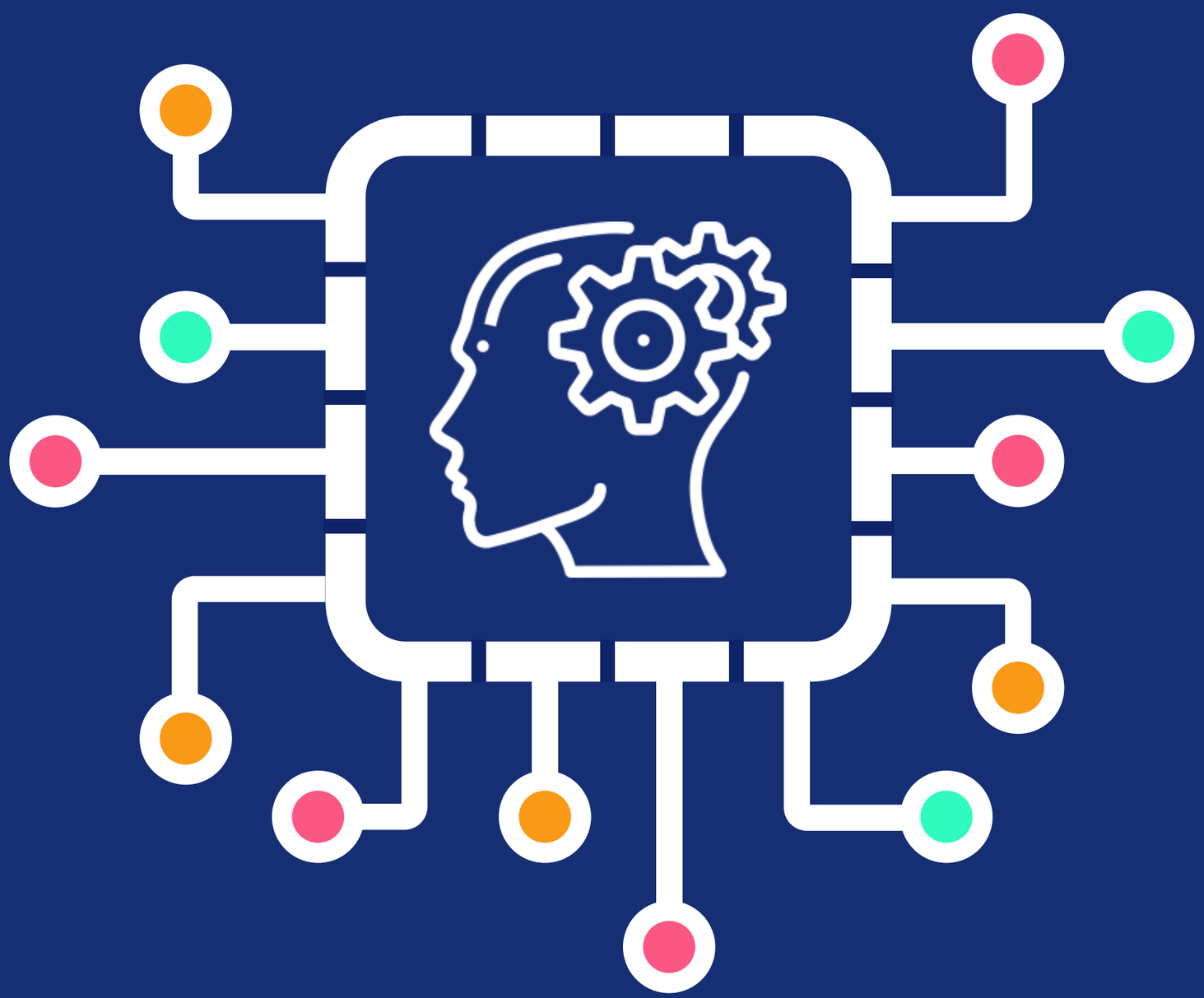
BOSSCODER ACADEMY

Top 10

Machine Learning

Algorithms



Complete Study Plan Inside

1. Linear Regression

Overview

Linear regression is a foundational algorithm in machine learning, used for predicting a continuous variable.

