AI Course

Chapter 6. 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 learning is incorrect?
2. Supervised learning uses both input data and learning data given a result, but unsupervised learning is used in the form of learning using only input data without learning data.
3. Unsupervised learning can be said to be a learning method that informs the problem but does not tell the answer.
4. As an unsupervised learning algorithm, hierarchical clustering analysis and K-means clustering can be used for clustering.
5. K-means clustering has the advantage of being easy to determine the initial number of clusters and easy to interpret the results.

Answer. 4

It has the disadvantage that it is difficult to determine the initial number of clusters and to interpret the results.

1. Which of the following statements about clustering is incorrect?
2. The most commonly used clustering technique is K-Means.
3. Hierarchical Clustering is decided after clustering without determining the number of groups.
4. The algorithm of hierarchical clustering has K-Medoids.
5. Clustering is a technique that allocates and analyzes groups based on the similarity or dissimilarity between each entity.

Answer. 3

Clustering can be divided into hierarchical and non-hierarchical methods, and K-medoids belongs to non-hierarchical methods.

1. The EM algorithm belongs to the greedy algorithm. Please explain the reason.

Answer

The E and M steps are iterated recursively till the parameters converge to the optimal solution. The E and M steps make the choice that seems to be the best at the given moment. Each step it chooses the optimal choice, without knowing the future or the past.

1. Write three or more examples of clustering algorithm. (ex. Data analysis)

Answer

Data analysis, customer classification, recommendation system, searching engine, image segmentation, semi-supervised learning, anomaly detection, spam filter, and etc.

※ Scoring guide

Only students answering three or more examples out of the suggested answers or other clustering algorithm cases can get score.

1. Explain about the two methods of choosing optimal number of clusters when using k-means. Describe how to find the optimal number of clusters using the two methods.

Answer

To find the optimal number of clusters using the k means, the **elbow method** is the simplest way. When drawing a graph of the inertia (the average of squared distances of the samples to their nearest centroids) as a function of the number of clusters, the curve has roughly the shape of an arm, and it contains an inflexion point called the elbow near which the inertia curve does not rapidly decrease. This point is close to the optimal number of clusters.

**The silhouette method** is also a method to find the optimal number of clusters. By plotting silhouette scores as a function of number of clusters, there is often a highest value (pick) of the graph. This highest value is mostly the optimal number of clusters. Silhouette score for the data set is used for measuring the mean of the Silhouette Coefficient for all samples. The value of Silhouette score is near +1 when the samples are well matched to their own cluster and well-separated from other clusters. When the samples are near the wrong clusters, the value of Silhouette score gets near -1. By generating the silhouette diagram, it is possible to perform more analysis.

※ Scoring guide

This quiz is to check whether a student well understands the two methods (elbow method and silhouette method) of selecting optimal number of clusters. In order to get score, students have to well describe both of the elbow method and the silhouette method for choosing the optimal number of clusters.