Support Vector Machines-2 quiz

8 out of 8 correct

1. Which of the following is a kernel function used in SVM?

- ◯ Linear kernel
- ◯ Polynomial kernel
- ◯ Radial basis function kernel
- ◉ All of the above

Explanation: SVM can use a variety of kernel functions, including the linear kernel, polynomial kernel, and radial basis function kernel.

2. The parameter "p" in the polynomial kernel controls:

- ◯ The width of the kernel
- ◉ The degree of the polynomial
- ◯ The regularization parameter
- ◯ The margin of the decision boundary

Explanation: The parameter "p" in the polynomial kernel controls the degree of the polynomial. A higher value of "p" results in a higher degree polynomial and a more complex decision boundary.

3. The Gaussian/RBF kernel is prone to overfitting when:

- ◯ The value of "γ" is too small

- [x] The value of "γ" is too large
- [ ] The dataset is not properly normalized
- [ ] Both a) and b)

4. Data leakage can occur in SVM when:

- [ ] The test set is used for model training
- [ ] The training set is not representative of the entire population
- [ ] The same data is used for feature selection and model training
- [x] All of the above

Explanation: Data leakage can occur in SVM when the test set is used for model training, the training set is not representative of the entire population, or the same data is used for feature selection and model training. Data leakage can result in overestimating the performance of the model and can lead to poor generalization to new data.

5. Which of the following is a technique to prevent data leakage in SVM?

- [ ] Stratified sampling
- [x] Cross-validation
- [ ] Feature scaling
- [ ] Regularization

Explanation: Cross-validation is a technique to prevent data leakage in SVM and other machine learning models. It involves splitting the data into training and validation sets multiple times and using the validation set to evaluate the performance of the model. This helps to ensure that the model does not overfit the training data and can generalize well to new data.

6. Which of the following is the correct formula to calculate accuracy from a confusion matrix?

- (TP + TN) / (TP + TN + FP + FN)
- TP / (TP + TN + FP + FN)
- TP / (TP + TN + FP + FN)
- (TP + TN) / (P + N)

7. Which of the following is the correct interpretation of the false positive rate (FPR) in a confusion matrix?

- The proportion of actual positive instances that are correctly identified by the model
- The proportion of actual negative instances that are incorrectly identified as positive by the model
- The proportion of actual negative instances that are correctly identified by the model
- The proportion of actual positive instances that are incorrectly identified as negative by the model

Explanation: The false positive rate (FPR) is calculated as the number of false positive predictions divided by the total number of actual negative instances in the data, which is equal to FP / (FP + TN). FPR measures the rate at which the model incorrectly identifies negative instances as positive.

8. Which of the following is the correct formula to calculate F1 Score from true positive (TP), false positive (FP), and false negative (FN) values?

- 2 * TP / (2 * TP + FP + FN)
- TP / (TP + TN)
- TN / (TP + TN + FP + FN)

- ◯ TP / (TP + FN)

Explanation: F1 Score is the harmonic mean of precision and recall, and is calculated as 2 * precision * recall / (precision + recall), which can be simplified to 2 * TP / (2 * TP + FP + FN).

Submit