Machine Learning Concepts Summary

1) K-Nearest Neighbors (KNN) Algorithm

Definition:

KNN is a supervised machine learning algorithm used for classification or regression.

It predicts the class of a new data point by looking at the K nearest data points (neighbors)

from the training dataset and choosing the majority class among them.

Steps:

1. Choose the value of K (number of neighbors).

2. Calculate the distance between the new data point and all training data points.

3. Select the K nearest points.

4. Perform majority voting - the most common class among K neighbors is the prediction.

Formula (Euclidean Distance):

Distance = sqrt((x2 - x1)^2 + (y2 - y1)^2)

---

2) Binary Classification

Definition:

Binary classification means classifying data into two categories (classes).

Example:

0 -> Customer will stay

1 -> Customer will leave

Examples:

- Spam (1) or Not Spam (0)

- Disease (1) or No Disease (0)

- Churn (1) or Retain (0)

---

3) Support Vector Machine (SVM)

Definition:

SVM is a supervised machine learning algorithm used for classification and regression.

It finds the best boundary (hyperplane) that separates data points of different classes

with the maximum margin.

Formula for Decision Boundary:

$w . x + b = 0$

Where:

w = weight vector

x = input features

b = bias

---

4) TP, TN, FP, FN - Definitions and Example

TP (True Positive): Model correctly predicts positive (Predicted leave, actually left)

TN (True Negative): Model correctly predicts negative (Predicted stay, actually stayed)

FP (False Positive): Model wrongly predicts positive (Predicted leave, actually stayed)

FN (False Negative): Model wrongly predicts negative (Predicted stay, actually left)

---

5) Confusion Matrix & Evaluation Metrics

Confusion Matrix:

	Predicted: Yes	Predicted: No
Actual: Yes	TP	FN
Actual: No	FP	TN

Formulas:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Error Rate = (FP + FN) / (TP + TN + FP + FN)

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

F1 Score = 2 x (Precision x Recall) / (Precision + Recall)

Support = Number of actual occurrences of each class

Summary Table:

Metric	Formula	Meaning
Accuracy	(TP + TN) / (TP + TN + FP + FN)	Overall correctness
Error Rate	(FP + FN) / (TP + TN + FP + FN)	Overall wrong predictions
Precision	TP / (TP + FP)	How many predicted positives were right
Recall	TP / (TP + FN)	How many actual positives were found
F1 Score	2 x (Precision x Recall) / (Precision + Recall)	Balance between precision and recall
Support	Count of actual samples per class	Shows data distribution