Machine Learning Questions – Answers & Justifications

1. Inputs & Outputs

1.1. A Self-Driving Car

Input: Real-time camera images, LIDAR sensor data, GPS coordinates.
Output: Steering angle, acceleration or braking commands.
Justification: The car needs to 'see' the environment (input) to decide how to control its movement (output) safely.

1.2. Netflix Recommendation System

Input: User viewing history, ratings, watch time, and genre preferences.
Output: A ranked list of recommended movies/TV shows.
Justification: Recommendations are based on past behaviour (input) and should produce personalised content suggestions (output).

1.3. Signature Recognition

Input: Image or digital trace of a signature.
Output: Verification result: genuine or forged.
Justification: The system compares the input signature against stored examples and outputs a classification.

1.4. Medical Diagnosis

Input: Patient data (symptoms, lab results, imaging scans).
Output: Diagnosis (e.g., “Positive for diabetes” or “Negative for diabetes”).
Justification: Diagnosis maps input features to a predicted condition, helping doctors make decisions.

2. Regression vs. Classification

2.1. Classifying Emails

Classification – The output is a category (e.g., promotional, social), not a continuous value.

2.2. Forecasting Stock Price

Regression – The output is a continuous numeric value (predicted price).

2.3. Sorting Animal Images

Classification – Each image belongs to a discrete category (dog, cat, bird, etc.).

2.4. Predicting Disease Likelihood

Classification (for yes/no disease) OR Regression (for probability/risk score).
Most medical use cases use classification (positive/negative), but regression can be used to predict a risk score between 0–1.

3. Supervised vs. Unsupervised Learning

3.1. Detecting Anomalies

Unsupervised – No labelled data is available; the system must find anomalies by itself.

3.2. Predicting Customer Lifetime Value

Supervised – We have historical data with known outcomes to train on.

3.3. Segmenting Customers

Unsupervised – No predefined labels exist; the algorithm should group customers into natural clusters.

3.4. Categorising Social Media Posts

Supervised (if labelled data exists) OR Unsupervised (if discovering themes without labels).

4. Semi-Supervised Learning

4.1. Predicting Fraudulent Transactions

Inappropriate – Most transactions are labelled, so supervised learning is better.

4.2. Customer Satisfaction Surveys

Appropriate – Few labels exist; semi-supervised learning can leverage both labelled and unlabelled data.

4.3. Identifying Spam Emails

Inappropriate – With most data labelled, supervised classification is preferred.

4.4. Predicting Credit Card Default

Inappropriate – Fully labelled data is available; supervised learning is most suitable.