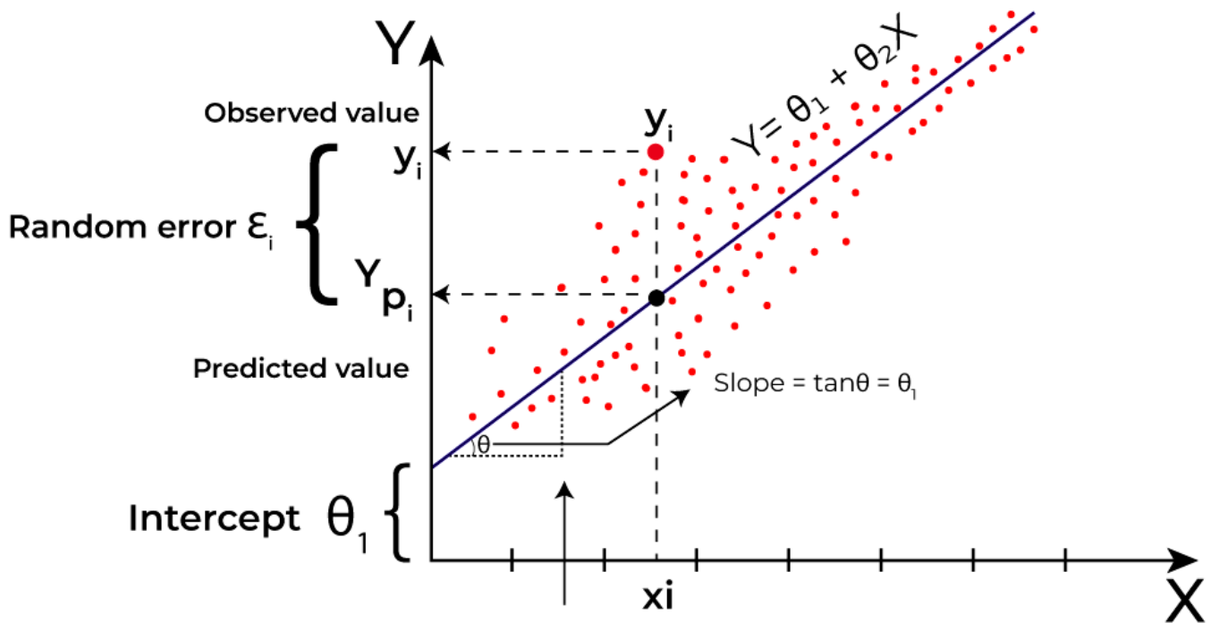
Learning Objectives

- Understand the theory behind linear regression.
- Learn to implement linear regression in Python.

Practice Questions

1. Implement linear regression to predict housing prices using a given dataset.

2. How would you evaluate the performance of your linear regression model?



2. Logistic Regression

📋 Overview

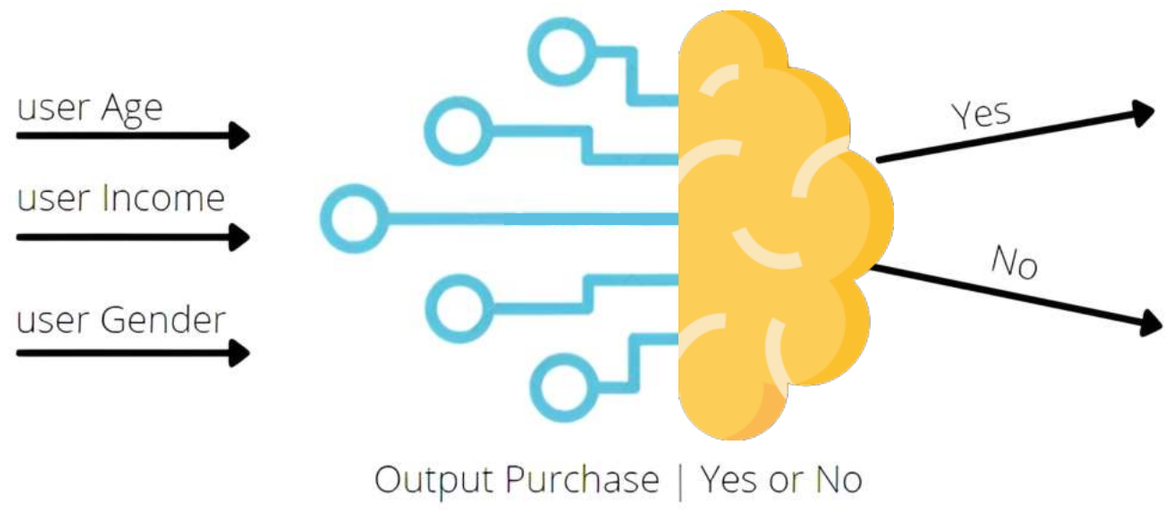
Logistic regression is used for binary classification problems, such as spam detection or predicting whether a customer will make a purchase.

