1. Explain Supervised vs. Unsupervised Learning

**Supervised Learning**:

* 1. You have input features XXX and labels YYY.
  2. Choose a model (e.g. decision tree, neural network, linear regression, SVM).
  3. Define a loss function that measures the error between predicted label and actual label (e.g. Mean Squared Error, Cross-Entropy).
  4. Optimize the model by minimizing the loss over the training data (e.g. via gradient descent).
  5. Evaluate on unseen (test) data.

**Unsupervised Learning**:

* 1. You have input data XXX only.
  2. Choose a model or method (e.g. clustering, dimensionality reduction, density estimation).
  3. The model tries to find structure: e.g. group similar data points together, reduce dimensions to capture major variation, find probability distributions.
  4. Evaluate via internal metrics (e.g. clustering coherence, reconstruction error), or sometimes by how useful the learned structure is for downstream tasks.

**Examples**

| **Supervised Learning** | **Unsupervised Learning** |
| --- | --- |
| Classifying emails as “spam” or “not spam” (labels provided) | Clustering customers into segments based on purchase behavior (no pre-defined segments) |
| Predicting house prices given features like area, number of rooms, etc | Dimensionality reduction of image data (e.g. using Principal Component Analysis) to compress images or visualize them |
| Recognizing handwritten digits (labels 0-9) | Detecting anomalies/outliers in data (e.g. fraud detection) |

**Strengths and Weaknesses**

| **Aspect** | **Supervised Learning** | **Unsupervised Learning** |
| --- | --- | --- |
| **Strengths** | • Accuracy can be very high if you have good labels and lots of data.  • Clear performance metrics (accuracy, precision, recall, etc.).  • Predictive power for specific tasks. | • Can discover unknown patterns or structure in data you didn’t anticipate.  • Doesn’t require costly labelling. |