Machine Learning:

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 method helps in grouping valuable data into clusters and from that picks appropriate results based on different techniques. For example, in information retrieval, the results of the query are grouped into small clusters, and each cluster has irrelevant results. By Clustering techniques, they are grouped into similar categories and each category is subdivided into sub-categories to assist in the exploration of queries output. clustering methods try ro split the data into groups that are supposed to be more homogeneous within and more different to another. Thrn you don't have to look at every object, but only at some of each cluster to hopefully learn something about the whole cluster (and your whole data set). Centroid methods such as k-means even can proviee a "prototype" for each cluster, albeit it is a good idea to also lool at other points within the cluster. You may also want to do outlier detection and look at some of the unusual objects. This scenario is somewhere inbetween of **sampling representative objects** and **reducing the data set size** to become more manageable. The key difference to above points is that the result is usually not "operationalized" automatically, but because explorative clustering results are too unreliable (and thus require many iterations) need to be analyzed manually.

14. Clustering is an unsupervised machine learning methodology that aims to partition data into distinct groups, or clusters. There are a few different forms including hierarchical, density, and similarity based. Each have a few different algorithms associated with it as well. One of the hardest parts of any machine learning algorithm is feature engineering, which can especially be difficult with clustering as there is no easy way to figure out what best segments your data into separate but similar groups.

We can then attempt to use Gradient Descent on this loss function to try and minimize it with respect to the similarity matrix. Gradient Descent is one of the most common optimization algorithms in machine learning that is used to find best parameters of a given function by using the function gradient, a combination of the partial derivatives. By taking steps proportional to the negative of the gradient, we can try to find the local minimum of the function. We will continually update the weights until either our maximum number of iterations has been met, or the function converges. So the gradient descent will be of our loss function with a partial derivative in respect to the weights. We will update the weights every iteration with respect to the gradient and learning rate.

A good representation of its effectiveness is fuzzy c-means, a relative of the commonly used k-means algorithm. It works in a very similar fashion to k-means, but rather results in something called the fuzzy partition matrix instead of just a cluster label.The fuzzy partition matrix is a set of weights that measure how similar a single point is to a given cluster center, close to how our similarity matrix is used previously. It can also be calculated using a weighted distance metric which we can feed our new found optimal weights. This will also then go back into updating the cluster centers. Like K-means, this results in the cluster centers shifting with each iteration, until the maximum number of iterations or a certain improvement threshold has been met.In fuzzy c-means, you would have a very similar goal as to our original loss function. You would like less “fuzzyness” from points, and you want them all to be as close as possible to their cluster centers, and further away from others. A good measure of the fuzzy clustering algorithm is Dunn’s partition coefficient, a sum of all components of the fuzzy partition matrix.

SQL:

1. CREATE Table Customers(

CustomerNumber INT,

CustomerName VARCHAR(25),

contactLastName VARCHAR(20),

contactFirstName VARCHAR(20),

phone BIGINT,

adressLine1 VARCHAR(100),

adressLine2 VARCHAR(100),

city VARCHAR(20),

state VARCHAR(20),

postalCode INT,

country VARCHAR(20),

salesRepEmployeeNumber VARCHAR(20),

creditLimit INT

);

**2)**

CREATE Table Orders(

orderNumber VARCHAR(50),

orderDate DATETIME,

requiredDate DATETIME,

shippedDate DATETIME,

status VARCHAR(50),

comments VARCHAR(50),

customerNumber VARCHAR(50)

);

3.SELECT \* FROM Orders;

4.SELECT comments FROM Orders;

5.SELECT COUNT(\*) ,orderDate FROM orders

GROUPBY (orderDate);

6.SELECT employeeNumber,firstName,lastName from employees;

7.

8.

9.

10. SELECT products productName, MSRP, productDescription FROM Products;

11.

12.

13.

14.

15.

Statistics:

1.B

2.C

3.A

4.A

5.A

6.B

7.B

8.D

9.A

10. Bayes Theorem provides a principled way for calculating a conditional probability.It is a deceptively simple calculation, although it can be used to easily calculate the conditional probability of events where intuition often fails.Although it is a powerful tool in the field of probability, Bayes Theorem is also widely used in the field of machine learning. Including its use in a probability framework for fitting a model to a training dataset, referred to as maximum a posteriori or MAP for short, and in developing models for classification predictive modeling problems such as the Bayes Optimal Classifier and Naive Bayes.