📖 Learning Objectives

- Grasp the concept of logistic regression and its difference from linear regression.
- Practice implementing logistic regression on a binary classification problem.

🎯 Practice Questions

1. Use logistic regression to classify emails as spam or not spam.
2. Discuss how changing the threshold value affects the model's performance.



3. Decision Trees

📋 Overview

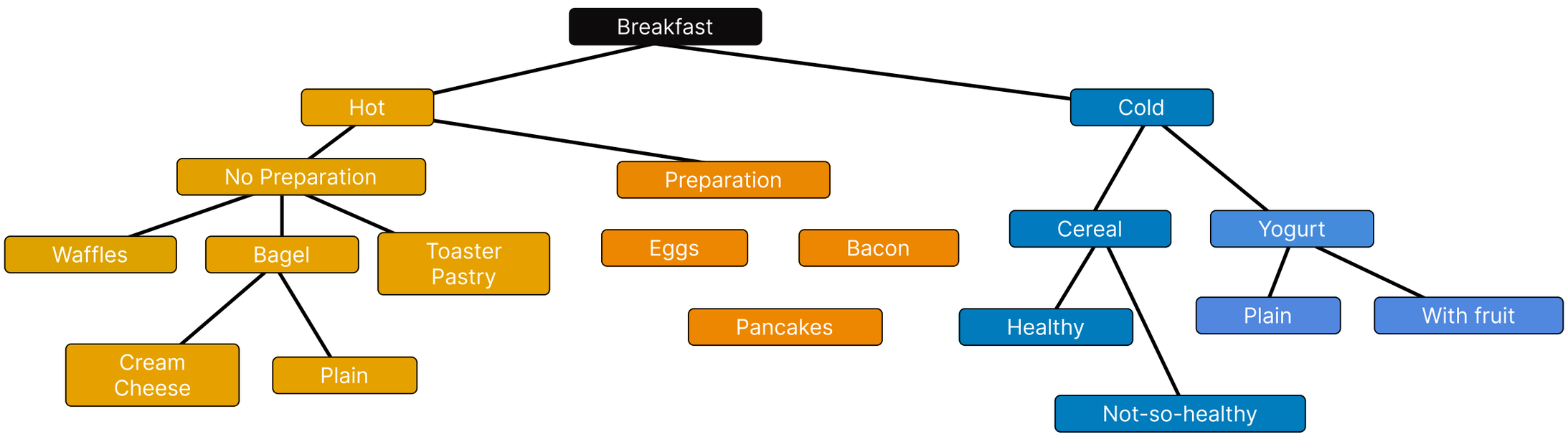
Decision trees are powerful for both classification and regression tasks. They are intuitive and easy to interpret.

📖 Learning Objectives

- Learn how decision trees are built, including the concepts of entropy and information gain.
- Implement a decision tree on a classification problem.

🎯 Practice Questions

1. Create a decision tree to predict customer churn.
2. How do decision trees handle overfitting?



4. Random Forest

Overview

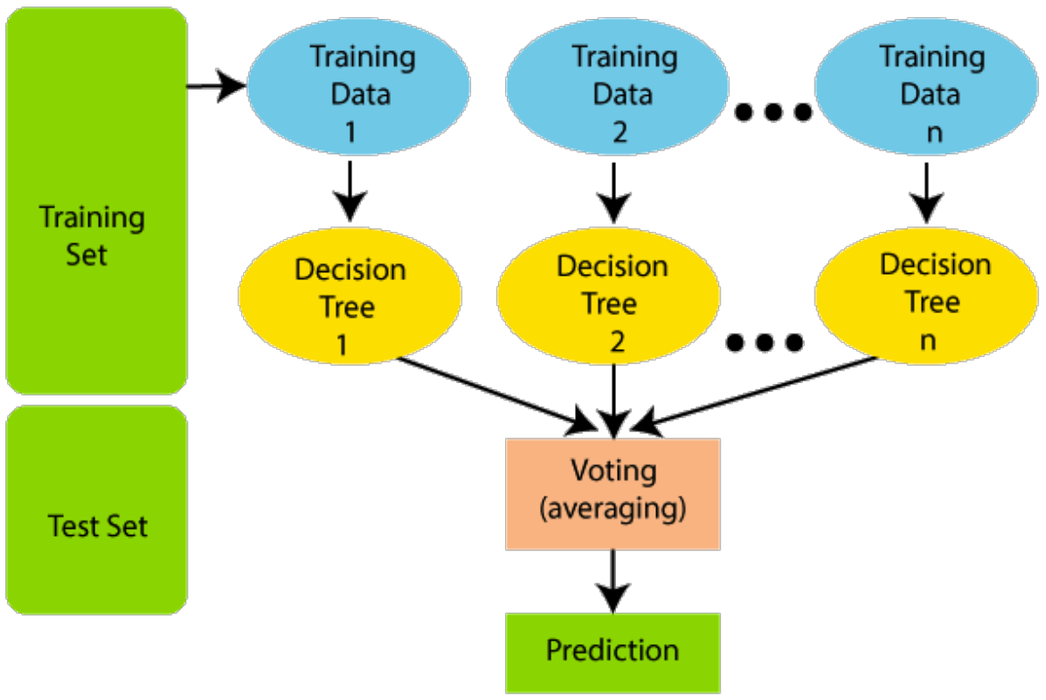
Random Forest is an ensemble learning method that operates by constructing multiple decision trees during training time.

