Question 1: Explain the bias–variance tradeoff in machine learning. Why is it important for model generalization?

Answer 1:

The **bias–variance tradeoff** explains how a model's learning capacity affects its performance.
A model with **high bias** is too simple and underfits it cannot capture the underlying patterns in the data, leading to consistently high errors on both training and test sets.
A model with **high variance** is too complex and overfits it memorizes the training data, performs well on it, but fails to generalize to new, unseen data. The goal is to find a balance where the model is expressive enough to learn important patterns but not so flexible that it captures noise.
Achieving this tradeoff leads to a model that generalizes well and performs reliably on new inputs, such as unseen images or text samples.

Question 2: What is a Support Vector Machine (SVM), and how does it classify data using hyperplanes and kernel functions?

Answer 2:

A **Support Vector Machine (SVM)** is a supervised learning algorithm used primarily for classification but also applicable to regression tasks. SVM works by finding an optimal **decision boundary (hyperplane)** that maximally separates data points of different classes. The key idea is to maximize the **margin**, which is the distance between the hyperplane and the closest data points, known as **support vectors**. A larger margin typically leads to better generalization. SVM can also handle non-linearly separable data through **kernel functions** such as the polynomial kernel or radial basis function (RBF), which map the data into a higher-dimensional feature space where separation becomes possible. This makes SVM robust, effective in high-dimensional spaces, and particularly strong in problems with limited training samples.

Question 3: What is bagging in ensemble learning? Describe how it works and why it improves model performance.

Answer 3:
Bagging (Bootstrap Aggregating) is an ensemble learning technique that improves model stability and accuracy by reducing variance. It works by generating multiple bootstrap samples from the training dataset each sample is created by randomly selecting data points *with replacement*.
A separate model (usually a high-variance model like a decision tree) is trained on each subset.
During prediction, the outputs of all models are combined, typically through majority voting for classification or averaging for regression. This aggregation smoothens out fluctuations caused by individual models and significantly reduces overfitting.