**Q1. Which of the following method is best suited to detect outliers in a n-dimensional space, where n > 1:**

A. Normal Probability plot

B. Boxplot

C. Mahalonobis distance

D. Scatter Plot

**Solution: C**

Mahalanobis’ distance is a statistical measure of the extent to which cases are multivariate outliers, based on a chi-squared distribution. For more detail refer this[link](http://eurekastatistics.com/using-mahalanobis-distance-to-find-outliers/).

**Q2. Logistics regression differs from multiple regression analysis in following ways ?**

A. It is specifically designed to predict the probability of an event

B. The goodness-of-fit indices

C. In the estimation of regression coefficients

D. All of the above

**Solution: D**

**A:** Logistic regression designed for classification problem where we can calculate what is the probabilities of an event occurring.

**B:** A goodness-of-fit test, in general, refers to measuring how well do the observed data corresponds to the fitted (assumed) model. We use logistic regression as a way of checking the model fit.

**C:** After fitting the logistic regression model, we can also observe the relationship (positive and negative relation) between the independent features to target using their coefficients.

**Q3. What does it mean to bootstrap data?**

A. To sample m features with replacement from the total M.

B. To sample m features without replacement from the total M.

C. To sample n examples with replacement from the total N.

D. To sample n examples without replacement from the total N.

**Solution: C**

If we don’t have enough data to train our algorithm then we can increase the size of our training set by randomly selecting items and duplicating them (with replacement).

**Q4. “Overfitting is a challenge with supervised learning, but not with unsupervised learning.” Is the above statement True or False?**

A. True

B. False

**Solution: B**

We can evaluate an unsupervised machine learning algorithm with the help of unsupervised metrics. For example, we can evaluate the clustering model using “adjusted rand score”.

**Q5. Which of the following are true with regards to choosing “k” in k-fold cross validation?**

A. Higher value of “k” is not always better, choosing higher k may slow your process to evaluate your results.

B. Choosing a higher value of “k” leads to a lower bias towards true expected error (as training folds will become similar to the total dataset).

C. Select a value of k which always minimizes the variance in CV.

D. All of the above

**Solution: D**  
Larger k value means less bias towards overestimating the true expected error (as training folds will be closer to the total dataset) and higher running time (as you are getting closer to the limit case: Leave-One-Out CV). We also need to consider the variance between the k folds accuracy while selecting the k.

**Q6. A regression model is suffering with multicollinearity. How will you deal with this situation without losing much information?**

1. Remove both collinear variables.  
2. Instead of removing both variables, we can remove only one variable.  
3. We can calculate VIF (variance inflation factor) to check the presence of multicollinearity and take action accordingly.  
4. Removing correlated variables might lead to loss of information. In order to retain those variables, we can use penalized regression models like ridge or lasso regression.

**Which of the above statements are true?**

A. 1

B. 2

C. 2 and 3

D. 2,3 and 4

**Solution: D**

To check multicollinearity, we can create a correlation matrix to identify & remove variables having correlation above 75% (deciding a threshold is subjective). In addition, we can use calculate VIF (variance inflation factor) to check the presence of multicollinearity. VIF value <= 4 suggests no multicollinearity whereas a value of >= 10 implies serious multicollinearity. Also, we can use tolerance as an indicator of multicollinearity.

But, removing correlated variables might lead to loss of information. In order to retain those variables, we can use penalized regression models like ridge or lasso regression. Also, we can add some random noise in correlated variable so that the variables become different from each other. But, adding noise might affect the prediction accuracy, hence this approach should be carefully used.

**Q7. After evaluation of model, it is identified that we have high bias in our model. What could be the possible ways to reduce it?**

A. Reduce the number of features in the model.

B. Add more features to the model.

C. Add more data points to the model.

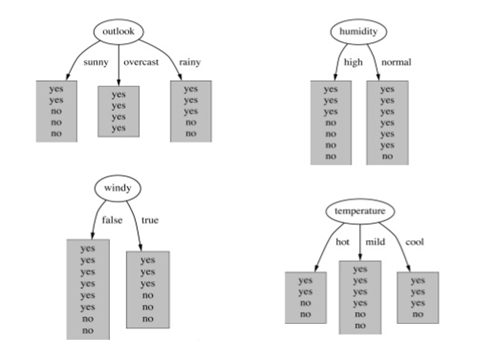
D. B and C

E. All of the above

**Solution: B**

If a model is suffering from high bias, it means that model is less complex, to make the model more robust, we can add more features in feature space. Adding data points will reduce the variance.