Learning Objectives

- Understand the concept of ensemble learning and how random forests improve upon single decision trees.
- Implement a random forest model to solve a problem.

Practice Questions

1. Use a random forest to improve the model created on Day 3 for predicting customer churn.
2. Compare the performance of the decision tree and random forest models.



5. Support Vector Machines

Overview

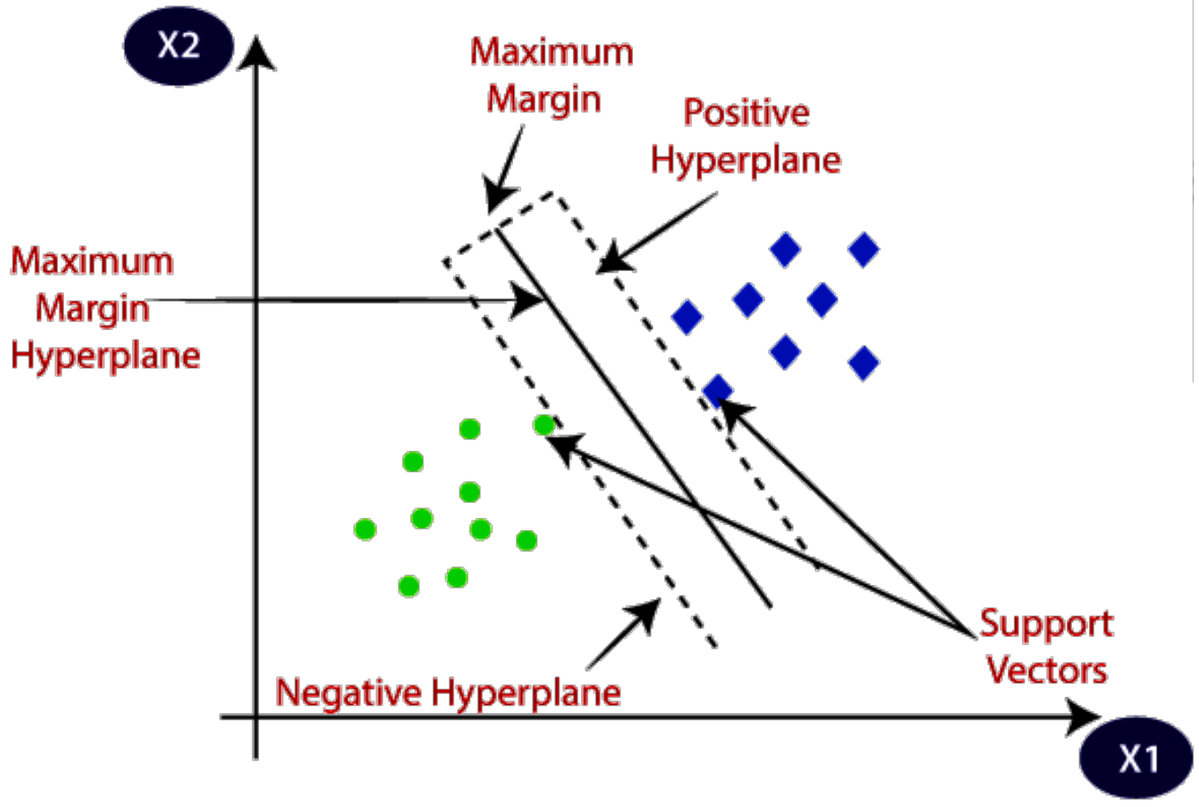
SVMs are powerful for high-dimensional spaces and are used for classification and regression tasks.

