Machine Learning for Beginners

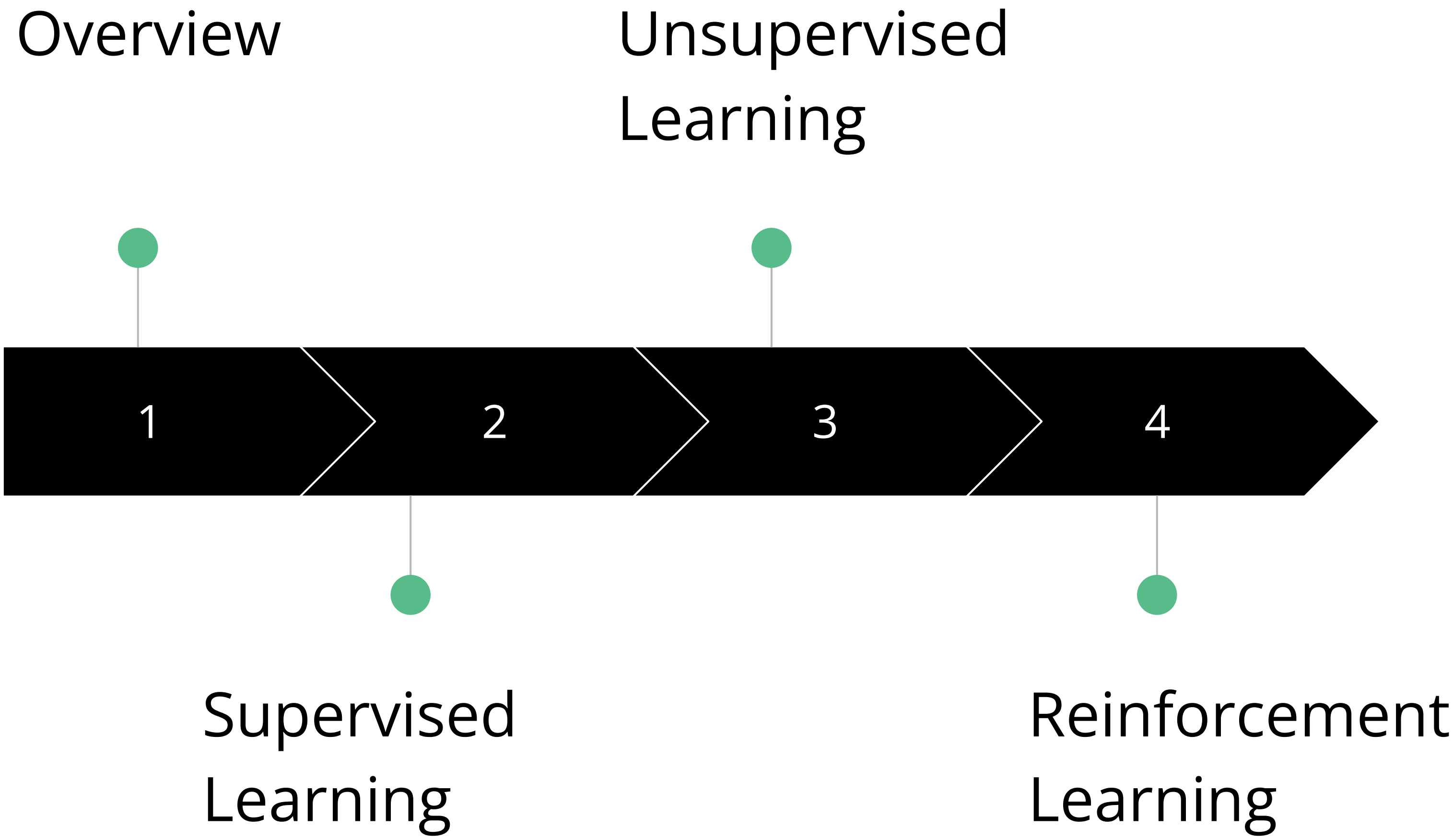
Victor Verma
October 29, 2024

Audience

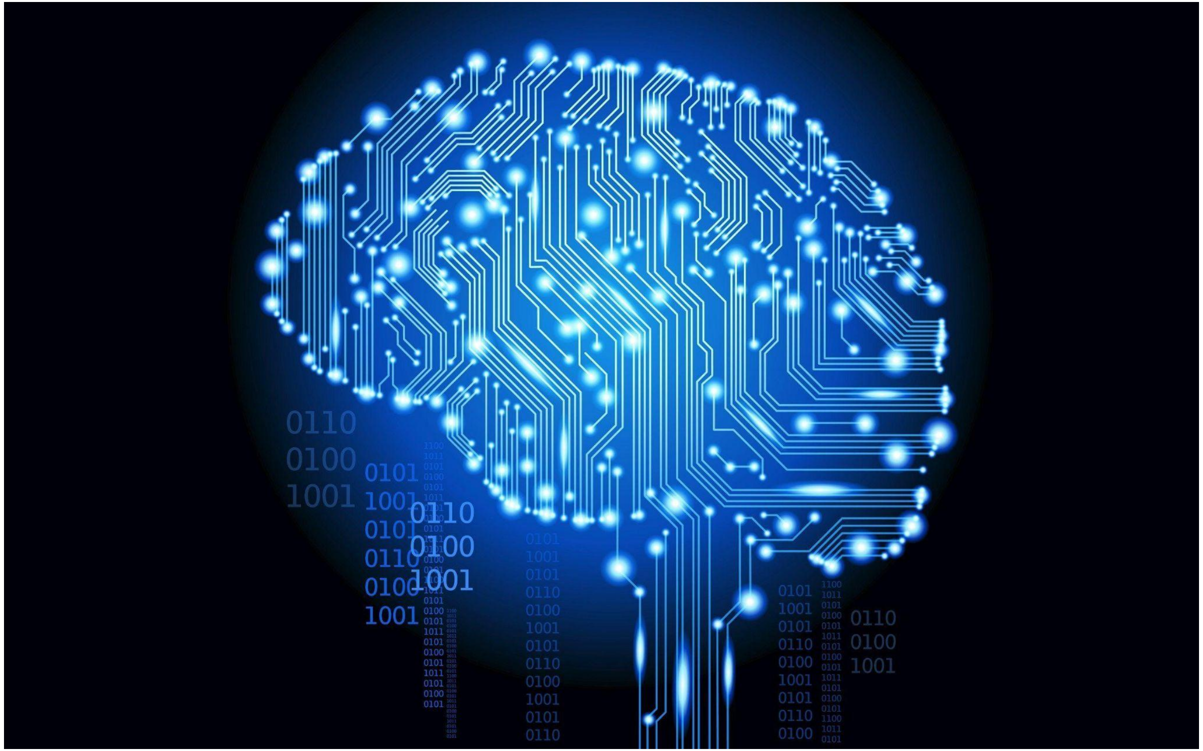
- People studying DS, CS, Math or CE with minimal exposure to machine learning.
- People in fields unrelated to machine learning, with a general interest.

NO MATH KNOWLEDGE REQUIRED!*

*Only for this presentation. Machine Learning is entirely math.


Overview
Unsupervised Learning
1
2
3
4
Supervised Learning
Reinforcement Learning

WHAT IS MACHINE LEARNING?



“A branch of computer science that focuses on using data and algorithms to enable AI to imitate the way that humans learn, gradually improving its accuracy” - IBM

The General Workflow

1. Identify a decision-making process for the model.
2. Define a loss function to evaluate the accuracy of the model's prediction.
3. Optimize the model using the loss function.

Train - Test - Validate

- Separate ~70% of the data to form a training dataset, ~15% of the data to form a testing dataset, and ~15% of the data to form a validation dataset.
- Allocating a validation dataset is extremely important and helps prevent overfitting.

Supervised Learning

(we are given labeled data)

Two Main Objectives

Classification

- Predicting discrete categorical variables.

