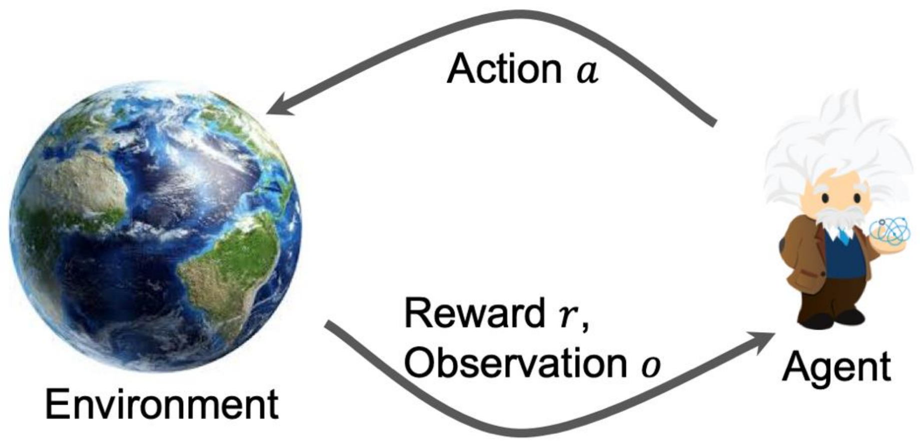
Not in this course

What is Reinforcement Learning (RL)

RL is a general-purpose framework for sequential decision-making

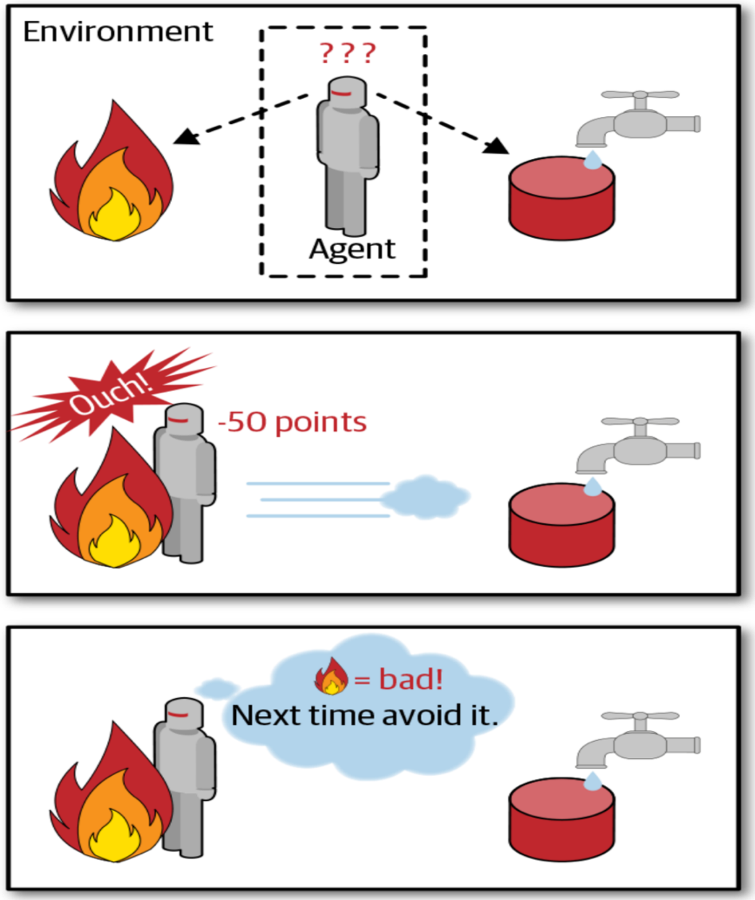
Usually described as an agent interacting with an unknown environment

Goal: select the action to maximize a future cumulative reward



https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/lectures/lecture16-guest.pdf