Learning Objectives

- Learn the theory behind SVMs, including the kernel trick.
- Practice implementing SVMs on a dataset.

Practice Questions

1. Implement an SVM to classify images of handwritten digits.
2. Experiment with different kernels and compare their effects on the model's performance.



6. K-Nearest Neighbors

📋 Overview

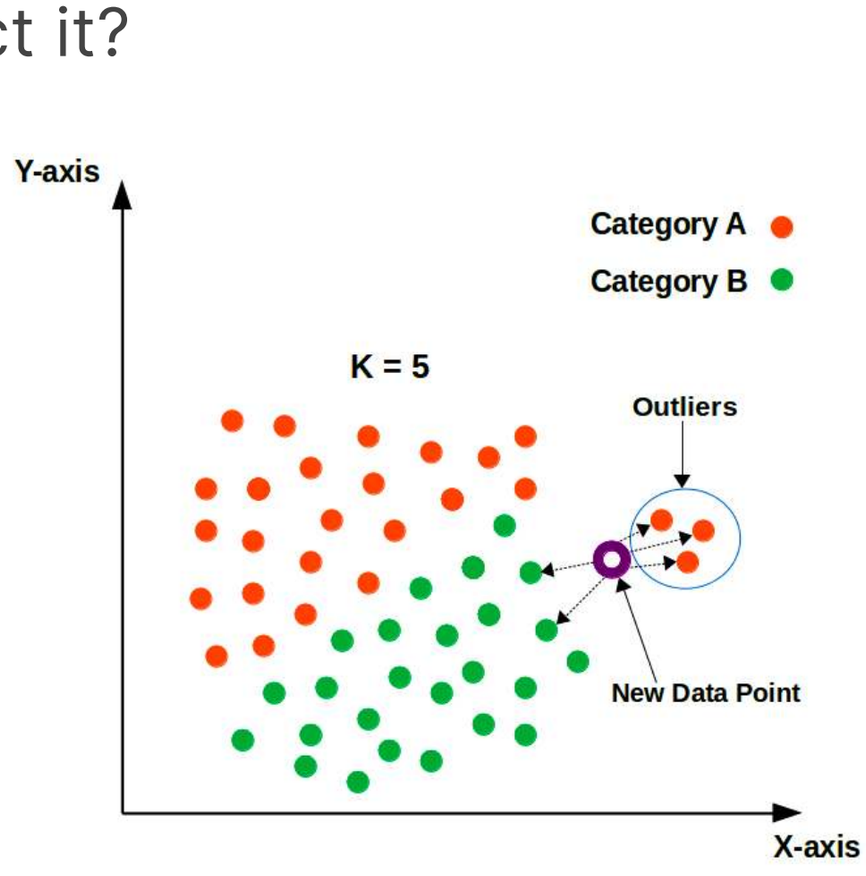
KNN is a simple, instance-based learning algorithm where the class of a sample is determined by the majority class among its k nearest neighbors.

📖 Learning Objectives

- Understand the KNN algorithm and its application.
- Implement KNN for a classification problem.

🧪 Practice Questions

1. Use KNN to classify patients as either having a disease or not based on their medical records.
2. How does the choice of 'k' affect the performance and how would you select it?



