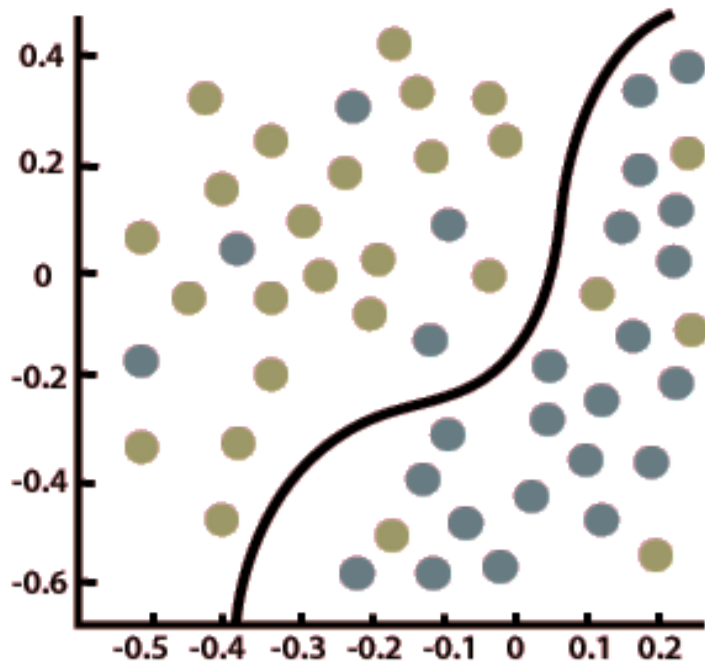


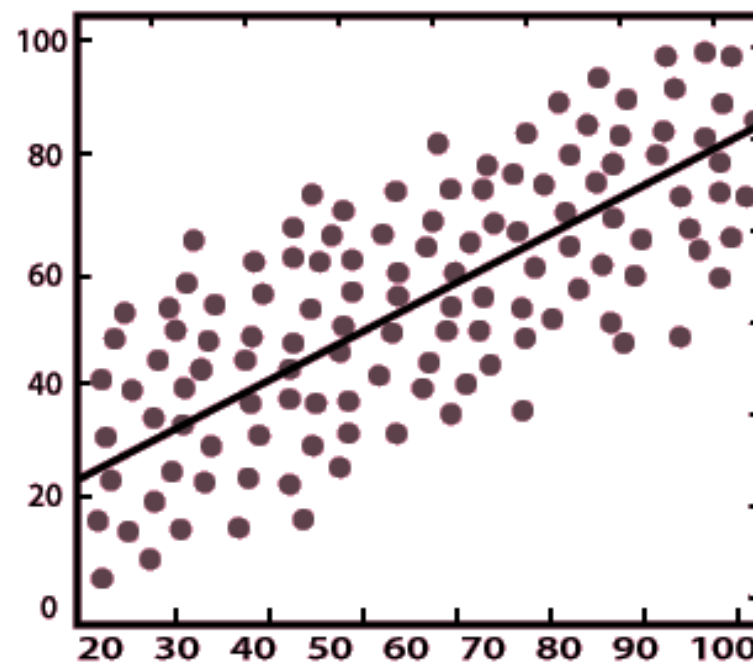
Classification Models

- **Classification vs. Regression**
- **Binary vs. multi classification**
- **The basic algorithms**

Regression predicts continuous outcomes, while **Classification** focuses on predicting discrete categories, like determining whether an email is spam or not.



Classification



Regression

Aspect	Regression	Classification
Type of Outcome	Continuous values (e.g., real numbers)	Discrete categories (e.g., labels)
Goal	Predict numerical values	Predict categories or labels
Visualization	Scatter plots with regression lines	Decision boundaries dividing data into categories

Classification vs. Regression

מספר האנשים שיגיעו לאירוע על סמך מזג האוויר.

לקבוע אם לקוח יעזוב את השירות של חברה או יישאר.

לחזות את כמות החשמל שצורכת עיר בשעה מסוימת.

לזהות איזה גזע של כלב בתמונה

לחזות את הטמפרטורה של עיר בשעה מסוימת ביום.

לחזות את כמות הגשם שתירד במהלך יום נתון.

