**ANSWER 1** - The value of correlation coefficient will always be: between -1 and 1.

**ANSWER 2** – Regularization is not considered as dimensionality reduction technique so, both Lasso regularization and ridge regularization will not be considered as a dimensionality reduction technique.

**ANSWER 3** – Hyper plane is not a kernel support in support vector machine, kernel is used when the data is not linearly separable, kernel in support vector machine helps to transform and create new features with the existing ones so that the model can separate the non-linear points.

**ANSWER 4** – Logistic Regression is not suitable for non linear classification, as the assumptions for logistic regression is that the data is linearly separable and the features are independent of each other. Support vector machines are linear separable, however by using kernel it can work on non-linear datasets

**ANSWER 5** – new coefficient of X will be old coefficient of ‘X’ ÷ 2.205.

**ANSWER 6** – As we increase the number of estimators is ADABOOST the model accuracy will increase, (assuming that there are only week learners (stump) decision trees). There might come appoint where the model starts to get overfit.

**ANSWER 7** – Random forest are not easy to interpret, as in random forest bagging concept is used where we create many dense decision trees, it becomes hard to interpret all the dense connection and the results.

**ANSWER 8** – D) All of the above

**ANSWER 9** – C) and D)

**ANSWER 10** – A),B) and D)

**ANSWER 11** – Outliers are the point in a data set who are very far away or varies a lot from other data points. Usually are those points that lies an abnormal distance from all the values in a random sample of population. For example: weights in kgs of all the students in a class = 40,54,52,4,52,51,43,15,105, here 105 will be an outlier as it is very far away from other data points. Mean gets affected because of outliers hence having an outlier in a data set is not good if you are doing some testing with the data.  
Outliers can be identified using z-score or interquartile range.  
To find the outliers using interquartile range steps are:  
1st find the 1st and 3rd quantile of the data then subtract 1st quantile which is 25 percentiles from 3rd quantile which is 75 percentile you will get the interquartile range. Then find the threshold value which shows that this is the limit of the data points beyond this point the point will be an outlier.  
Any observations that are more than 1.5 IQR below Q1 or more than 1.5 IQR above Q3 are considered outliers. If a point > 1.5 IQR + 3rd quantile it will be an outlier or if a point < 1.5 IQR – 1st quantile it will be an outlier.

**ANSWER 12** – In Bagging (boot strap aggregation) we train multiple models using dense decision tree, each decision tree shows high variance, number of Decision trees are known as number of estimators. Each Decision tree will train on random subset of data. Then the result from each decision tree is calculated and based on the majority final output is classified.  
  
However, in Boosting technique, we train shallow/weak decision trees (stump decision trees). In this method the accuracy gets improved with each iteration, that iteratively adjusts the weight of observation as per the last classification done by the model. If an observation is incorrectly classified,  
it increases the weight of that observation.

**ANSWER 13** – Before adjusted R2 let me explain what is R2. When we make a linear model, we try to make a line, plane or hyperplane such that all the predicted points are on hyperplane and sum of squared distance between the predicted and actual target variable data points should be minimum, that will be considered as best fit line, hyperplane to our model. Sum of squared distance between the predicted and actual target variable data points are called sum of residual error.  
Then find sum of average total by taking the square distance of each data from the mean (average of target) parallel to x-axis.   
Now R2 = 1 - [ Var(mean) – Var(best fit line) ] / Var(mean) this will make R2 ranging from 0-1 where , if the R2 value is near to 1 then it means that the variables are correlated and showing high variance.  
As the number of independent features increases the R2 value will increase whether the independent variable is correlated to target variable or not.   
Adjusted R2 solve this problem. Formula for adjusted R2 is : 1 – [(1-R2) (1-N)/N-P-1] where N is sample size, P is number of independent variables. Adjusted R2 tries to find out the features that are not correlated with the output. If the adjusted R2 value is close to 1 that means the features are correlated and when the variables are not correlated the adjusted R2 value will be less.

**ANSWER 14** – Normalization means transforming the data points in such a way that they should scale n a range of 0-1, where as standardisation means scaling of data in such a way that the mean value should be 0 and standard deviation should be 1. Standardisation means mean centring of data with standard deviation of 1.

**ANSWER 15** – Cross validation is a technique of training a model using random subset of training data so that the model result will not be biased. Cross validation can be used to detect overfitting, however for a perfect cross validation the compute time is more, it is computationally expensive.