Q1. The basic distinction between a linear regression model and generalised (1) linear regression model is the following

1. **The errors in the linear regression are normally distributed while they can have a more general distribution for the generalised linear model
2. The errors in the linear regression model are homoskedstic while they are heteroskedstic in a generealised linear model
3. The generalised linear model is not used for continuous dependent variable while that is not the case with the linear regression model
4.   The linear regression model is easy to estimate while the generalised linear regression model is not easy to estimate.

Q2. The process of training a predictive model with well defined target values is known as (1)

1. Unsupervised learning
2. **Supervised learning
3. Model estimation
4. Model testing

Q3. The following is true of the Knn algorithm (2)

1. It has slow training phase
2. It has a fast classification phase
3. **Makes no assumptions about the data distribution
4. Produces a predictive model

Q4. The trade off between over fitting and under fitting training data is called (3)

1. **The bias-variance tradeoff
2.   The residual sum of squares
3. The tradeoff curve
4. The null deviance

Q5. The following transformation is called the z-score transformation (2)

1. (X-max(X))/(Max(X)-Min(X)
2. (X-Mean(X))/(Max(X)-Min(X))
3. **(X-Mean(X))/Standard Deviation of X
4. Mean(X)/Max(X)

Q6. How do you deal with Euclidean distance for nominal data in the context of Knn (3) classification?

1. **Using dummy coding
2. Ignoring such data
3. Deleting those observations
4. Replacing these observations by 0

Q7. The following is the correct code for a function that normalizes the data (1)

1. normalize <- function(x) { return ((x - max(x)) / (max(x) - min(x))) }
2. **normalize <- function(x) { return ((x - min(x)) / (max(x) - min(x))) }
3. normalize <- function(y) { return ((x - min(x)) / (max(x) - min(x))) }
4. normalize <- function(x) { return ((x - min(x)) / (min(x) - max(x))) }

Q8. The following is the correct code to execute a Knn model (where train is the (2) training data, test is the testing data, labels are stored in train_labels and we have a 7 nearest neighbour classifiction

1. wbcd_test_pred <- knn(train =train, test = test, cl =train_labels, k=21)
2. **wbcd_test_pred <- knn(train =train, test = test, cl = train_labels, k=07)
3. wbcd_test_pred <- knn(train = wbcd_train, test = wbcd_test, cl = wbcd_train_labels, k=21)
4. wbcd_test_pred <- knn(train = test, test = train, cl = test, k=21)

Q9. The following is not true of the naïve Bayes classifier (2)

1. It is easy to obtain the estimated probability for a prediction
2. It deals well with noisy and missing data
3. It is simple, fast and effective
4. **It is ideal for data sets with large number of numeric variables

Q10 Suppose 40% of spam messages contain the word "free". 10% of (3) messages are spam.10% of messages contain the word "free". The probability that a message is spam, given that it contains the word "free" is

1. 0.5
2. 0.1
3. 0.2
4. **0.4

Q11. The assumption of Class conditional independence means (3)

1. **Event are independent provided that they are conditioned on the same class value
2. Events are independent in general
3. Events are disjoint in general
4. The classes are unrelated

Q12. What is the purpose of the Laplace estimator in the context of Naïve Bayes classifiers (3)

1. To ensure that probabilities are not negative
2. To ensure that probabilities sum to one
3. **To ensure non –zero probabilities
4. To ensure that cdf integrates to 1

Q13. Binning in the context of naïve Bayes classifiers refers to (2)

1. Deleting incorrect values
2. Averaging missing values
3. Extrapolating missing values
4. **Discretizing numeric values

Q15. The command corpus_clean <- tm_map(x, tolower) would achieve the following (1)

1. Clean the data
2. **Convert the collection x to lower case
3. Convert the collection x to lower magnitude
4. Create maps

Q16. In order to create a cross table of predicted and actual values, the below set of (1)
code will work
   a) CrossTable(predicted, actual.prop.chisq = FALSE, prop.t = FALSE, dnn = c('predicted', 'actual'))
   b) **CrossTable(predicted, actual, prop.chisq = FALSE, prop.t = FALSE, dnn = c('predicted', 'actual'))
   c) CrossTable(predicted, actual,, prop.chisq = FALSE, prop.t = FALSE, dnn = c('predicted', 'actual'))
   d) Crosstable(predicted,actual, sms_raw_test$type, prop.chisq = FALSE, prop.t = FALSE, dnn = c('predicted', 'actual'))

Q17. The following is a weakness of the tree based classifiers (3)

1. **They are liable to overfit or under fit models
2. They always overfit models
3. They always underfit models
4. They take a lot of time to fit models

Q18. Let S be the split before partition on feature A, and let M be the split after partition (3)
 on feature A. The ionformation gin from the split is measured as:
1. Entropy(M)-Entropy(S)
2. Entropy(M)+Entropy(S)
3. Entropy(M)/Entropy(S)
4. **Entropy(S)-Entropy(M)

Q19. Pruning a decision tree involves: (3)

1. **Reducing its size so that it fits well to unseen data
2. Reducing the tree size to fit well to training data
3. Increasing the tree size
4. Reducing the number of nodes in the tree

Q20. Suppose Y is a binary valued dependent variable (1 and 0) and x1 and x2 (1)
are explanatory variables. Predicting the probability Y=1 involves estimating a
1. Linear regression
2. **A logistic regression
3. A Poisson regression
4. A linear probability model

Q21. Let m.rpart be a regression tree object, which we need to visualise.   (3)
The following is the correct code
1. **rpart.plot(m.rpart, digits = 4, fallen.leaves = TRUE, type = 3, extra = 101)
2. plot.rpart(m.rpart, digits = 4, fallen.leaves = TRUE, type = 3, extra = 101)
3. rpart(m.rpart, digits = 4, fallen.leaves = TRUE, type = 3, extra = 101)
4. rpart.plot(rpart, digits = 4, fallen.leaves = TRUE, type = 3, extra = 101)

Q22. The following statement is valid in the case of standard deviation reduction (3)
in the context of regression trees
1. It is used to reduce the length of the nodes
2. **Measures the reduction in standard deviation from the original value to the weighted standard deviation post-split.
3. Measures the reduction in standard deviation from the original value to the un-weighted standard deviation post-split.
4. Measures the reduction in standard deviation from the original value to the weighted standard deviation pre-split.

Q23. The following is a strength of the tree based classification compared to (3)

linear regression model:

1. Does not require the user to prespecify a model
2. Is always correct
3. Is easier to implement
4. Has higher predictive probability

Q24. The following is one of the strengths of support vector machines (SVM) (3)

1. Not very prone to overfitting
2. Not very prone to underfitting
3. **Is generally quick to train
4. Easy to interpret

Q25. In the context of SVM with linear separability, the class boundries are called (3)

1. The concave hull
2. The symmetric box
3. **The convex hull
4. The quadratic optimiser

1. In the context of SVM, the kernel trick is used to
   a) Linearise relationships
   b) Estimate models
   c) Normalise the data
   d) Equalise the data

Answer key

Question no	Answer	Question No	Answer
1	a	14	b
2	b	15	b
3	c	16	a
4	a	17	d
5	c	18	a
6	b	19	b
7	b	20	a
8	d	21	b
9	d	22	a
10	a	23	a
11	c	24	c
12	d	25	a
13	b		