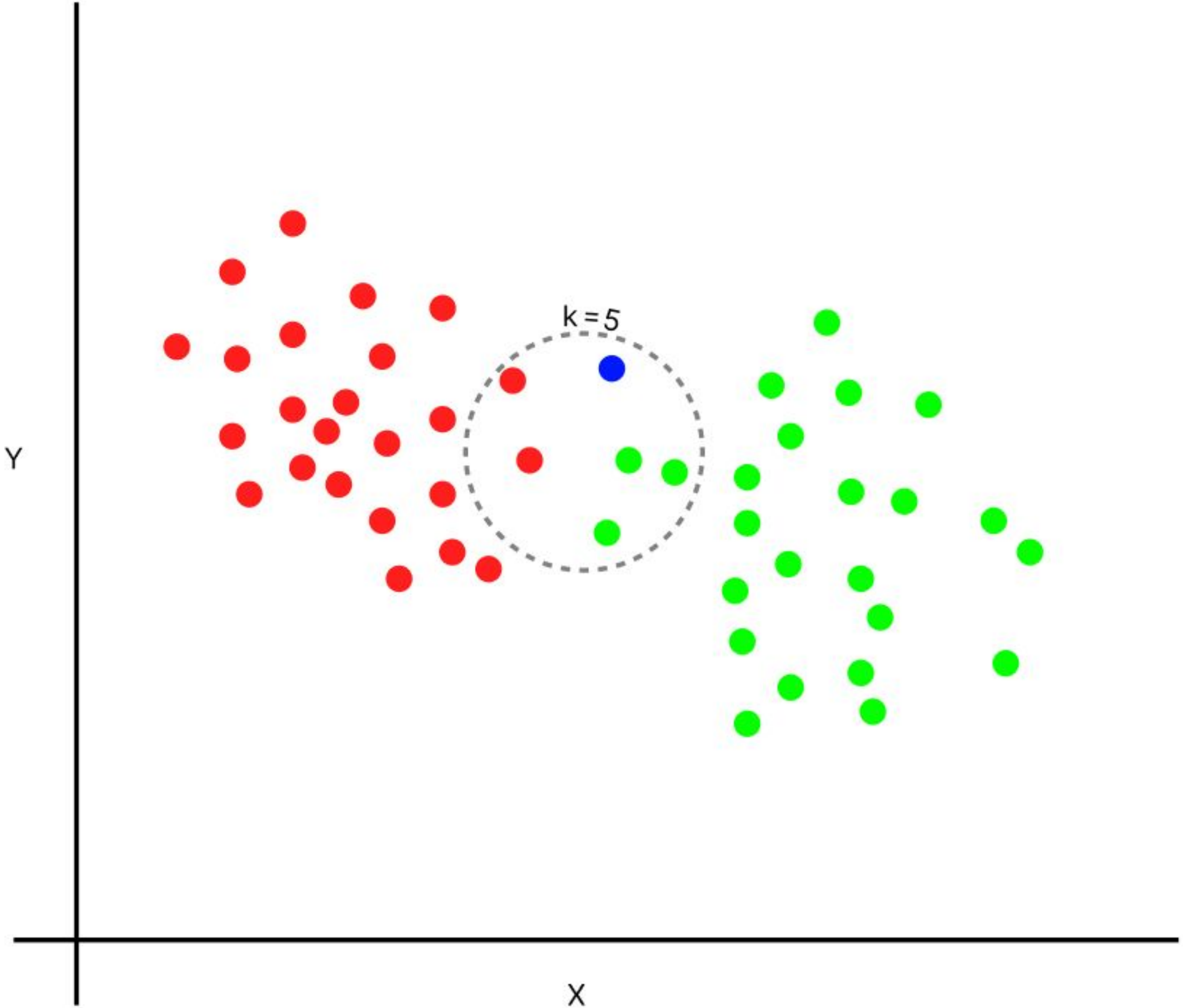
Regression

- Predicting continuous numerical variables.

