

Machine Learning

Assignment 3

1.(d)

2.(d)

3.(c)

4.(b)

5.(d)

6.(c)

7.(d)

8.(a)

9.(a)

10.(a)

11.(a)

12.(b)

13. Clustering is important in data analysis and data mining applications . It is the task of grouping a set of objects so that objects in the same group are more similar to each other than to those in other groups.

14. Applying unsupervised feature learning to input data using either RICA or SFT improves clustering performance. Surprisingly for some cases , high clustering performance can be achieved by simply performing K- means clustering on the ICA components after PCA dimension reduction on the input data.

Statistics worksheet 3

1. (b)

2.(c)

3.(a)

4.(a)

5.(c)

6.(b)

7.(b)

8.(d)

9.(a)

10. Baye's theorem states that conditional probability of an event based on the occurrence of another event is equal to the likelihood of the second event given the first event multiplied by the probability of the first event.

11. Z – score is a statistical measurement that describes a value's relationship to the mean of a group of values. Z – score is measured in terms of standard deviations from the mean. If z – score is 0 it's indicates that the data point's score is identical to the mean score. If z – score is 1 it's indicates a value that is one standard deviation from the mean. Z – score maybe positive or negative , with positive value indicating the score is above the mean and a negative score indicating it is below the mean.

12. T – test is used to measure and compare the means of two groups of data . $T = \frac{\text{mean} - \text{theoretical mean}}{\text{sd}/\sqrt{n}}$, n is the sample size.

13. Percentile is a score which helps us to compare a particular score and scores of rest of the group.

14. ANOVA is a analysis of variance , it has some statistical models and some estimations that helps to analyses the different among means.

15. ANOVA is helpful in finding difference between the means of two groups on a single variable.