**Q8. While building a decision tree based model, we split a node on the attribute, which has highest information gain. In the image above, select the attribute which has the highest information gain?**

[](https://www.analyticsvidhya.com/wp-content/uploads/2016/11/Q8_Image123.png)

A. Outlook

B. Humidity

C. Windy

D. Temperature

**Solution: A**

Information gain increases with the average purity of subsets. To understand the calculation of information gain, read [here](https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/). You can also check out this[slide](http://www.slideshare.net/marinasantini1/lecture-4-decision-trees-2-entropy-information-gain-gain-ratio-55241087).

**Q9. Which of the following is a correct statement about “Information Gain”, while splitting a node in a decision tree?**

1. Less impure node requires more information to describe the population  
2. Information gain can be derived using entropy as 1-Entropy  
3. Information gain is biased towards choosing the attribute with large number of values

A. 1

B. 2

C. 2 and 3

D. All statements are correct

**Solution: C**

For better understanding read this [article](https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/) and [slide](http://www.slideshare.net/marinasantini1/lecture-4-decision-trees-2-entropy-information-gain-gain-ratio-55241087).

**Q10. An SVM model is suffering with under fitting. Which of the following steps will help to improve the model performance?**

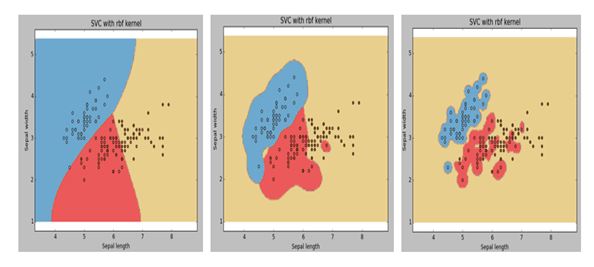
A. Increase penalty parameter “C”

B. Decrease penalty parameter

C. Decrease kernel coefficient (gamma value)

**Solution: A**  
In the case of underfitting, we need to increase the complexity of a model. When we increase the value of C,it means that we are making decision boundary more complex. Hence ‘A’ is the right answer.

**Q11. Suppose, we were plotting the visualization for different values of gamma (Kernel coefficient) in SVM algorithm. Due to some reason, we forgot to tag the gamma values with visualizations. In that case, which of the following option best explains the gamma values for the images below (1,2,3 left to right, so gamma values are g1 for image1, g2 for image2 and g3 for image3 ) in case of rbf kernel.**

[](https://www.analyticsvidhya.com/wp-content/uploads/2016/11/Q11_Image2.jpg)

A. g1 > g2 > g3

B. g1 = g2 = g3

C. g1 < g2 < g3

D. g1 >= g2 >= g3

E. g1 <= g2 <= g3

**Solution: C**  
Higher the value of gamma, will try to exact fit the as per training data set i.e. generalization error and cause over-fitting problem. Hence ‘C’ is the right answer.

**Q12. We are solving a classification problem (Binary class prediction) and predicting the probabilities of classes instead of actual outcome (0,1). Now suppose, I have taken the model probabilities and applied a threshold of 0.5 to predict the actual class (0,1). Probabilities more than or equals to 0.5 will be consider as positive class (say 1) and below 0.5 will be considered negative class (say 0). Next, if I use a different threshold, which is higher than 0.5 for classification of positive and negative class, what is the most appropriate answer below you can think of?**

1. The classification will have lower or same recall after increasing threshold.  
2. The classification will have higher recall after increasing threshold.  
3. The classification will have higher or same precision after increasing threshold.  
4. The classification will have lower precision after increasing threshold.

A. 1

B. 2

C. 1 and 3

D. 2 and 4

E. None of the above

**Solution: C**

To understand the impact on precision and recall rate after changing the probability threshold check out this [article](http://nerds.airbnb.com/confidence-splitting-criterions/).