7. Naive Bayes

📋 Overview

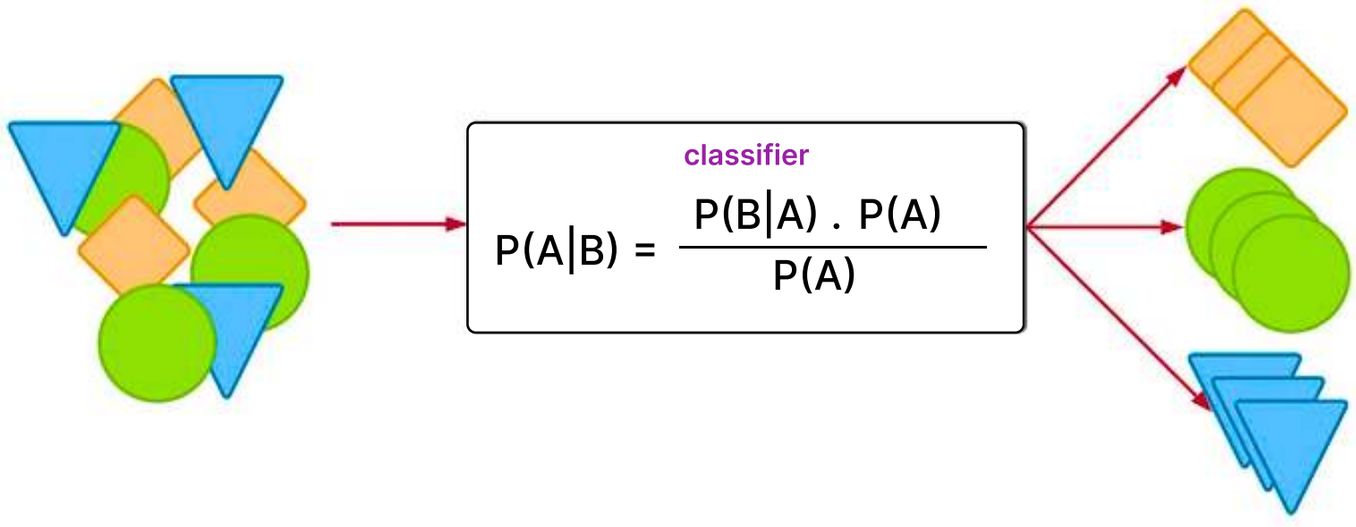
Naive Bayes classifiers are a family of simple probabilistic classifiers based on applying Bayes' theorem with strong independence assumptions between the features.

Learning Objectives

- Learn about the theory and assumptions behind Naive Bayes.
- Apply Naive Bayes to a text classification problem.

Practice Questions

1. Classify news articles into categories using Naive Bayes.
2. Discuss the assumption of feature independence in Naive Bayes. Is it always valid?



8. K-Means Clustering

Overview

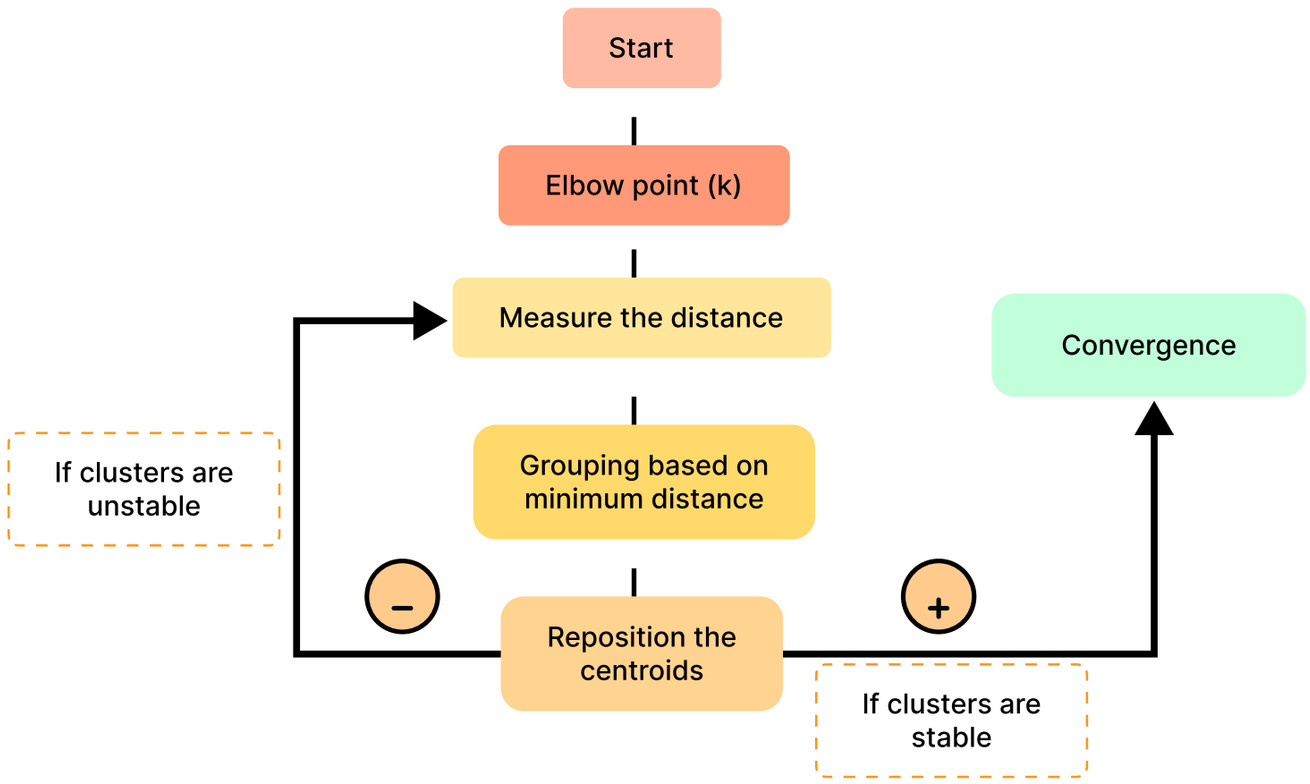
K-Means is a type of unsupervised learning algorithm used for clustering.