Bayes' theorem thus gives the probability of an event based on new information that is, or may be related, to that event. The formula can also be used to see how the probability of an event occurring is affected by hypothetical new information, supposing the new information will turn out to be true. For instance, say a single card is drawn from a complete deck of 52 cards. The probability that the card is a king is four divided by 52, which equals 1/13 or approximately 7.69%. Remember that there are four kings in the deck. Now, suppose it is revealed that the selected card is a face card. The probability the selected card is a king, given it is a face card, is four divided by 12, or approximately 33.3%, as there are 12 face cards in a deck.

11. A Z-score is a numerical measurement that describes a value's relationship to the mean of a group of values. Z-score is measured in terms of [standard deviations](https://www.investopedia.com/terms/s/standarddeviation.asp) from the mean. If a Z-score is 0, it indicates that the data point's score is identical to the mean score. A Z-score of 1.0 would indicate a value that is one standard deviation from the mean. Z-scores may be positive or negative, with a positive value indicating the score is above the mean and a negative score indicating it is below the mean. The Z-score, by contrast, is the number of standard deviations a given data point lies from the mean. For data points that are below the mean, the Z-score is negative. In most large data sets, 99% of values have a Z-score between -3 and 3, meaning they lie within three standard deviations above and below the mean.

12. A t-test is a type of inferential [statistic](https://www.investopedia.com/terms/s/statistics.asp) used to determine if there is a significant difference between the means of two groups, which may be related in certain features. It is mostly used when the data sets, like the data set recorded as the outcome from flipping a coin 100 times, would follow a normal distribution and may have unknown variances. A t-test is used as a hypothesis testing tool, which allows testing of an [assumption](https://www.investopedia.com/ask/answers/073115/what-assumptions-are-made-when-conducting-ttest.asp) applicable to a population. Essentially, a t-test allows us to compare the average values of the two data sets and determine if they came from the same population. In the above examples, if we were to take a sample of students from class A and another sample of students from class B, we would not expect them to have exactly the same mean and standard deviation. Similarly, samples taken from the placebo-fed control group and those taken from the drug prescribed group should have a slightly different mean and standard deviation.

13. Percentile” is in everyday use, but there is no universal definition for it. The most common definition of a percentile is a number where **a certain percentage of scores fall below that number.** You might know that you scored 67 out of 90 on a test. But that figure has no real meaning unless you know what percentile you fall into. If you know that your score is in the 90th percentile, that means you scored better than 90% of people who took the test. Percentiles are a great tool to use when you need to know the relative standing of a value. Where does a value fall within a distribution of values? While the concept behind percentiles is straight forward, there are different mathematical methods for calculating them. In this post, learn about percentiles, special percentiles and their surprisingly flexible uses, and the various procedures for calculating them.

14. Analysis of variance (ANOVA) is an analysis tool used in statistics that splits an observed aggregate variability found inside a data set into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not. Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study. The ANOVA test is the initial step in analyzing factors that affect a given data set. Once the test is finished, an analyst performs additional testing on the methodical factors that measurably contribute to the data set's inconsistency. The analyst utilizes the ANOVA test results in an f-test to generate additional data that aligns with the proposed [regression](https://www.investopedia.com/terms/r/regression.asp) models.The ANOVA test allows a comparison of more than two groups at the same time to determine whether a relationship exists between them. The result of the ANOVA formula, the F statistic (also called the F-ratio), allows for the analysis of multiple groups of data to determine the variability between samples and within samples.

If no real difference exists between the tested groups, which is called the [null hypothesis](https://www.investopedia.com/terms/n/null_hypothesis.asp), the result of the ANOVA's F-ratio statistic will be close to 1. The distribution of all possible values of the F statistic is the F-distribution. This is actually a group of distribution functions, with two characteristic numbers, called the numerator [degrees of freedom](https://www.investopedia.com/terms/d/degrees-of-freedom.asp) and the denominator degrees of freedom.

15. ANOVA is helpful for testing three or more variables. It is similar to multiple two-sample [t-tests](https://www.investopedia.com/terms/t/t-test.asp). However, it results in fewer [type I errors](https://www.investopedia.com/terms/t/type_1_error.asp) and is appropriate for a range of issues. ANOVA groups differences by comparing the means of each group and includes spreading out the variance into diverse sources. It is employed with subjects, test groups, between groups and within groups. You might use Analysis of Variance (ANOVA) as a marketer, when you want to test a particular hypothesis. You would use ANOVA to help you understand how your different groups respond, with a null hypothesis for the test that the means of the different groups are equal. If there is a statistically significant result, then it means that the two populations are unequal (or different).