Classification Methods

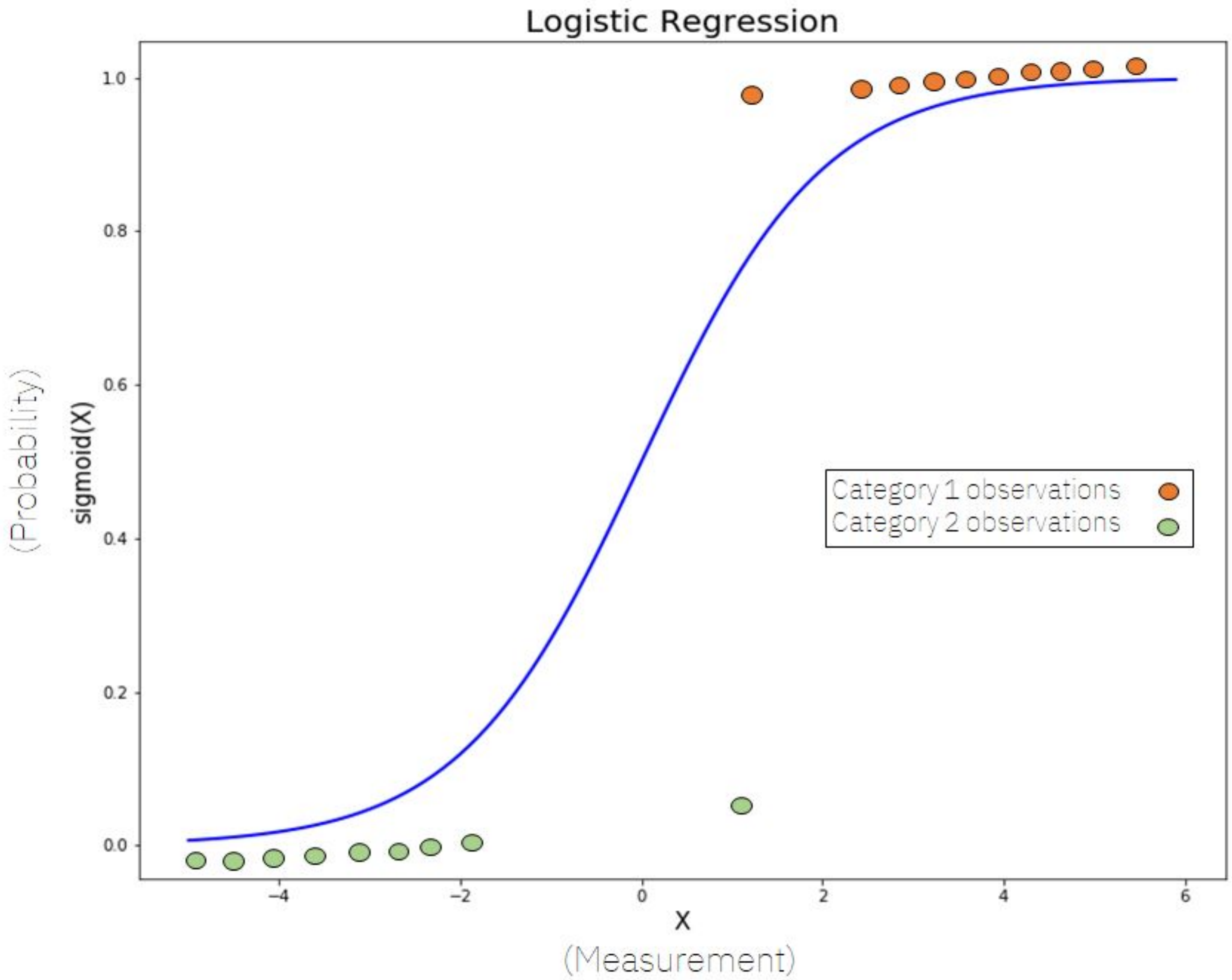
K-Nearest Neighbors

1. Pick an arbitrary unlabeled data point.
2. Find the k-closest pieces of labeled data to the chosen point, where closeness is determined by a defined distance function.
3. Assign the unlabeled piece of data to be the majority class across the k-closest pieces of labeled data.


k = 5
Y
X

Logistic Regression

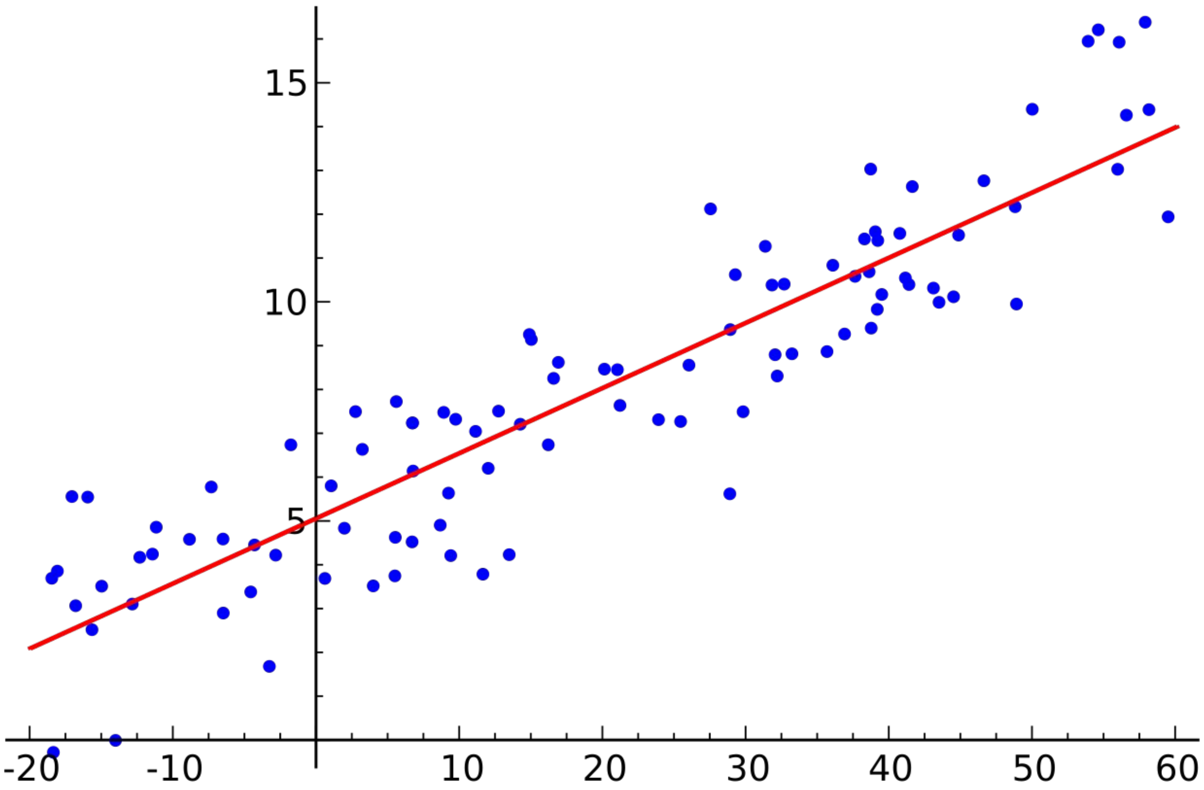
1. Assign weights to each feature and calculate the linear combination of the weights of the labeled data points.
2. Apply the sigmoid function to convert the sums into probabilities.
3. Compare the probabilities of the labels with the true labels.
4. Update the weights and repeat steps 1-3 as necessary.