**Q13. “Click through rate” prediction is a problem with imbalanced classses (say 99% negative class and 1% positive class in our training data). Suppose, we were building a model on such imbalanced data and we found our training accuracy is 99%. What could be the conclusion?**

A. Model accuracy is very high we don’t need to do anything further.

B. Model accuracy is not good and we should try to build a better model

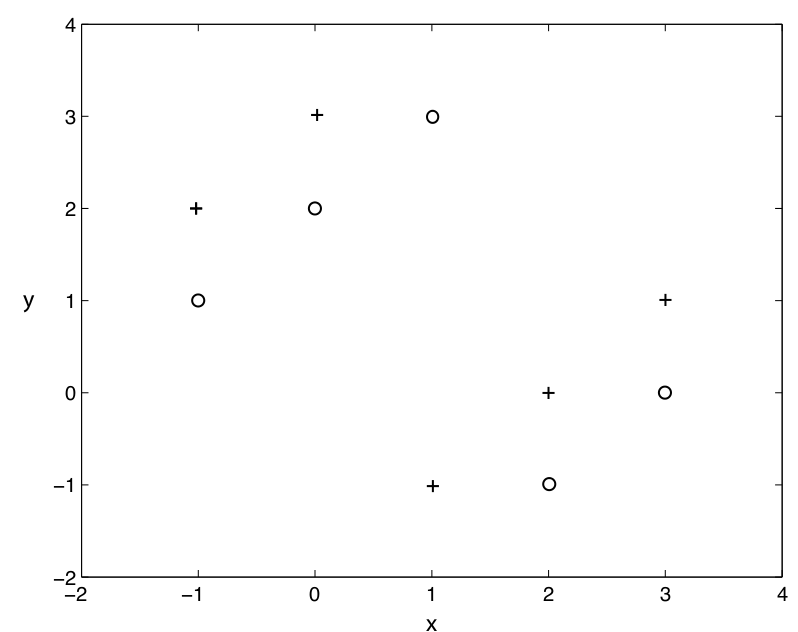
C. Cannot say anything about the model

D. None of these

**Solution: B**

In an imbalanced data set, accuracy should not be used as a measure of performance because 99% (as given) might only be predicting majority class correctly, but our class of interest is minority class (1%). Hence, in order to evaluate model performance, we should use Sensitivity (True Positive Rate), Specificity (True Negative Rate), F measure to determine class wise performance of the classifier. If the minority class performance is found to to be poor, we take the necessary steps. For more about to deal with imbalance class problem, you can refer this [article](https://www.analyticsvidhya.com/blog/2016/03/practical-guide-deal-imbalanced-classification-problems/).

**Q14. Let’s say we are training a model using kNN where training data has less number of observations (below is a snapshot of training data with two attributes x, y and two labels as “+” and “o”). Now for k=1 in kNN, what would be the leave one out cross validation error?**

[](https://www.analyticsvidhya.com/wp-content/uploads/2016/11/14-2.png)

A. 0%

B. 100%

C. Between 0 to 100%

D. None of the above

**Solution: B**  
In Leave-One-Out cross validation, we will select (n-1) observations for training and 1 observation of validation. Consider each point as a cross validation point and then find the nearest point to this point which will always give the opposite class. Hence for each observation will get misclassified which means that we will get the 100% error.

**Q15. We want to train a decision tree on a large dataset. What options could you consider for building a model which will take less time for training?**

1. Increase the depth of tree.  
2. Increase learning rate.  
3. Decrease the depth of tree.  
4. Decrease the number of tree.

A. 2

B. 1 and 2

C. 3

D. 3 and 4

E. 2 and 3

F. 2, 3 and 4

**Solution: C**

If remaining parameters are fixed for decision tree then we can conclude following options:

* Increase the depth of the tree will cause all nodes are expanded until all leaves are pure. Hence increase the depth will take more training time.
* Learning rate, it is not a valid parameter to tune with single decision tree
* In decision tree we build only one tree.

**Q16. Which of the following options are true regarding neural network?**

1. Increasing the number of layers may increase the classification error in test data.  
2. Decreasing the number of layers always decreases the classification error in test data.  
3. Increasing the number of layers always decreases the classification error in training data.

A. 1

B. 1 and 3

C. 1 and 2

D. 2

**Solution: A**  
It is generally observed that increasing the number of layers will make the model more generalized. Therefore it would perform better on both the train and test data. But this is not always true. In this [paper](https://arxiv.org/pdf/1512.03385v1.pdf), the authors have observed that deeper network (a neural network with more layers) has higher training error in comparison to a shallow network (a neural network with lesser layers).  
Therefore options 2 and 3 are not true, because the hypotheses is not “always” correct, whereas option 1 “may” be true.

