Q1. The mathematical formula for a linear Support Vector Machine (SVM) can be represented as:
$f(x)$=sign($w \cdot x+b$)$f(x)$=sign($w \cdot x+b$)
Here, $f(x)f(x)$ represents the predicted class label for the input vector xx, ww is the weight vector, bb is the bias term, and signsign is the sign function.

Q2. The objective function of a linear SVM is to maximize the margin between the decision boundary (hyperplane) and the closest data points (support vectors), subject to the constraint that the data points are correctly classified. Mathematically, the objective function is formulated as:
min$w$,b12||w||2min$w$,b21||w||2
subject to $yi(w \cdot xi+b) \geq 1yi(w \cdot xi+b) \geq 1$ for i=1,2,...,ni=1,2,...,n, where ww is the weight vector, bb is the bias term, $xixi$ are the training samples, $yiyi$ are their corresponding class labels, and nn is the number of training samples.

Q3. The kernel trick in SVM allows it to handle nonlinearly separable data by implicitly mapping the input features into a higher-dimensional space where the data becomes linearly separable. Instead of explicitly computing the transformation, the kernel function computes the dot product between the mapped vectors, avoiding the need to compute and store the transformed feature vectors explicitly.

Q4. Support vectors are the data points that lie closest to the decision boundary (hyperplane) and influence the position and orientation of the hyperplane. They are the critical elements in defining the margin of the SVM. Support vectors are the only data points that contribute to the determination of the decision boundary. In essence, they are the most informative points for the classification task.
Example: Consider a binary classification problem with two classes, represented by circles and squares in a 2D feature space. The support vectors are the data points that lie on the margin or are misclassified. In the graph, these points are the ones closest to the decision boundary (hyperplane), and they define the position and orientation of the separating hyperplane.

Q5.

- **Hyperplane**: In a linear SVM, the hyperplane is the decision boundary that separates the classes. It is a flat, high-dimensional plane determined by the weights and bias terms. In a 2D feature space, the hyperplane is a line. Example: A line that separates circles from squares in a 2D feature space.
- **Marginal Plane**: The marginal plane is the boundary parallel to the hyperplane that defines the margin of the SVM. It is equidistant from the hyperplane and passes through the support vectors. Example: In a 2D feature space, the marginal planes are lines parallel to the hyperplane and touching the support vectors.
- **Soft Margin**: In a soft-margin SVM, the margin is allowed to have some misclassified data points within it, which are penalized by a slack variable. This allows for a more flexible decision boundary that can handle noisy or overlapping data. Example: When the classes are not perfectly separable, and some data points are allowed to be within the margin.
- **Hard Margin**: In a hard-margin SVM, the margin is rigid, and no data points are allowed within it. This is suitable for linearly separable data with no noise. Example: When the classes are perfectly separable, and there are no misclassified data points within the margin.