Learning Objectives

- Understand the concept and algorithm behind K-Means clustering.
- Implement K-Means to segment a dataset into clusters.

Practice Questions

1. Use K-Means to cluster customers based on their shopping habits.
2. How do you determine the optimal number of clusters?



9. Principal Component Analysis

Overview

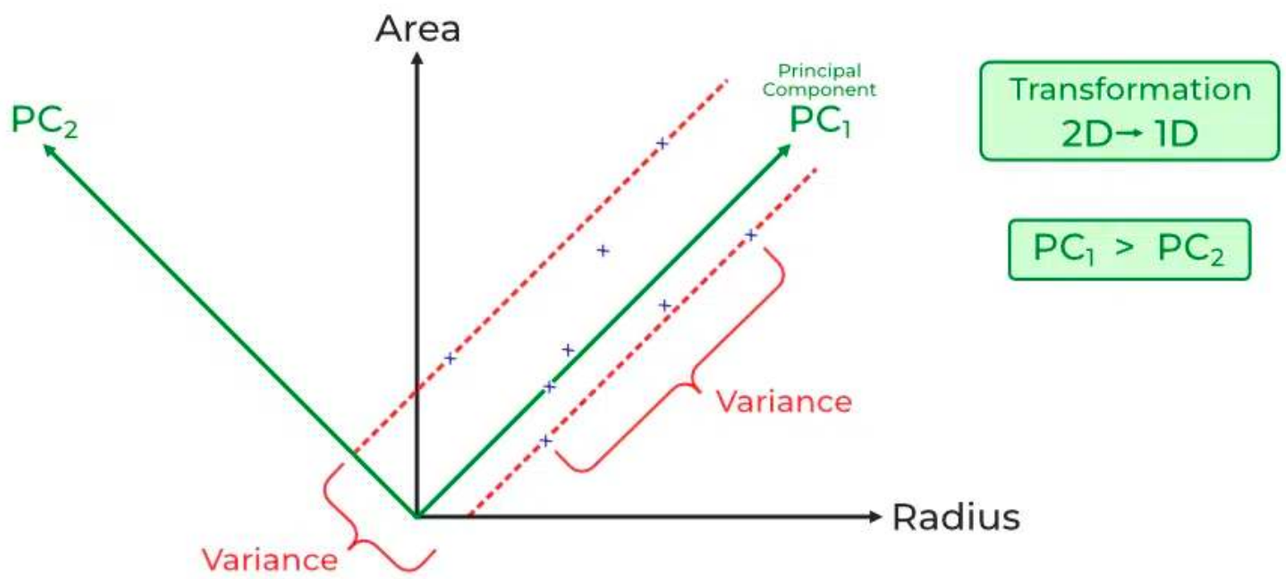
PCA is a dimensionality reduction technique used to reduce the dimensionality of large datasets.

Learning Objectives

- Grasp the concept of dimensionality reduction and the algorithm behind PCA.
- Apply PCA on a dataset and visualize the results.

Practice Questions

1. Implement PCA to reduce the dimensionality of a dataset containing images of faces.
2. How does PCA affect the performance of a classifier trained on the reduced dataset?