**Q17. Assume we are using the primal non-linearly separable version of the SVM optimization target function. What do we need to do to guarantee that the resulting model is linearly separable?**

A. C = 1

B. C = 0

C. C = Infinite

D. None of the above

**Solution: C**

If we are using the primal non-linearly separable version of the SVM optimization target function, we need to set C = Infinite to guarantee that the resulting model is linearly separable. Hence option ‘C’ is correct.

**Q18. After training an SVM, we can discard all examples which do not support vectors and can still classify new examples?**

A. TRUE

B. FALSE

**Solution: A**  
This is true because only support vectors affect the boundary.

**Q19. Which of the following algorithm(s) can be constructed with the help of a neural network?**

1. K-NN  
2. Linear Regression  
3. Logistic Regression

A. 1 and 2

B. 2 and 3

C. 1, 2 and 3

D. None of the above

**Solution: B**  
1. K-NN is an instance-based learning algorithm and does not have any parameters to train. So it can not be constructed with the help of neural network.  
2. The simplest neural network performs least squares regression.  
3. Neural networks are somewhat related to logistic regression. Basically, we can think of logistic regression as one layer neural network.

**Q20. Please choose the datasets / problems, where we can apply Hidden Markov Model?**

A. Gene sequence datasets

B. Movie review datasets

C. Stock market price datasets

D. All of above

**Solution: D**  
All are the examples of time series dataset where Hidden Markov Model can be applied.

**Q21. We are building a ML model on a dataset with 5000 features and more than a million observations. Now, we want to train a model on this dataset but we are facing a challenge with this big size data. What are the steps will you consider to train model efficiently?**

A. We can take random samples from the dataset and will build the model on them

B. We will try to use online machine learning algorithms

C. We will apply PCA algorithm to reduce number of features

D. B and C

E. A and B

F. All of the above

**Solution: F**

Processing a high dimensional data on a limited memory machine is a strenuous task, Following are the methods you can use to tackle such situation:

* We can randomly sample the data set. This means, we can create a smaller data set, let’s say, having 1000 variables and 300000 rows and do the computations.
* Using online learning algorithms like those present in Vowpal Wabbit is a possible option.
* Also, we can use PCA and pick the components which can explain the maximum variance in the data set.

Hence all are correct.

**Q22. We want to reduce the number of features in a data set. Which of the following steps can you take to reduce features (choose most appropriate answers)?**

1. Will use Forward Selection method  
2. Will use Backward Elimination method  
3. We train our model on all features once and get the accuracy of a model on test. Now, we will take 1 feature at a time and shuffle feature values for test data set and apply prediction on shuffled test data. After taking the prediction on this we will evaluate the model. If it increases the accuracy of a model, we will remove this feature.  
4. Will look at the correlation table to remove features with high correlation

A. 1 and 2

B. 2, 3 and 4

C. 1, 2 and 4

D. All

**Solution: D**

1. Forward Selection and backward Elimination are the two main methods, we use for feature selection.
2. Instead of using these two algorithms, we can use the algorithm which is defined in 3rd option which is very useful in case of a large dataset where Forward Selection and backward Elimination algorithms fail.
3. We can also use correlation based feature selection, where we can remove multicollinear features.

Hence option D is correct.

**Q23. Please choose options which are correct in case of RandomForest and GradientBoosting Trees.**

1. In RandomForest intermediate trees are not independent of each other’s but in case on Gradient Boosting intermediate trees are independent.2. Both use
2. Both use random subset of features for creating intermediate trees.
3. We can generate parallel trees in case Gradient boosting because trees are independent of each other.  
   Gradient boosting always outperform RandomForest regardless of any data.

A. 2

B. 1 and 2

C. 1, 3 and 4

D. 2 and 4

**Solution: A**

1. Random Forest is based on bagging whereas Gradient Boosting is based on boosting. Incase of boosting intermediate trees are not independent of each other because result of previous tree is used by next tree but in case of bagging they are independent.
2. In both algorithms, we use random set of feature for generating intermediate trees.
3. Since Intermediate trees are independent of each others in Random Forest so we can generate these tree parallel but in case of Gradient Boosting it is not possible.
4. It is not true because this is depend on data.

