

## MACHINE LEARNING

1. D
2. D
3. A
4. D
5. A
6. C
7. D
8. –
9. D
1. A
2. –
3. –
4. 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.
5. –

## STATISTICS WORKSHEET 3

1. A
2. C
3. A
4. A
5. –
6. A
7. B
8. –
9. A
10. Bayes theorem states that the 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. A z score is a numerical measurement that describes a values relationship to the mean of a group of values.
12. A t test is a statistical test that is used to compare the means of two groups. It is often used in hypothesis testing to determine whether a process or treatment actually has an effect on the population of interest or whether two groups are different from one another.
13. A percentile is a term that describes how a scores from the same set. While there is no universal definition of percentile it is commonly expressed as the percentage of values in a set of data scores that fall below a given value.
14. ANOVA is a collection of statistical models and their associated estimation procedures used to analyze the difference among means.

15. ANOVA is helping for testing three or more variables. It is similar to multiply two sample t sets, however it results in fewer type I errors and is appropriate for a range of issues. ANOVA groups difference by comparing the means of each group and includes spreading out the variance in to diverse sources.