10. Gradient Boosting Machines

Overview

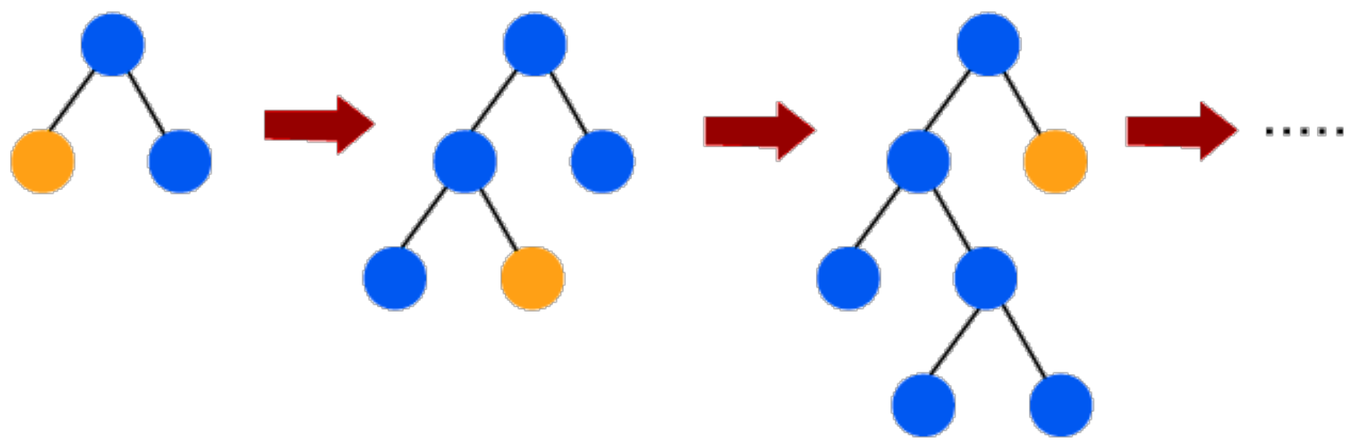
GBMs are a group of machine learning algorithms that use boosting techniques to improve prediction accuracy.

Learning Objectives

- Understand the principles of boosting and how GBMs build upon it to minimize loss.
- Implement a GBM model to tackle a complex regression or classification problem.

Practice Questions

1. Apply a GBM model to improve prediction accuracy on a dataset used in previous days. Compare its performance against models like decision trees and random forests.
2. Explore how changing parameters (such as learning rate, number of trees) affects the model's performance and overfitting.



ROADMAP TO FOLLOW

Begin by building your foundational knowledge:

So, make sure you focus on these first:

- **Mathematics and Statistics:**

Dedicate few weeks to have good understanding of Maths & Stats concepts necessary for understanding ML algorithms, such as linear algebra, calculus, probability, and statistics.

- **Programming Skills:**

Once this is done, focus on become proficient in a Programming language commonly used in data science, like Python, etc.

Week-by-week Plan

Once foundational knowledge is in place, get into the ML algorithms. Here's a rough estimate:

- **Weeks 1-3:**

Begin with simpler algorithms that have broad applications, such as linear regression, logistic regression, and decision trees.

This phase includes understanding the theory, mathematical underpinnings, practical implementations, and getting comfortable with relevant data science libraries (e.g., Scikit-learn).

- **Weeks 4-6:**

Move on to more complex algorithms, including ensemble methods like random forests and gradient boosting machines, as well as support vector machines.

Focus on grasping the concepts, learning how to tune their parameters, and exploring their applications.

- **Weeks 7-9:**

Dive into algorithms that require a good understanding of similarity measures and optimization, such as K-nearest neighbors (KNN), and clustering algorithms like K-means.

Also, start exploring dimensionality reduction techniques such as PCA.

- **Weeks 10-12:**

Wrap up with algorithms and models that involve probabilistic models and more complex decision-making, like Naive Bayes and neural networks (if included in your top 10 list).

Begin to explore specialized topics within machine learning that interest you, such as deep learning, reinforcement learning, or specific applications like natural language processing or computer vision.

Continuous Learning

Beyond 12 Weeks: Machine learning is a vast and rapidly evolving field.

Continuous learning through projects, online courses, attending workshops, reading research papers, and participating in competitions like Kaggle can help deepen and update your knowledge.

Practice and Projects

Throughout this period, it's crucial to apply what you've learned through hands-on projects. Implementing the algorithms on different datasets, participating in competitions, or working on personal projects can significantly reinforce your understanding and skills.

BOSSCODER ACADEMY

WHY BOSSCODER?

- **750+** Alumni placed at Top Product-based companies.
- More than **136% hike** for every **2 out of 3** working professional.
- Average package of **24LPA.**

> The syllabus is most **up-to-date** and the list of problems provided covers **all important topics.**
>
> **Lavanya**
> Meta

> Course is very well **structured** and **streamlined** to crack any **MAANG** company
>
> **Rahul .**
> Google

EXPLORE MORE