לחזות האם מחר יהיה יום גשום

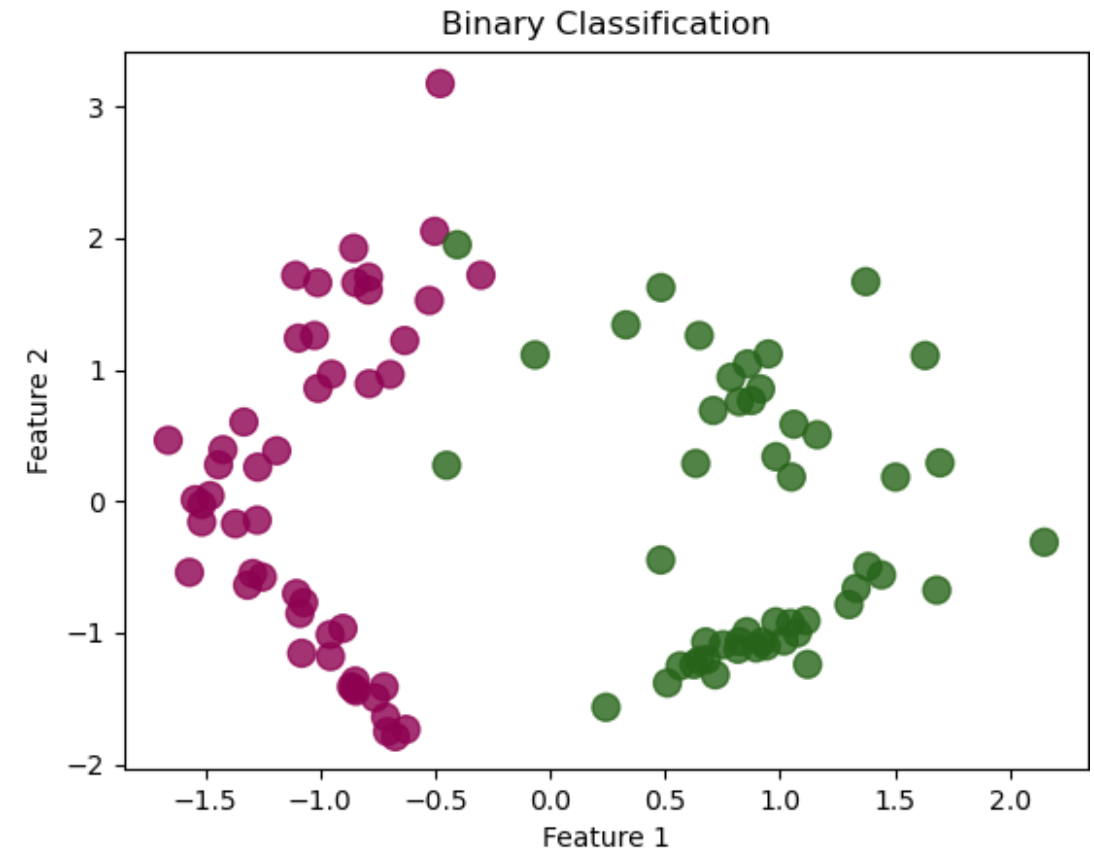
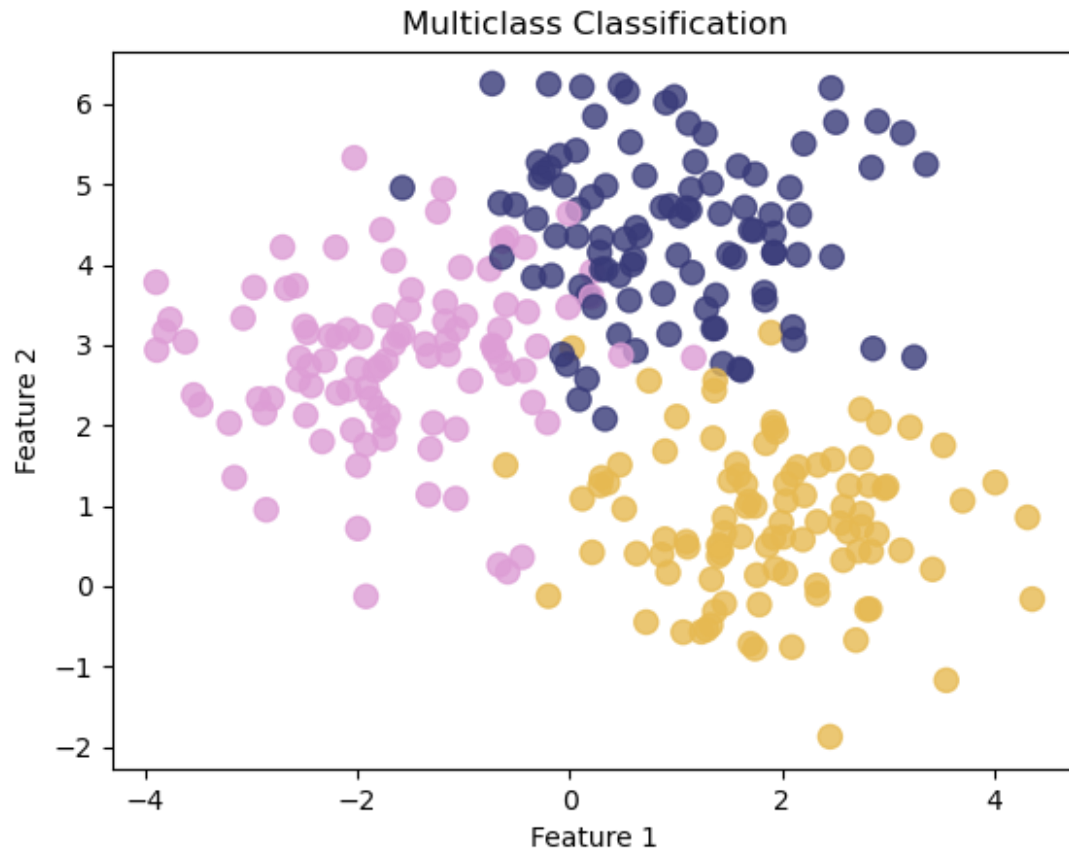
לקבוע אם סטודנט יעבור או ייכשל בקורס על סמך הציונים שלו עד כה.

