**SVM and Grid Search – 20 Questions (20 mins):**

### **Q1. What is a Support Vector Machine (SVM)?**

**Answer:**  
SVM is a supervised machine learning algorithm used for classification and regression tasks. It finds the optimal hyperplane that best separates the data into classes with the maximum margin.

### **Q2. How does SVM differ from logistic regression?**

**Answer:**  
SVM aims to maximize the margin between classes using support vectors, whereas logistic regression models probabilities and uses a sigmoid function to make predictions.

### **Q3. What is a hyperplane in the context of SVM?**

**Answer:**  
A hyperplane is a decision boundary that separates data points of different classes in a feature space.

### **Q4. Define margin in SVM.**

**Answer:**  
The margin is the distance between the hyperplane and the nearest data points from each class. SVM aims to maximize this margin.

### **Q5. What is the kernel trick in SVM, and why is it useful?**

**Answer:**  
The kernel trick transforms input data into a higher-dimensional space to make it possible to separate data that is not linearly separable in the original space.

### **Q6. List four common types of kernels used in SVM.**

**Answer:**

1. Linear
2. Polynomial
3. Radial Basis Function (RBF)
4. Sigmoid

### **Q7. Explain the concept of soft margin in SVM.**

**Answer:**  
A soft margin allows some misclassifications or violations of the margin to improve generalization and handle overlapping classes.

### **Q8. What is the role of the C parameter in SVM?**

**Answer:**  
The C parameter controls the trade-off between maximizing the margin and minimizing classification errors. A smaller C allows more margin violations.

### **Q9. How does the Radial Basis Function (RBF) kernel work?**

**Answer:**  
The RBF kernel measures the similarity between data points using a Gaussian function, effectively mapping them to an infinite-dimensional space to allow non-linear separation.

### **Q10. Define Grid Search.**

**Answer:**  
Grid Search is a method to systematically work through multiple combinations of hyperparameter values to determine the best-performing set.

### **Q11. Why is cross-validation used with Grid Search?**

**Answer:**  
Cross-validation is used to assess the generalization performance of each parameter combination and prevent overfitting.

### **Q12. What is GridSearchCV in Scikit-Learn?**

**Answer:**  
GridSearchCV is a function in Scikit-Learn that performs Grid Search combined with cross-validation to find the best hyperparameters for a model.

### **Q13. What are the types of machine learning algorithms?**

**Answer:**

1. Supervised Learning
2. Unsupervised Learning
3. Semi-Supervised Learning
4. Reinforcement Learning

### **Q14. How is C parameter related to the soft margin in SVM?**

**Answer:**  
The C parameter determines the softness of the margin; lower C allows a softer margin with more tolerance to misclassifications.

### **Q15. What is the purpose of the** cv **parameter in GridSearchCV?**

**Answer:**  
The cv parameter specifies the number of folds for cross-validation, controlling how the training data is split during parameter tuning.

### **Q16. What are support vectors in SVM?**

**Answer:**  
Support vectors are the data points that lie closest to the decision boundary (hyperplane) and directly influence its position.

### **Q17. How would you interpret** best\_score\_ **from GridSearchCV?**

**Answer:**  
best\_score\_ indicates the highest average cross-validation score achieved by any parameter combination during the grid search.

### **Q18. Give an example of when Grid Search might not be suitable.**

**Answer:**  
Grid Search is not suitable for high-dimensional hyperparameter spaces due to its computational cost. Randomized Search or Bayesian optimization may be better in such cases.

### **Q19. Why might SVM not perform well on large datasets?**

**Answer:**  
SVMs are computationally intensive in both training and prediction, especially with non-linear kernels, making them less suitable for large datasets.

### **Q20. What is Supervised Machine Learning and how is it different from Unsupervised Machine Learning?**

**Answer:**  
Supervised learning uses labeled data to train models, while unsupervised learning finds patterns or groupings in data without labels.