Task 7 - Advanced SVM Interview Q&A

1. What is a support vector?

Support vectors are the data points closest to the decision boundary (hyperplane). They are the most critical elements in defining the position and orientation of the boundary.

2. What does the C parameter do?

The C parameter controls the trade-off between having a wide margin and classifying training points correctly. A low C allows some misclassifications (simpler model), while a high C tries to classify all points correctly (risk of overfitting).

3. What are kernels in SVM?

Kernels are functions that transform data into a higher-dimensional space to make it easier to separate with a linear boundary. Common kernels include linear, polynomial, and RBF (Gaussian).

4. What is the difference between linear and RBF kernel?

Linear kernel creates a straight-line boundary and works well when data is linearly separable. RBF kernel creates curved boundaries and is better for non-linear data.

5. What are the advantages of SVM?

SVMs are effective in high-dimensional spaces, use memory efficiently (support vectors only), work well with clear margin separation, and are versatile with kernel functions.

6. Can SVMs be used for regression?

Yes, using a variant called Support Vector Regression (SVR). Instead of finding a boundary between classes, SVR tries to fit a line or curve within a margin of tolerance.

7. What happens when data is not linearly separable?

SVM can use soft margins (with C) to allow some misclassifications or apply kernel tricks to transform the data into a space where linear separation is possible.

8. How is overfitting handled in SVM?

SVM handles overfitting by adjusting the C parameter and using kernels carefully. Lower C values allow the model to generalize better. Cross-validation also helps select the best parameters.