לקבוע ציון של סטודנט יקבל במבחן הבא על סמך הציונים שלו עד כה

התרגיל הזה הוא ...

Classification or Regression

Binary classification predicts two outcomes, while multi-class predicts three or more. This distinction matters because multi-class tasks often require model adaptations, more complex metrics, and additional computational resources



Binary vs Multi - classification

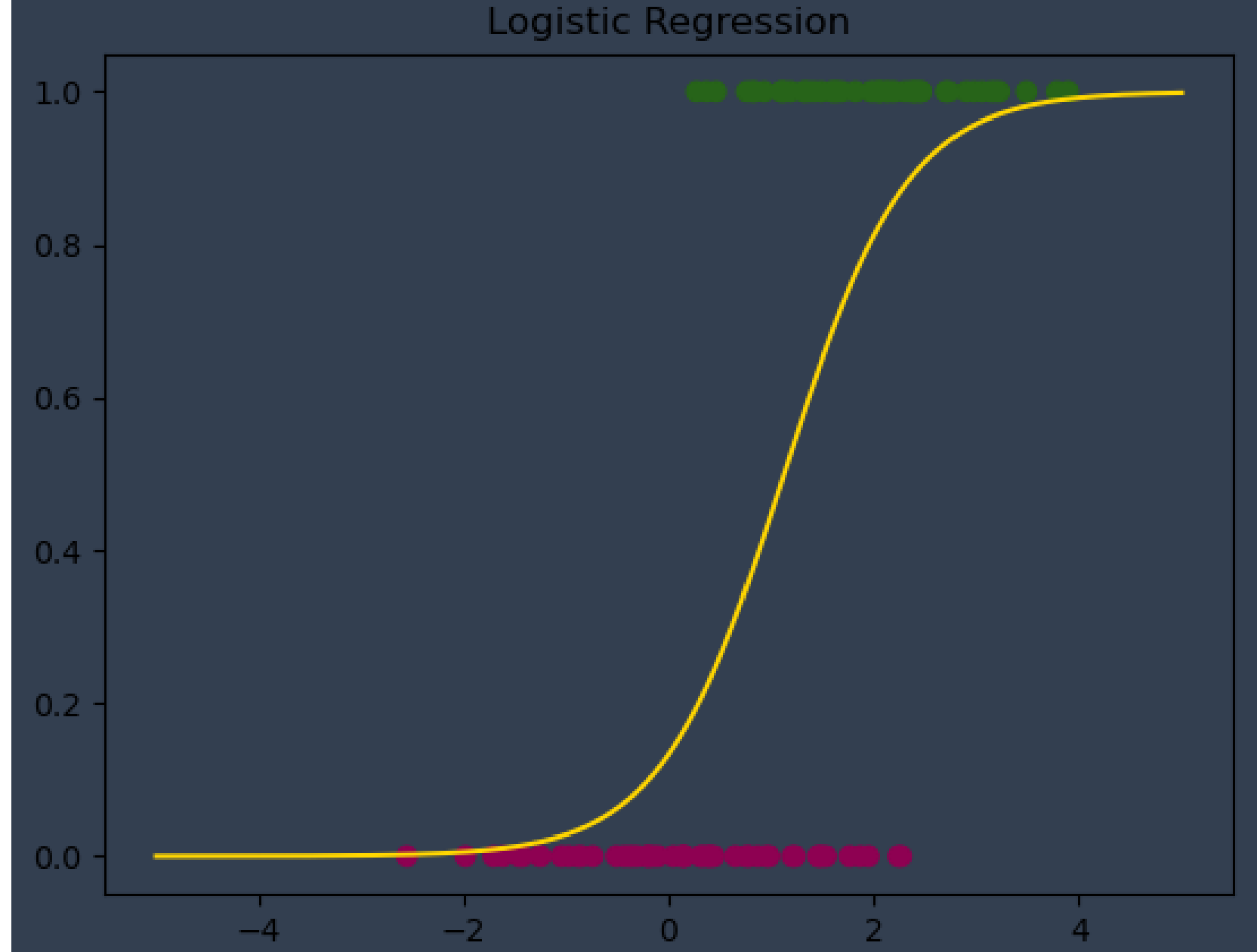
- ✓ Classification vs. Regression
- ✓ Binary vs. multi classification

→ The basic algorithms

1. Logistic Regression
2. Decision Trees
3. Random Forests
4. Support Vector Machines (SVM)
5. KNN
6. Perceptrons

Logistic Regression is a statistical model used primarily for binary classification problems.

the sigmoid function outputs a value between 0 and 1 making it ideal for transforming to probability. This probability value is then used to classify the data into one of the two classes. For example, in a binary classification problem, we could say that if the probability is below 0.5, the class is 0, and if it's above 0.5, the class is 1.