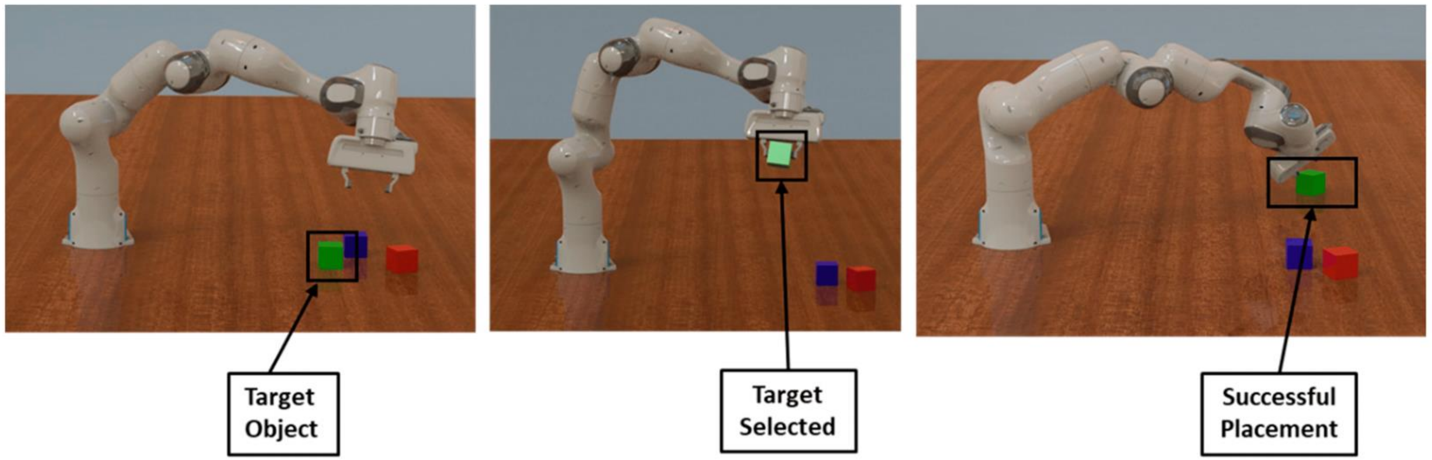
Reinforcement Learning



1. Observe
2. Select action using policy
3. Action!
4. Get reward or penalty
5. Update policy (learning step)
6. Iterate until an optimal policy is found

Applications of Reinforcement Learning

Automated Robots



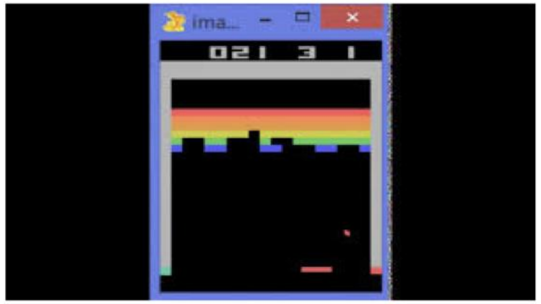
https://bair.berkeley.edu/blog/2020/04/27/ingredients/