Logistic Regression
1.0
0.8
0.6
0.4
0.2
0.0
(Probability)
sigmoid(X)
−4
−2
0
2
4
6
X
(Measurement)
Category 1 observations
Category 2 observations

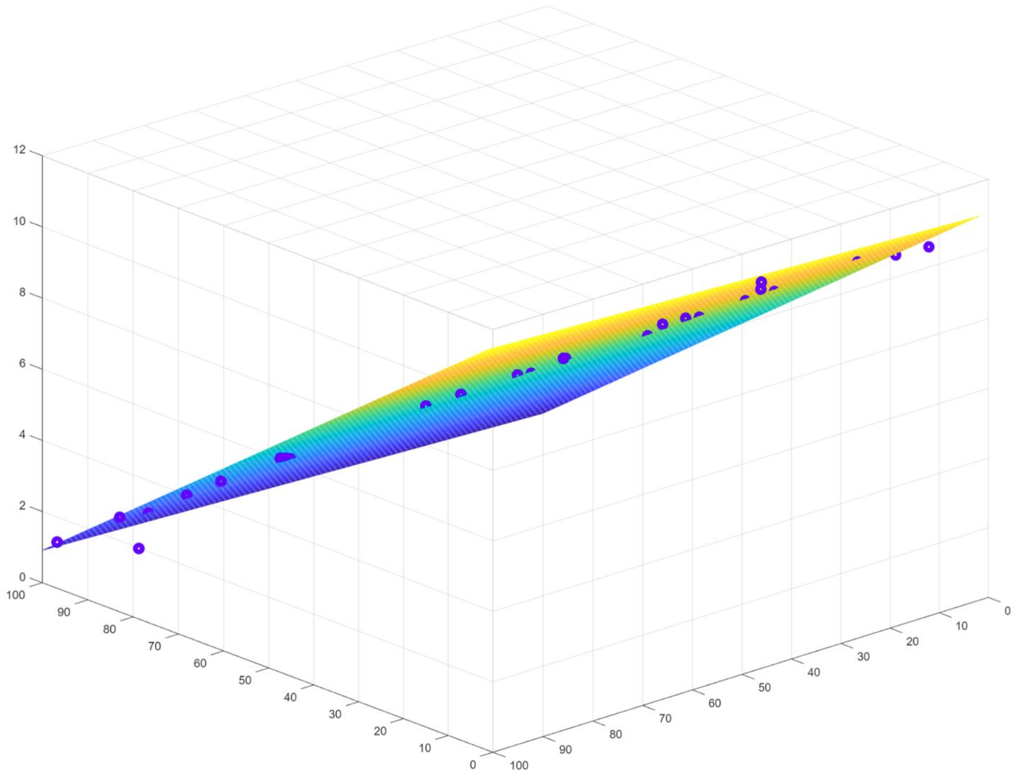
Regression Methods

Linear Regression

1. Assign weights to each feature and calculate the predicted target variable as a linear combination of the features.
2. Calculate the mean squared error and update the weights as necessary.



