**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