Motor Control

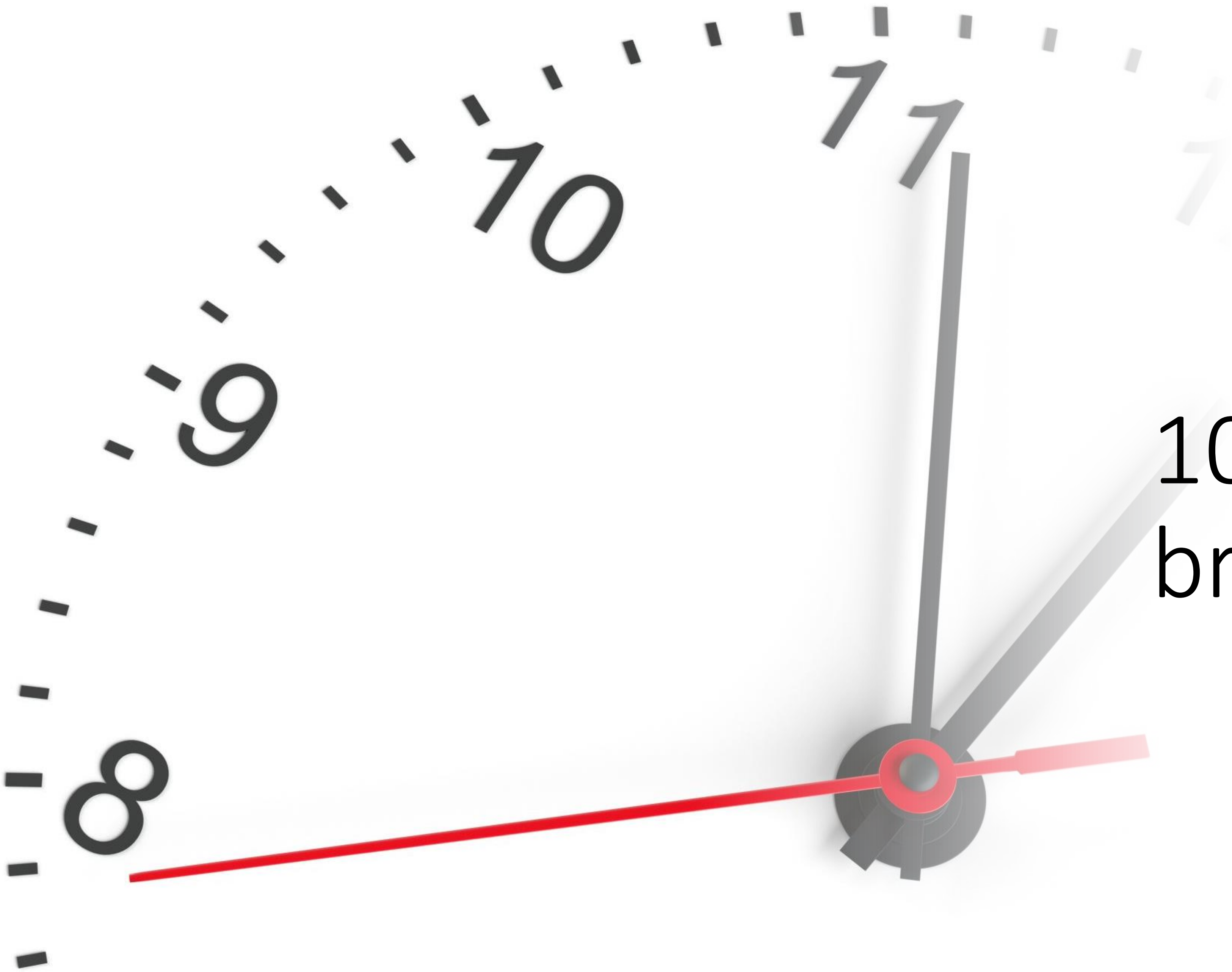
- Observations: images from the camera, joint angle
- Actions: joint torques
- Rewards: navigate to target location, serve and protect humans

https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/lectures/lecture16-guest.pdf

Games



https://web.stanford.edu/class/archive/cs/cs224n/cs224n.1184/lectures/lecture16-guest.pdf



