AI Course

Chapter 5. Quiz

For instructors (with answers)

ⓒ2022 SAMSUNG. All rights reserved.

Samsung Electronics Corporate Citizenship Office holds the copyright of this document.

This document is a literary property protected by copyright law so reprint and reproduction without permission are prohibited.

To use this document other than the curriculum of Samsung Innovation Campus, you must receive written consent from copyright holder.

1. Which of the following statements about supervised learning is incorrect?
2. Supervised learner is to make predictions correctly with given data from training data.
3. Precision is the ratio of what the model classifies as true to what it actually classifies as true. This can be called the answer rate.
4. Recall ratio is the ratio of what model predicts to be true out of what is actually true.
5. Recall ratio is calculated as . (where tp is true positive and fp is false positive.)

Answer. 4

is the expression for precision.

1. What is the method of estimating regression coefficients in regression analysis?

Answer. Least Squares Estimation

1. Which of the following statements about overfitting is incorrect?
2. A characteristic of a pattern found in a specific dataset that exists only in that dataset, or a characteristic of a pattern that does not generalize and exists only in a specific data is called overfitting
3. The extraction methods used to solve the overfitting problem include the holdout method, the cross-validation method, the bootstrap method, etc.
4. In the holdout method, cross-validation is performed by dividing the training data for model learning and construction and the verification data for performance evaluation, and the results of the verification data are used only for performance measurement without affecting the model.
5. The cross-validation method is a method to reselect training data repeatedly and based on restoration extraction, and it is suitable when the total amount of data is not large.

Answer. 4

1. Which of the following about the decision tree model is incorrect?
2. Decision trees are used for corporate bankruptcy prediction, stock price cap prediction, exchange rate prediction, and economic outlook prediction.
3. CHAID(Chi-squared Automatic Interaction Detection) algorithm is an algorithm that performs separation using chi-squared or F-test.
4. CART(Classification and Regression Trees) algorithm is an algorithm that performs separation using the Gini index, and 100 represents a perfectly pure node between 0 and 100.
5. The entropy index of the C4.5 algorithm can be obtained by using the likelihood ratio test statistic in the polynomial distribution and subtracting the entropy of the child node from the entropy of the parent node.

Answer. 3

CART algorithm is an algorithm that performs separation using the Gini index, and 1 represents a perfectly pure node between 0 and 1. The Gini index measures the impurity or diversity of each node.

1. When I trained a SVM classifier using the RBF kernel, it seems to underfit the training set.

Should I increase gamma or decrease gamma? What about C?

Answer.

If the SVM classifier trained on the RBF kernel underfits the training set, the regulation may be too large. To reduce the regulation, either gamma or c (or both) should increase.