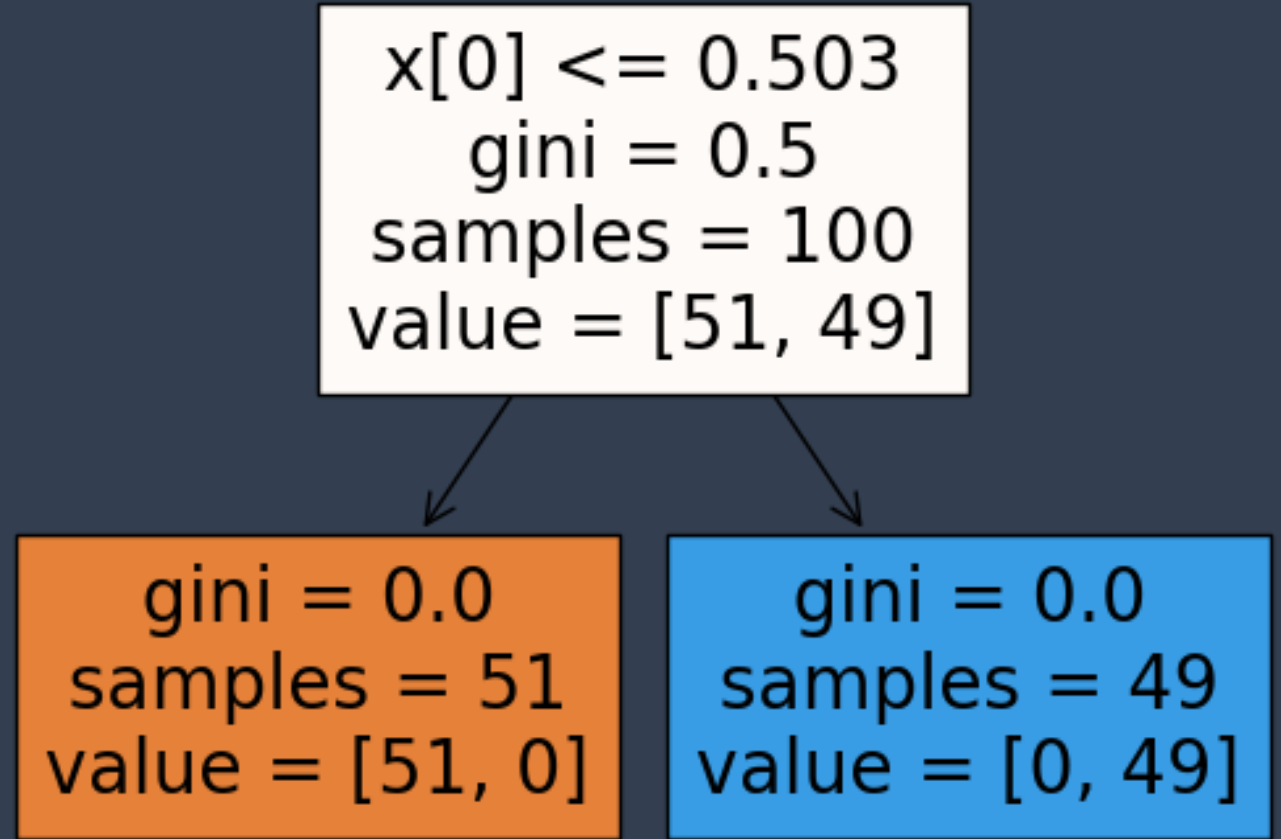


Decision Trees are **flow-chart models** where each node represents a feature-based decision.