10 minutes break

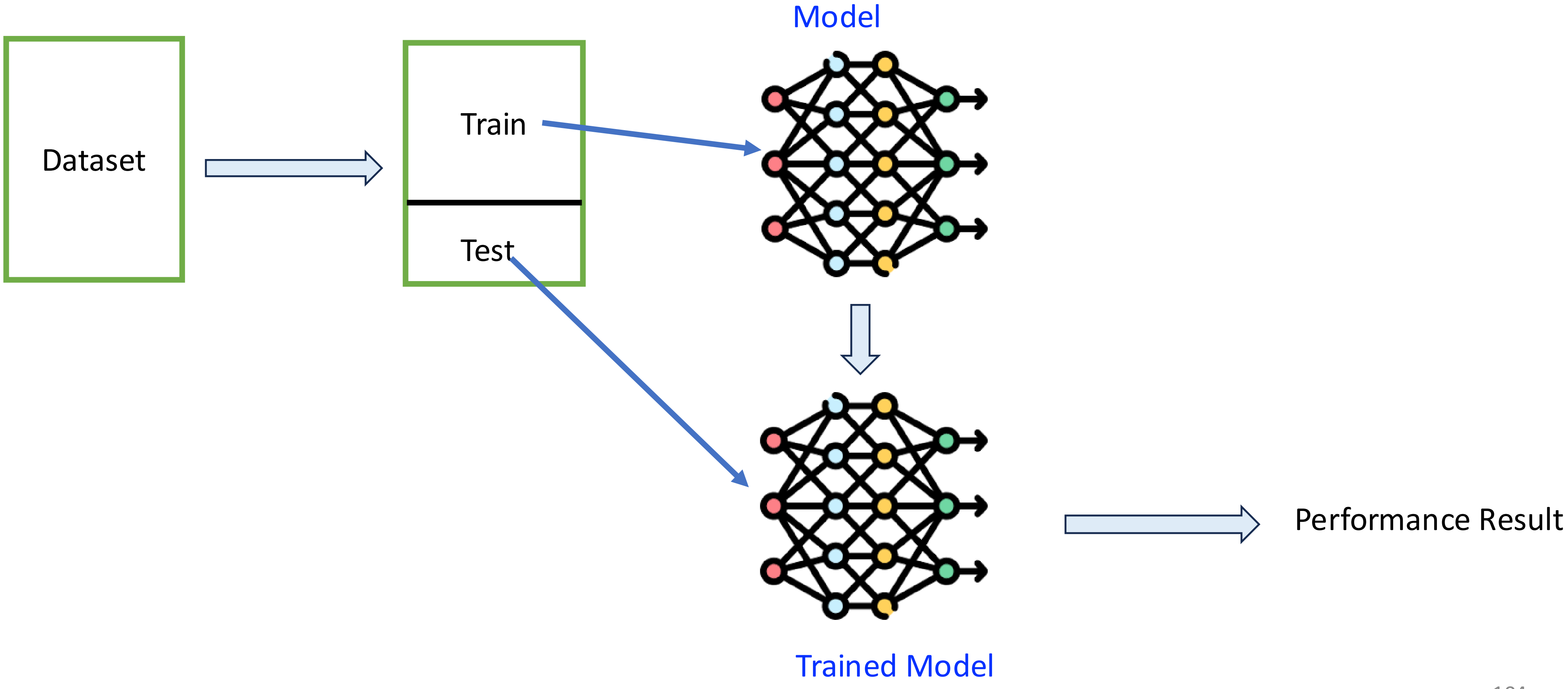
Evaluation

Model evaluation

- Instead of training the model directly on all available training data, split the data in a **training set** and a **test set**
- Next, train the model only using the training set, and use the examples in the test set to assess the goodness of the model
- Generalization of this idea: **k-fold cross validation**
  - Partition data into k disjoint subsets
  - for i = 1 to k

    hold out subset i for testing and train on the remaining subsets
  - Report average performance

Model evaluation



10-fold cross validation

ENTIRE DATASET

	fold 1	fold 2	fold 3	fold 4	fold 5	fold 6	fold 7	fold 8	fold 9	fold 10
1	test	train	train	train	train	train	train	train	train	train
2	train	test	train	train	train	train	train	train	train	train
3	train	train	test	train	train	train	train	train	train	train
4	train	train	train	test	train	train	train	train	train	train
5	train	train	train	train	test	train	train	train	train	train
6	train	train	train	train	train	test	train	train	train	train
7	train	train	train	train	train	train	test	train	train	train
8	train	train	train	train	train	train	train	test	train	train
9	train	train	train	train	train	train	train	train	test	train
10	train	train	train	train	train	train	train	train	train	test

compute average accuracy across the 10 folds

Evaluation metrics

- Regression:
  - R^2
  - Adjusted R^2
  - Mean Absolute Error (MAE)
  - Root Mean Absolute Error
- Classification
  - Accuracy
  - Precision
  - Recall
  - AUC
- Clustering
  - Silhouette Score
  - Inertia (for K-Means)

Overfitting and Underfitting

Overfitting:

- The model learns the noise in the training data, leading to poor generalization to unseen data.
- Symptoms: High accuracy on training data but low accuracy on test data.
- Solution:
  - Use regularization (L1, L2),
  - increase training data,
  - simplify the model, or use cross-validation.

Underfitting:

- The model is too simple to capture the underlying structure of the data.
- Symptoms: Poor performance on both training and test data.
- Solution:
- Increase model complexity,
- use more features, or
- reduce regularization.

Machine Learning in a nutshell

Tens of thousands of machine-learning algorithms	Several Application Areas	Broadly four types of learning	Every machine learning algorithm has four components:
• SVM • DT • KNN • NN • etc	• NLP • Computer Vision • Healthcare • etc	• Supervised • Unsupervised • Semi-Supervised • Reinforcement	• Representation (data, features, etc) • Method • Evaluation • Optimization

Recap

- Course Logistics
- Expectations
- Introduction to ML
- Types of Learning
  - Supervised Machine Learning
  - Unsupervised Machine Learning
  - Reinforcement Learning
- Evaluation
- Next Class: Optimization and Linear Regression
  - Formulation
  - Cost Function
  - Gradient Descent
- Python Example