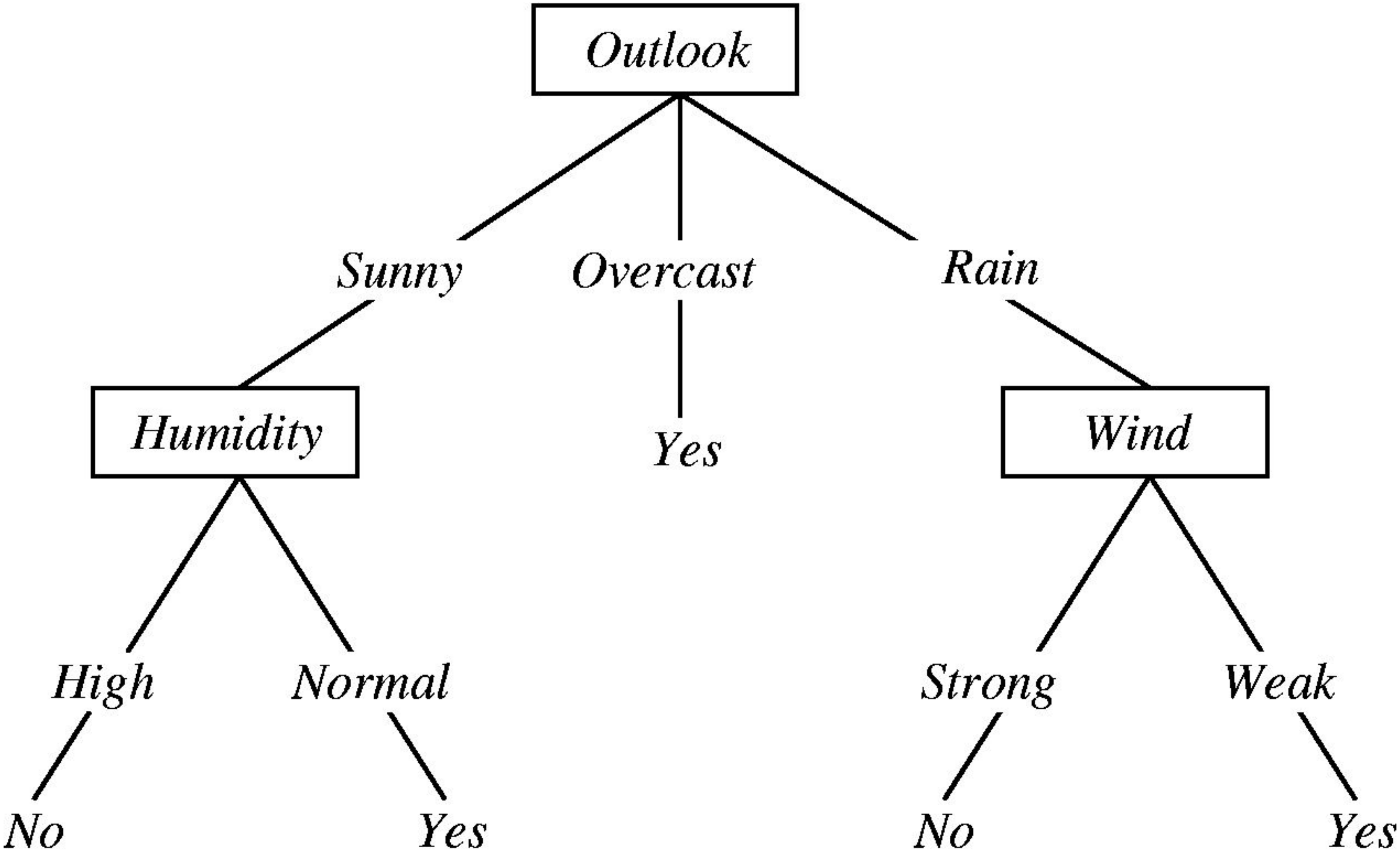
Two-Dimensional Case (1 feature + 1 target)



Three-Dimensional Case (2 features + 1 target)

Decision Trees

- Create a flow-chart like tree, where each node split is the result of a test on a subset of attributes.
- Decide the split based on
    - the likelihood of incorrect classification.
    - the amount of uncertainty.
    - the information gain from the split.


Outlook
Sunny
Overcast
Rain
Humidity
Yes
Wind
High
Normal
Strong
Weak
No
Yes
No
Yes

Bagging vs Boosting

Random Forest

- Creating decision trees in parallel and aggregating the results by voting or averaging.

XGBoost

- Creating decision trees iteratively, improving upon the previous, and using the final tree.