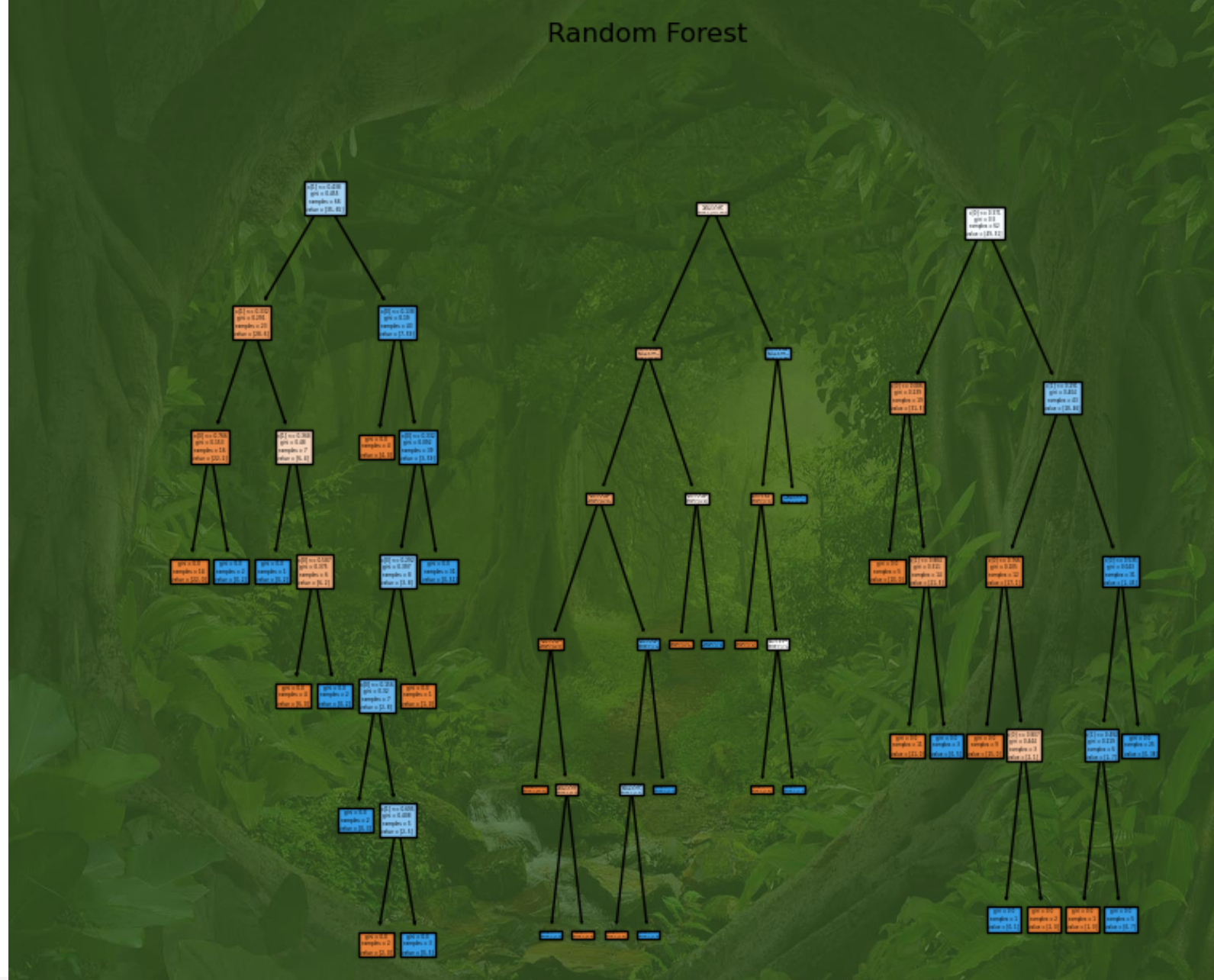
They handle binary and multi-class classification well but are **sensitive to overfitting and bias**.

