**Question 1**

1. **Differentiate between supervised learning and unsupervised learning. Provide example of each. (3 marks).**

Supervised learning uses labeled datasets to train algorithms to classify data or predict outcomes. As input data or predict outcomes. As input data is inputted into the model, its weight modify until it fits into the model; this process is known as cross validation which ensures the model is not over fitted or under fitted.

Example;

Classification: Email spam detection (classify emails as “spam” or “not spam”).

Unsupervised learning analyses and clusters unlabeled datasets using machine learning methods. The algorithms find hidden patterns or data groupings without human interaction. This method is useful for exploratory data analysis, cross-selling, consumer segmentation, and image and pattern recognition.

Examples:

Clustering: Customer segmentation (grouping customers by purchasing behavior).

1. **Explain the concept of overfitting in machine learning models and explain how it can be mitigated. (3 marks).**

Overfitting occurs when a machine learning model is too complex and fits the training data too complex and fits the training data too closely. This can lead to poor performance on new, unseen data because the model is too specialized to the training dataset. To prevent overfitting, it is important to use a validation dataset to evaluate the model’s performance and to use regularization techniques to simplify model.

**Reference**

TutorialsPoint. *Machine learning tutorial.* TutorialsPoint. Retrieved from https://www.tutorialspoint.com/machine\_learning

**Applications:**

* Predicting house prices based on features like area, number of rooms, etc.
* Forecasting stock prices or sales.

**ii. K-Nearest Neighbors** (2 marks)

K-Nearest Neighbors (KNN) is a supervised learning algorithm used for both classification and regression. It is a non**-**parametric method, meaning it makes no assumptions about the underlying data distribution.

**Working Principle:**

1. **Data Representation**:
   * Each data point in the training set is plotted in feature space.
   * The algorithm stores all training data and uses it during prediction.
2. **Classification**:
   * For a new data point, the algorithm identifies the kkk closest data points (neighbors) using a distance metric such as:
     + **Euclidean Distance**: d(p,q)=∑i=1n(pi−qi)2d(p, q) = \sqrt{\sum\_{i=1}^{n} (p\_i - q\_i)^2}d(p,q)=i=1∑n​(pi​−qi​)2​
     + Manhattan or Minkowski distances can also be used.
   * It assigns the class label that is most frequent among these kkk neighbors (majority voting).
3. **Regression**:
   * For regression tasks, KNN predicts the output as the average of the outputs of the kkk nearest neighbors.
4. **Key Parameters**:
   * **Value of kkk**: Determines the number of neighbors considered.
   * **Distance Metric**: Determines how neighbors are measured.
5. **Training Process**:
   * No explicit training phase; all computations are deferred to prediction (lazy learning).
6. **Prediction**:
   * The algorithm searches for neighbors and applies the majority vote (classification) or averaging (regression).

**Question 3: (6 marks)**

a) Discuss the importance of feature scaling in machine learning and compare two common methods used: Min-Max Scaling and Standardization. (3 marks)

**Feature Scaling:** Feature scaling is a crucial preprocessing step in machine learning. It ensures that the features (input variables) in a dataset have the same scale or range, which is essential for many machine learning algorithms. Without scaling, algorithms that use distance metrics (e.g., K-Nearest Neighbors, Support Vector Machines) or gradient-based optimization (e.g., neural networks, linear regression) may not perform well because some features with larger numeric ranges can dominate others, leading to biased results or poor convergence. Scaling helps in speeding up the convergence of algorithms and leads to better model performance.

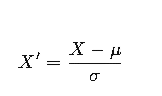
**Why Feature Scaling is Important:**

1. **Improved Model Performance:** Machine learning models, especially those that rely on calculating distances (like KNN or SVM) or gradients (like logistic regression or neural networks), can be sensitive to differences in feature scales. Features with larger ranges can overshadow smaller ones, leading to skewed results.
2. **Faster Convergence:** Many optimization algorithms (such as gradient descent) converge faster when features are scaled, reducing the number of iterations needed to find an optimal solution.
3. **Prevents Numerical Instability:** Models that use mathematical computations may face numerical instability when features are on very different scales, leading to inaccurate results.
4. **Ensures Equal Weight:** Feature scaling ensures that all features contribute equally to the model, preventing features with larger magnitudes from having an outsized influence.

**Common Methods of Feature Scaling:** Two widely used methods of feature scaling are **Min-Max Scaling** and **Standardization**. Both methods are designed to transform the data into a specific range or distribution, but they differ in their approach.

1. **Min-Max Scaling (Normalization):**

* **Explanation:** This method scales each feature to a specific range, often [0, 1]. It works by subtracting the minimum value of the feature and dividing by the range (maximum - minimum) of the feature. The result is that all values are transformed to fit within the desired range.
* **When to Use:** Min-Max scaling is effective when the data is known to follow a specific range, especially when the features are bound within a known minimum and maximum. It is sensitive to outliers, as the scaling is based on the minimum and maximum values of the data.
* **Advantage:**
* Simple and effective.
* Retains the original distribution of the data.
* **Disadvantage:**
* Sensitive to outliers, which can skew the scaling process.
* Not suitable for datasets with a wide range of values.

1. **Standardization (Z-score Normalization):**

**Formula:**

Where μ is the mean of the feature and σ is the standard deviation.

Standardization transforms the data into a distribution with a mean of 0 and a standard deviation of 1. This ensures that the data has no units, and it scales the data based on how far the individual data points are from the mean.

Standardization is ideal when the dataset follows a normal distribution or when there are no specific bounds on the features. It is less sensitive to outliers compared to Min-Max scaling because it uses mean and standard deviation rather than the range.

Advantages

* Less sensitive to outliers.
* Useful when data has an unknown or varying range.

**Disadvantages:**

* Does not guarantee that the scaled data will lie within a specific range (e.g., [0, 1]).
* May not work well with highly skewed data.

**Comparison:**

* **Outliers:** Min-Max scaling is highly affected by outliers because the scale depends on the minimum and maximum values. In contrast, standardization is more robust to outliers since it relies on mean and standard deviation.
* **Range of Scaled Data:** Min-Max scaling forces the data to lie within a specific range, while standardization does not impose a fixed range on the data.
* **Distribution Assumptions:** Standardization works best when the data follows a normal distribution, whereas Min-Max scaling does not make any assumptions about the data distribution.

### b) Explain the role of training, validation, and test datasets in machine learning projects. (3 marks)

* **Training Dataset:** This is where the model learns and adapts, making it essential for building a functioning model. If too small, the model may not learn adequately; if too large, it might overfit, especially if not complemented with validation and test sets.
* **Validation Dataset:** Used for tuning the model's hyperparameters and avoiding overfitting. It allows for model improvement without using the test data, which ensures that the evaluation remains objective.
* **Test Dataset:** The test dataset is crucial for final model evaluation. It ensures that the model's performance is not overly optimistic and provides an accurate measure of its predictive ability in real-world scenarios

**Question 4: (10 marks)**