Hence answer ‘A’ is true.

**Q24. For PCA (Principle Component Analysis) transformed features, the independence assumption of Naive Bayes would always be valid because all principal components are orthogonal and hence uncorrelated. Is this statement True or False?**

A True

B. False

**Solution: B**

This statement is false. Firstly, uncorrelation is not equivalent to independence. And secondly, transformed features are not necessarily uncorrelated.

**Q25. Which of the following statements is true about PCA?**

1. We must standardize the data before applying PCA  
2. We should select the principal components which explain the highest variance  
3. We should select the principal components which explain the lowest variance  
4. We can use PCA for visualizing the data in lower dimensions

A. 1, 2 and 4

B. 2 and 4

C. 3 and 4

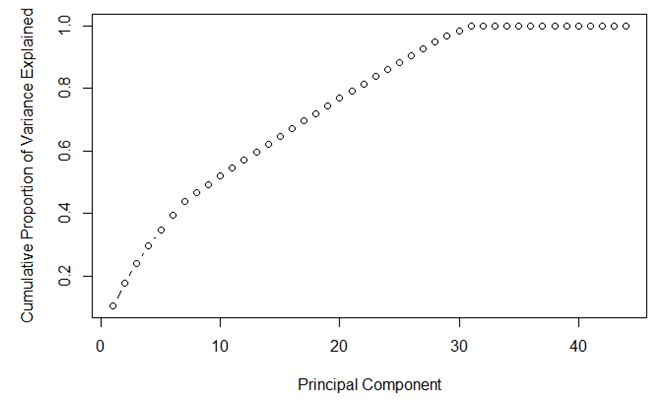
D. 1 and 3

E. 1, 3 and 4

**Solution: A**

* PCA is sensitive to scale of variables in data. So we need to standardize the data before applying PCA. For example if you change one variable from km to cm (increasing its variance), it may go from having little impact to dominating the first principal component.
* Second is true because we always select the principal component which has highest variance.
* Sometimes it is very useful to plot the data in lower dimensions. We can take the first 2 principal components and then visualize the data using scatter plot.

**Q 26: What would be the optimum number of principle components in the figure given**

[](https://www.analyticsvidhya.com/wp-content/uploads/2016/11/Q26_Image4.jpg)

A. 7

B. 30

C. 35

D. Can’t Say

**Solution: B**  
We can check in figure that the number of components = 30 is giving highest variance with lowest number of components. Hence option ‘B’ is the right answer.

**Q27. Data scientists always use multiple algorithms for prediction and they combine output of multiple machine learning algorithms (known as “Ensemble Learning”) for getting more robust or generalized output which outperform all the individual models. In which of the following options you think this is true (Choose best possible answer)?**

A. Base models having the higher correlation.

B. Base models having the lower correlation.

C. Use “Weighted average” instead of “Voting” methods of ensemble.

D. Base models coming from the same algorithm

**Solution: B**

Refer below ensemble guides to understand it in detail:

* [Basics of Ensemble Learning Explained in Simple English](https://www.analyticsvidhya.com/blog/2015/08/introduction-ensemble-learning/)
* [Kaggle Ensemble Guide](http://mlwave.com/kaggle-ensembling-guide/)
* [5 Easy questions on Ensemble Modeling everyone should know](https://www.analyticsvidhya.com/blog/2015/09/questions-ensemble-modeling/)

**Q28. How can we use clustering method in supervised machine learning challenge?**

1. We can first create the clusters and then apply supervised machine learning algorithm on different clusters separately.  
2. We can take cluster\_id as an extra feature in our features space before applying supervised machine learning algorithm.  
3. We can’t create the clusters before applying supervised machine learning.  
4. We can’t take cluster\_id as an extra feature in our features space before applying supervised machine learning algorithm.

A. 2 and 4

B. 1 and 2

C. 3 and 4

D. 1 and 3

**Solution: B**

* We can build the separate machine learning models for different clusters  and it may give a boost in prediction accuracy.
* Tacking cluster\_id as an extra feature can leads to higher accuracy results because it can summarize the data.

Hence ‘B’ is true.

**Q29. Which of the following statement(s) are correct?**

1. A machine learning model with higher accuracy will always indicate a better classifier.  
2. When we increase the complexity of a model, it will always decrease the test error.  
3. When we increase the complexity of a model, it will always decrease the train error.