Gini measures impurity, helping the model choose optimal splits.

Decision Tree

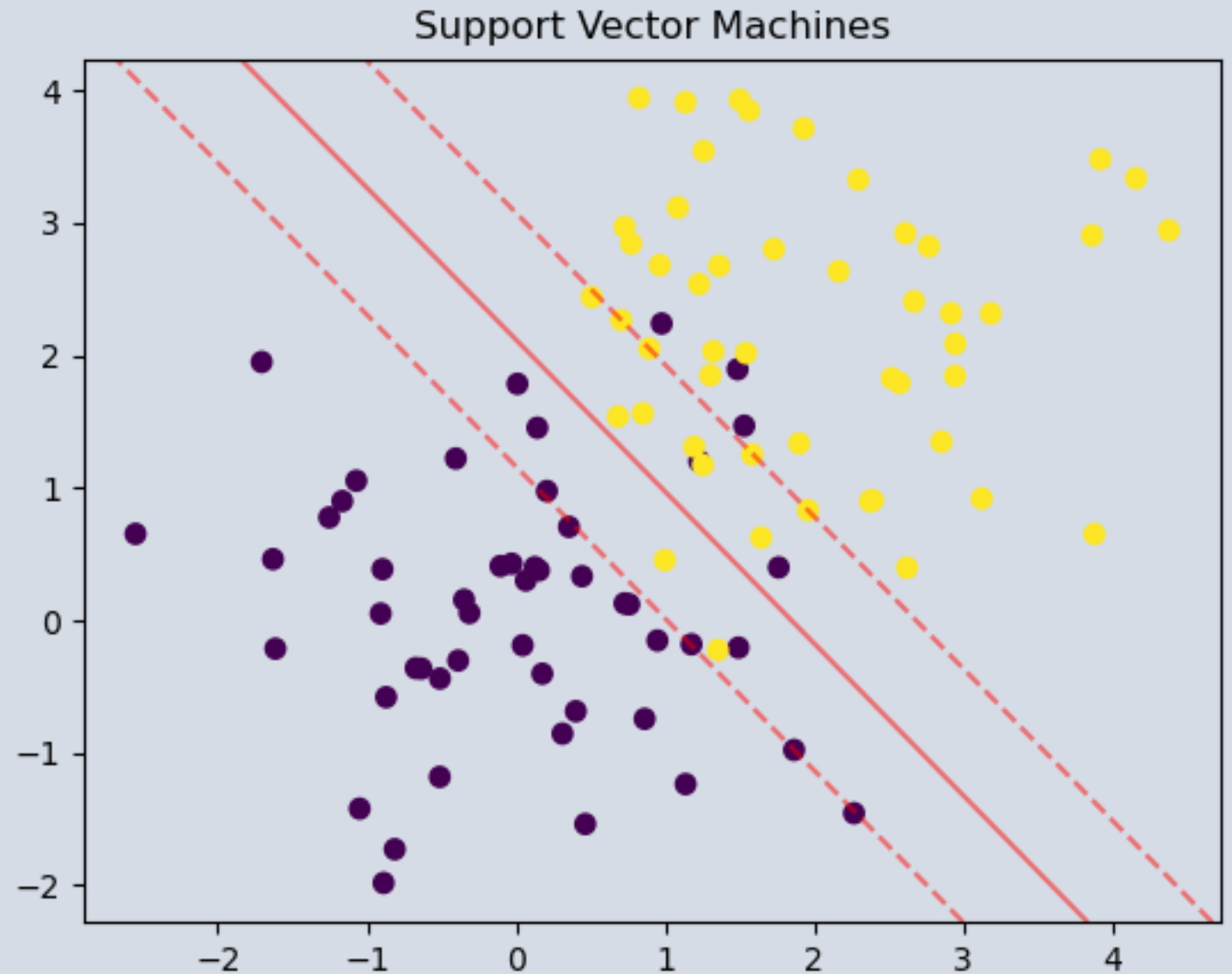


Random Forests are **ensembles of multiple decision trees** that work together to **improve accuracy** and **reduce overfitting**. Each tree is trained on a **random subset of data and features**, making the model more robust to noise and bias. The final prediction is based on the majority vote (classification) or the average (regression) of all trees.



Support Vector Machines
(SVM) are powerful
algorithms used for
classification and regression.

They work by finding the
optimal hyperplane that
separates classes with the
maximum margin, making
them **robust to noise**

