Q1 to Q15 are subjective answer type questions, Answer them briefly.

1. R-squared or Residual Sum of Squares (RSS) which one of these two is a better measure of goodness of fit model in regression and why?

Ans. R-squared is the proportion of variance in the dependent variable that can be explained by the independent variable.

The value of R-squared stays between 0 and 100%:

* 0% corresponds to a model that does not explain the variability of the response data around its mean. The mean of the dependent variable helps to predict the dependent variable and also the regression model.
* On the other hand, 100% corresponds to a model that explains the variability of the response variable around its mean.
* If your value of R2  is large, you have a better chance of your regression model fitting the observations.

1. What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and RSS (Residual Sum) of Squares) in regression. Also mention the equation relating these three metrics with each other.

Ans. The **sum of squares total**, denoted **SST**, is the squared differences between the observed dependent variable and its **mean**. It is a measure of the total variability of the dataset. There is another notation for the **SST**. It is **TSS** or **total sum of squares**.

If this value of **SSR** is equal to the **sum of squares total**, it means our **regression** **model** captures all the observed variability and is perfect. Once again, we have to mention that another common notation is **ESS** or **explained sum of squares**.

Thelasttermisthe**sumofsquareserror***,* or**SSE***.* Theerroristhedifferencebetweentheobservedvalueandthepredictedvalue*.*

We usually want to [minimize the error](https://365datascience.com/tutorials/statistics-tutorials/ols-assumptions/). The smaller the error, the better the estimation power of the **regression**. Finally, I should add that it is also known as **RSS** or **residual sum of squares**. Residual as in: remaining or unexplained.

1. What is the need of regularization in machine learning?

Ans. Regularization is one of the most important concepts of machine learning. It is a technique**to prevent the model from overfitting** by adding extra information to it. Sometimes the machine learning model performs well with the training data but does not perform well with the test data.

1. What is Gini–impurity index?

Ans. The Gini Impurity of a dataset is a number**between 0-0.5**, which indicates the likelihood of new, random data being misclassified if it were given a random class label according to the class distribution in the dataset. For example, say you want to build a classifier that determines if someone will default on their credit card.

1. Are unregularized decision-trees prone to overfitting? If yes, why?

Ans. In decision trees, over-fitting occurs when the tree is designed so as to perfectly fit all samples in the training data set. Thus it ends up with branches with strict rules of sparse data. Thus this effects the accuracy when predicting samples that are not part of the training set. One of the methods used to address over-fitting in decision tree is called **pruning**which is done after the initial training is complete. In pruning, you trim off the branches of the tree, i.e., remove the decision nodes starting from the leaf node such that the overall accuracy is not disturbed. This is done by segregating the actual training set into two sets: training data set, D and validation data set, V. Prepare the decision tree using the segregated training data set, D. Then continue trimming the tree accordingly to optimize the accuracy of the validation data set, V.

1. What is an ensemble technique in machine learning?

Ans. Ensemble methods are**techniques that create multiple models and then combine them to produce improved results**. Ensemble methods usually produces more accurate solutions than a single model would. This has been the case in a number of machine learning competitions, where the winning solutions used ensemble methods.

1. What is the difference between Bagging and Boosting techniques?

Ans. Bagging and Boosting: Differences. As we said already,**Bagging is a method of merging the same type of predictions. Boosting is a method of merging different types of predictions**. Bagging decreases variance, not bias, and solves over-fitting issues in a model. Boosting decreases bias, not variance.

1. What is out-of-bag error in random forests?

Ans. **Out**-**of**-**Bag** **Error** in **Random** **Forest** The **out**-**of**-**bag** **error** is the average error for each predicted outcome calculated using predictions from the trees that do not contain that data point in their respective bootstrap sample. This way, the **Random** **Forest** model is constantly being validated while being trained.

1. What is K-fold cross-validation?

Ans. k-Fold Cross-Validation Cross-validation is a**resampling procedure used to evaluate machine learning models on a limited data sample**. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into.

1. What is hyper parameter tuning in machine learning and why it is done?

Ans. Hyperparameters are the knobs or settings that can be tuned before running a training job to control the behavior of an ML algorithm. They can have a big impact on model training as it relates to training time, infrastructure resource requirements (and as a result cost), model convergence and model accuracy. It is rare that a model will perform at the level you need for production just in the first instance. To find the right solution for your business problem, often you have to go through an [iterative cycle](https://medium.com/analytics-vidhya/machine-learning-why-it-is-an-iterative-process-bf709e3b69f2). There are multiple pieces that come together to solve the intended machine learning puzzle. You may need to train and evaluate multiple models that include different data setup and algorithms, perform feature engineering a few times or even augment more data. This cycle also involves tweaking your model’s **hyperparameters**.

1. What issues can occur if we have a large learning rate in Gradient Descent?

Ans. The learning rate can seen as step size, η. As such, gradient descent is taking successive steps in the direction of the minimum. If the step size η is too large, it can (plausibly) "jump over" the minima we are trying to reach, ie. we overshoot. This can lead to**osculations around the minimum or in some cases to outright divergence**.

1. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?

Ans. **Logistic Regression is not suitable** for complex non-linear relationships between the dependent variable and independent variables. It is also not recommended for multi-class classification problems, as it can only handle binary classification.

1. Differentiate between Adaboost and Gradient Boosting.

Ans. The technique of Boosting uses various loss functions. In case of Adaptive Boosting or AdaBoost, it minimises the exponential loss function that can make the algorithm sensitive to the outliers. With Gradient Boosting, any differentiable loss function can be utilised. Gradient Boosting algorithm is more robust to outliers than AdaBoost.

1. What is bias-variance trade off in machine learning?

Ans. Bias is the simplifying assumptions made by the model to make the target function easier to approximate. Variance is the amount that the estimate of the target function will change given different training data. Trade-off istensionbetweentheerrorintroducedbythebiasandthevariance.