A. 1

B. 2

C. 3

D. 1 and 3

**Solution: C**

* In class imbalance dataset accuracy wouldn’t be a good evaluation metric. In such cases precision and recall are the best way to evaluate the models.
* Increase the complexity of a model will lead to overfitting. In case of overfitting, training error decrease and test error increase.

**Q30. Which options is / are true regarding the GradientBoosting tree algorithm.**

1. When we increase minimum number of sample for split, we always try to get algorithm which less over-fit the data.
2. When we increase the minimum number of sample for split, it gets overfitted on data.
3. When we decrease the fraction of samples to be used for fitting the individual base learners, we always want to reduce the variance.
4. When we decrease the fraction of samples to be used for fitting the individual base learners, we always want to reduce the bias.

A. 2 and 4

B. 2 and 3

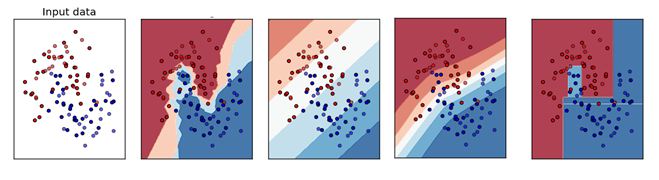
C. 1 and 3

D. 1 and 4

**Solution: C**

* Minimum number of samples (or observations), which are required in a node to be considered for splitting, are used to control overfitting. Too high values can lead to under-fitting hence, it should be tuned using CV.
* The fraction of observations to be selected for each tree is done by random sampling. Values slightly less than 1 make the model robust by reducing the variance. Typical values ~0.8 generally work fine but can be fine-tuned further.

**Q31. Which of the following is a decision boundary of KNN?**

[](https://www.analyticsvidhya.com/wp-content/uploads/2016/11/Q31_Image5.jpg)

A) B

B) A

C) D

D) C

E) Can’t Say

**Solution: B**  
The KNN algorithm classifies new observations by looking at the K nearest neighbors, looking at their labels, and assigning the majority (most popular) label to the new observation. So the decision boundaries may not be linear. Hence option ‘B’ is correct.

**Q32. If a trained machine learning model achieves 100% accuracy on test set, does this mean that the model will perform similar on a newer test set, i.e. give 100%?**

A. Yes, because now the model is general enough to be applied on any data

B. No, because there still are things which the model cannot account, such as noise

**Solution: B**  
The answer is No, Real world data will not always be noise free so in that case we won’t get 100% accuracy.

**Q33. Below are the common Cross Validation Methods:**

i. Bootstrap with replacement.

ii. Leave one out cross validation.

iii. 5 Fold cross validation.

iv. 2 repeats of 5 Fold cross validation

Arrange above four methods based on execution time required where sample size is 1000

A. i > ii > iii > iv

B. ii > iv > iii > i

C. iv > i > ii > iii

D. ii > iii > iv > i

**Solution: B**

* Bootstrapping is a statistical technique that falls under the broader heading of resampling. So there is only 1 validation set using random sampling.
* Leave-One-Out cross validation will take maximum time, because we need to train model n times(n is number of observations). If there are 1000 observations then we need to train 1000 models.
* 2 repeats of 5 Fold cross validation will have 10 models to train irrespective of number observations.
* 5 Fold cross validation will have 5 models to train irrespective of number of observations.

Hence Options B is correct option.

**Q34. Removed**

**Q35. Variable selection is intended to select the “best” subset of predictors. In case of variable selection what are things we need to check with respect to model performance?**

1. Multiple variables trying to do the same job  
2. Interpretability of the model  
3. Feature information  
4. Cross Validation

A. 1 and 4

B. 1, 2 and 3

C. 1,3 and 4

D. All of the above

**Solution: C**

* If multiple variables trying to do the same job then we will have multicollinearity.
* With respect to model performance we need not to see Interpretability of the model.
* If features are highly informative will always add value to the model.
* We need to apply cross validation to check the model generalizability.

Hence answer ‘C’ is correct.