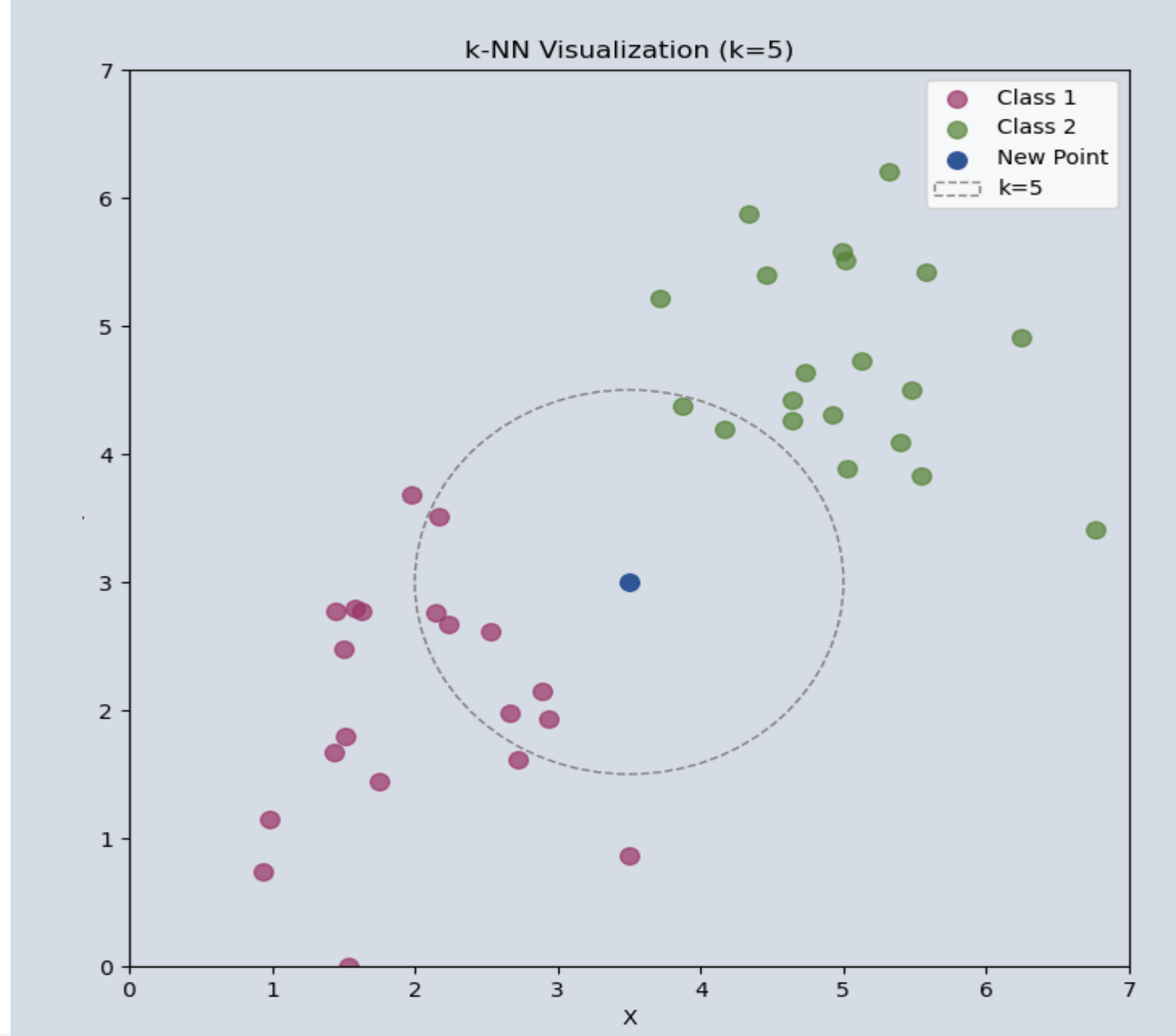


Support Vector Machines

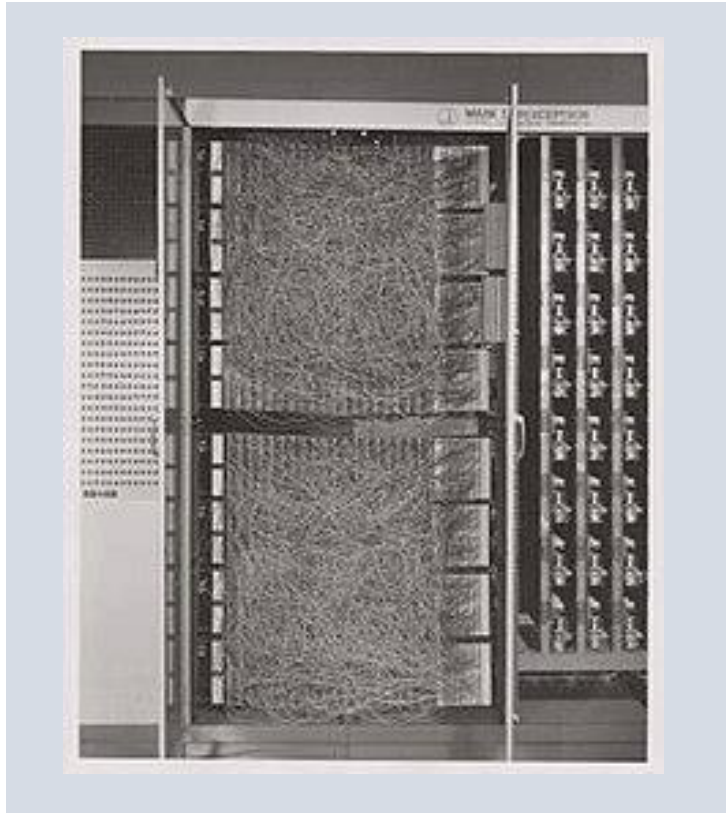
k-NN is a simple algorithm that classifies points based on the majority class of their k nearest neighbors.

Advantages: Easy to use, no assumptions about data.

Disadvantages: Slow on large datasets, sensitive to k and distance choice.



KNN



Perceptrons are simple neural networks for binary classification of linearly separable data. They compute a weighted sum of inputs, apply a step function, and output 000 or 111. While easy to understand, they cannot handle non-linear problems.

preceptron