IMPORTANT

https://www.recommendations.victorverma.com/

Unsupervised Learning

(we are given unlabeled data)

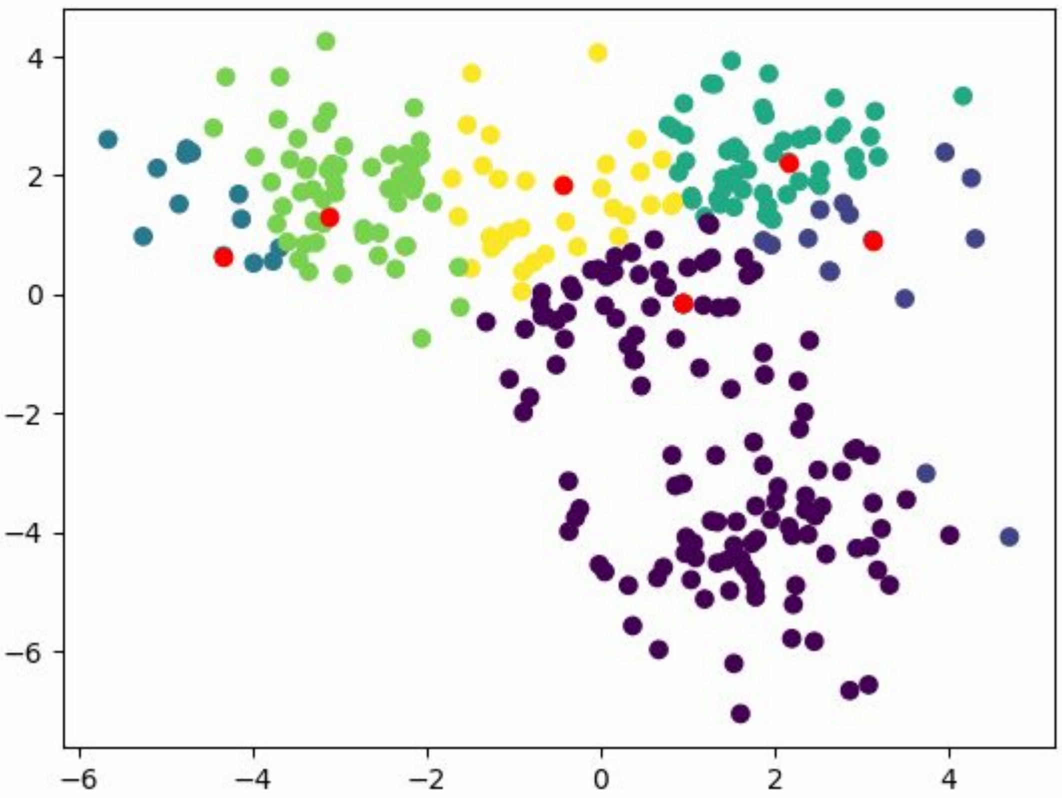
One Main Objective

- Exploring unlabeled data to find useful patterns.