**Q36. Which of the following statement(s) may be true post including additional variables in a linear regression model?**

1. R-Squared and Adjusted R-squared both are increase  
2. R-Squared is constant and Adjusted R-squared increase  
3. R-Squared decreases and Adjusted R-squared also decreases  
4. R-Squared decreases and Adjusted R-squared increases

A. 1 and 2

B. 1 and 3

C. 2 and 4

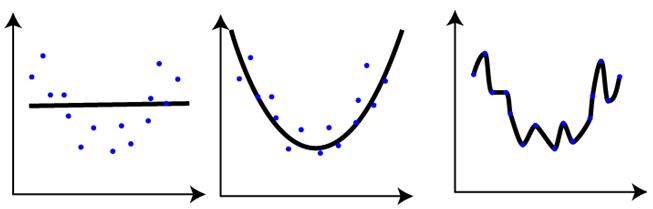
D. None of the above

**Solution: D**  
R-squared cannot determine whether the coefficient estimates and predictions are biased, which is why we must assess the residual plots. However, R-squared has additional problems that the adjusted R-squared and predicted R-squared are designed to address.  
Every time you add a predictor to a model, the R-squared increases or remains same.

The adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model. The adjusted R-squared increases only if the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected by chance.

For more detail, you can refer this [discussion.](https://discuss.analyticsvidhya.com/t/difference-between-r-square-and-adjusted-r-square/264/3)

**Q37. If we were evaluating the model performance with the help below visualizations which we have plotted for three different models for same regression problem on same training data. What do you conclude after seeing this visualization?**

[](https://www.analyticsvidhya.com/wp-content/uploads/2016/11/Q37_Image6.jpg)

1. The training error in first plot is maximum as compare to second and third plot.  
2. The best model for this regression problem is the last (third) plot because it has minimum training error (zero).  
3. The second model is more robust than first and third because it will perform best on unseen data.  
4. The third model is overfitting more as compare to first and second.  
5. All will perform same because we have not seen the testing data.

A. 1 and 3

B. 1 and 3

C. 1, 3 and 4

D. 5

**Solution: C**  
The trend in the graphs looks like a quadratic trend over independent variable X. A higher degree(Right graph) polynomial might have a very high accuracy on the train population but is expected to fail badly on test dataset. But if you see in left graph we will have training error maximum because it underfits the training data.

**Q38. What are the assumptions we need to follow while applying linear regression?**

1. It is important to check for outliers since linear regression is sensitive to outlier effects.  
2. The linear regression analysis requires all variables must have normal distribution  
3. Linear regression assumes that there is little or no multicollinearity in the data.

A. 1 and 2

B. 2 and 3

C. 1,2 and 3

D. None of these

**Solution: D**

* Outliers are the highly influential point in data that can change the slope of final regression line. So removing / treating the outliers is always important in regression analysis.
* It is useful, to understand the distribution of variables. Positive and negative skewed distribution of independent variables can impact the performance of model and converting/ transforming a highly skewed independent variable to normal will improve the model performance
* Multicollinearity occurs when your model includes multiple features that are correlated to each other. In other words, it results when you have factors that are a bit redundant. So linear regression assume that there should be low or if possible no redundancy in data.

**Q39. When we build linear models, we look at the correlation between variables. While searching for the correlation coefficient in the correlation matrices, if we found that correlation between 3 pairs of variables (Var1 and Var2 , Var2 and Var3 , Var3 and Var1) is -0.98, 0.45 and 1.23 respectively. What can we infer from this?**

1. Correlation between Var1 and Var2, it shows a high correlation  
2. Since correlation between Var1 and Var2 is very high we can consider it a case of multicollinearity so we can remove either of Var1 and Var2 from our model  
3. The correlation coefficient of 1.23 between Var3 and Var1 is not possible.

A. 1 and 3

B. 1 and 2

C. 1,2 and 3

D. 1

**Solution: C**

* Correlation between Var1 and Var2 is very high and it is negative so we can consider this as a case of multicollinearity. And also, when there are multicollinear features present in data we can remove one.
* Generally, if correlation is greater than 0.7 or less than -0.7 then we consider there is high correlation among features.
* Third option is self-explanatory, correlation coefficient fall between [-1,1]. -1 and 1 are also included in this range.

**Q40. If there is a high non-linearity & complex relationship between dependent & independent variables, a tree model is likely to outperform a classical regression method. Is this statement correct?**

A. TRUE

B. FALSE

**Solution: A**

When data is non-linear, classical regression models fail to generalize on the data, whereas tree based models generally perform better.