Crash Course in Statistical Learning on SVM Classification and Regression Quiz Questions

1. Which of the following is a disadvantage of using a linear SVM for classification?
   (a) Linear SVMs are not able to handle non-linearly separable data
   (b) Linear SVMs can be computationally expensive for large datasets
   (c) Linear SVMs require less tuning of hyperparameters compared to non-linear SVMs
   (d) Linear SVMs are more prone to overfitting than non-linear SVMs

   Answer: (a) Linear SVMs are not able to handle non-linearly separable data.

   Explanation: Linear SVMs assume that the classes are linearly separable, which can be a limitation in real-world data. Non-linear SVMs can be used to handle non-linearly separable data.

2. Which of the following is a hyperparameter that can be tuned in SVMs?
   (a) Intercept
   (b) Coefficient of variation
   (c) Regularization parameter
   (d) Standard deviation

   Answer: (c) Regularization parameter.

   Explanation: The regularization parameter, often denoted as C, is a hyperparameter that can be tuned in SVMs. It controls the trade-off between the complexity of the decision boundary and the amount of misclassification allowed.

3. Which of the following is a technique that can be used to handle imbalanced datasets in SVMs?
   (a) Oversampling the minority class
   (b) Undersampling the majority class
   (c) Both (a) and (b)
   (d) None of the above

   Answer: (c) Both (a) and (b).

Explanation: Imbalanced datasets can be handled in SVMs using oversampling the minority class, undersampling the majority class, or a combination of both.

4. Which of the following is not an advantage of using SVMs for regression?
   (a) SVMs can handle both linear and non-linear relationships between the input and output variables
   (b) SVMs are less prone to overfitting compared to other regression techniques
   (c) SVMs are computationally efficient for large datasets
   (d) SVMs can handle datasets with high-dimensional input spaces

   Answer: (c) SVMs are computationally efficient for large datasets.

   Explanation: SVMs can be computationally expensive for large datasets due to the quadratic programming problem that needs to be solved.

5. Which of the following is a common method used for tuning the hyperparameters in SVMs?
   (a) Grid search
   (b) Random search
   (c) Bayesian optimization
   (d) All of the above

   Answer: (d) All of the above.

   Explanation: Grid search, random search, and Bayesian optimization are commonly used methods for tuning the hyperparameters in SVMs.

6. Which of the following is not a similarity measure used in kernel functions for SVMs?
   (a) Euclidean distance
   (b) Cosine similarity
   (c) Radial basis function
   (d) Correlation coefficient
   Answer: (a) Euclidean distance.

   Explanation: Euclidean distance is not a similarity measure used in kernel functions for SVMs. The most commonly used similarity measures are radial basis function, cosine similarity, and correlation coefficient.

7. Which of the following is a limitation of using SVM for regression?
   a) SVM works only with linearly separable data
   b) SVM cannot handle categorical data

c) SVM is sensitive to outliers
d) SVM cannot handle high-dimensional data

Answer: c) SVM is sensitive to outliers. SVM tries to find the largest margin between the decision boundary and the data points, and outliers can significantly affect the location of the decision boundary.

Explanation: In SVM regression, the objective is to find a hyperplane that maximizes the margin while also minimizing the amount of violation (i.e., points that fall outside the margin). Outliers can have a significant impact on the location of the hyperplane, as they may force the hyperplane to be shifted in order to accommodate them. This can result in poor performance and overfitting.

8. What is the main difference between SVM regression and SVM classification?
   a) In regression, the output variable is continuous, while in classification, the output variable is discrete
   b) In regression, the objective is to minimize the error between the predicted and actual values, while in classification, the objective is to maximize the margin between the decision boundary and the data points
   c) SVM regression can only use linear kernel functions, while SVM classification can use nonlinear kernel functions
   d) None of the above

   Answer: a) In regression, the output variable is continuous, while in classification, the output variable is discrete.

   Explanation: In SVM regression, the objective is to find a function that can predict continuous output values. In contrast, in SVM classification, the objective is to find a function that can classify the input data into different classes (discrete output values).

9. Which of the following statements is true about regularization in SVM?
   a) Increasing the regularization parameter leads to higher model complexity
   b) Regularization is used to prevent overfitting in SVM
   c) Regularization is only applicable to linear SVM
   d) Both b) and c)

   Answer: b) Regularization is used to prevent overfitting in SVM.

   Explanation: Regularization is a technique used to prevent overfitting in machine learning models. In SVM, regularization is achieved by adding a penalty term to the objective function that penalizes large values of the weights. By increasing the regularization parameter, we can control the

amount of regularization applied and prevent the model from overfitting to the training data.

10. Which of the following is a limitation of using SVM for classification?
    a) SVM cannot handle imbalanced datasets
    b) SVM is computationally expensive for large datasets
    c) SVM can only handle binary classification
    d) SVM can handle all types of datasets without any limitations

    Answer: a) SVM cannot handle imbalanced datasets.

    Explanation: SVM works by maximizing the margin between the decision boundary and the data points. In cases where the dataset is imbalanced (i.e., one class has significantly more samples than the other), SVM may give more weight to the majority class,