Clustering

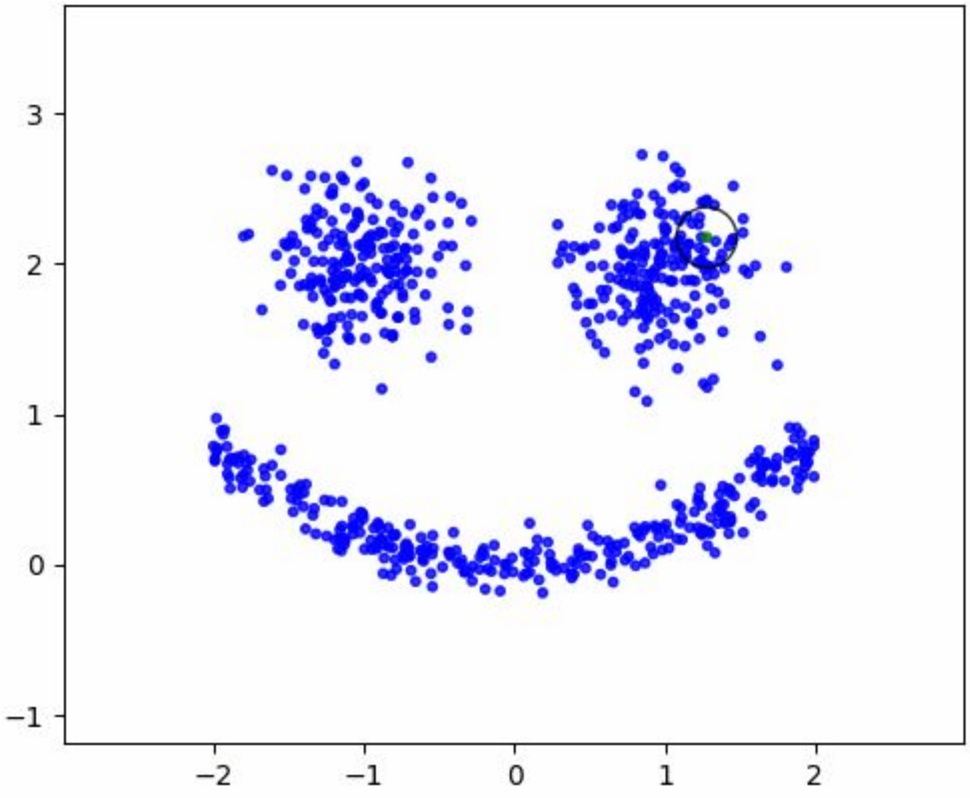
K-Means Clustering

1. Randomly choose k initial centroids.
2. Assign data points to their closest centroids.
3. Update the centroids with the average of the cluster assigned to that centroid.
4. Repeat steps 2-3 for a defined number of iterations.


4
2
0
−2
−4
−6
−6
−4
−2
0
2
4

Density-Based Clustering

- Core points have at least the minimum numbers of neighbors within radius epsilon.
- Border points are within radius epsilon of a core point, but do not have enough neighboring points.
- Assign core points in the same neighborhood to the same cluster, and border points to the nearest cluster.
- All unassigned points are considered noise.



3
2
1
0
−1
−2
−1
0
1
2

Other Unsupervised Learning Methods

Association Rules

- Identifying data points frequently appearing together.
- Finding conditional probabilities of data.

Dimensionality Reduction

- Reducing the number of features in a dataset, while still preserving important information from the dataset.

Association Rule Example

{Coffee, Sugar} → {Cream}

- Support: P(all 3 items) = 0.15.
- Confidence: P(cream|coffee, sugar) = 0.70.
- Lift: 3.5x more likely to buy cream given coffee and sugar.

REINFORCEMENT LEARNING

Key Components

Agent

- Takes actions resulting in rewards or penalties, to maximize the cumulative reward.

Environment

- Consists of the state space, action space, transition function, and a reward function.

Key Components

Actions

- Ways that the agent can move in from one state to the next.

Rewards

- Feedback telling the agent if the action was positive or negative.

Dog Training Example

- Agent → dog, environment → world, actions → dog behavior, rewards → treat.
- Over time, the dog will learn which actions result in the greatest rewards.

APPENDIX

Appendix: Key Terms

- Overview
    - Loss function, training dataset, testing dataset, validation dataset, overfitting.
- Supervised Learning
    - Distance function, majority class, weights, feature, sigmoid function, target variable, mean squared error, tree, node, split, bagging, boosting.
- Unsupervised Learning
    - Centroid, iterations, core points, border points, noise, support, confidence, lift.
- Reinforcement Learning
    - Agent, reward, penalty, cumulative reward, environment, state space, action space, transition function, reward function.

Appendix: Train - Test - Validate

- How should you split up the training and testing data?
- What happens if the train, test, and validation sets are not properly separated?
- How do you minimize bias in the training data?
- What is k-fold cross-validation?
- How do you handle imbalanced datasets?
- What is transfer learning?
- What is fine-tuning?

Appendix: Supervised Learning

- What are the commonly used distance functions?
- How do logistic and linear regression differ in the case of multiple target variables?
- What is the sigmoid function in the context of machine learning?
- What are the advantages and disadvantages of the different splitting criteria (gini impurity, entropy, and information gain) for decision trees?
- What is ensemble learning?
- What are the similarities and differences between bagging and boosting?

Appendix: Unsupervised Learning

- How do you choose the optimal number of clusters?
- What is K-means++, and what problem is it trying to solve?
- When does density-based clustering perform better than K-means?
- What are other useful clustering algorithms?
- What is principle component analysis (PCA) and singular value decomposition (SVD)?

Appendix: Reinforcement Learning

- What are policies and how do you balance exploration and exploitation?
- What is the value function?
- What are different types of transition functions?
- When is reinforcement learning useful?
- How do positive and negative reinforcement differ?
- What are the differences between adaptive dynamic programming, temporal difference learning, and q-learning?
- What are the advantages and disadvantages of reinforcement learning?

Mathematical Optimizations in Machine Learning