**Machine Learning**

1. [R-squared](https://statisticsbyjim.com/glossary/r-squared/) is a goodness-of-fit measure for linear [regression](https://statisticsbyjim.com/glossary/regression-analysis/) models. This statistic indicates the percentage of the variance in the dependent variable that the independent variables explain collectively. R-squared measures the strength of the relationship between your model and the dependent variable on a convenient 0 – 100% scale.

R-squared evaluates the scatter of the data points around the fitted regression line. It is also called the [coefficient](https://statisticsbyjim.com/glossary/regression-coefficient/) of determination, or the coefficient of multiple determination for multiple regression. For the same data set, higher R-squared values represent smaller differences between the observed data and the fitted values.

R-squared is the percentage of the dependent variable variation that a linear model explains.

{\displaystyle R^2 = \frac {\text{Variance explained by the model}}{\text{Total variance}}}

1. **TSS:** The Total SS (TSS or SST) tells you how much variation there is in the [dependent variable](https://www.statisticshowto.com/dependent-variable-definition/).  
   Total SS = Σ(Yi – mean of Y)2.  
   **Note**: Sigma (Σ) is a mathematical term for [summation](http://www.columbia.edu/itc/sipa/math/summation.html) or “adding up.” It’s telling you to add up all the possible results from the rest of the equation.

Sum of squares is a measure of how a data set varies around a central number (like the [mean](https://www.statisticshowto.com/mean/)). You might realize by the phrase that you’re summing (adding up) squares—but squares of what? You’ll sometimes see this formula:  
[ss2](https://www.statisticshowto.com/wp-content/uploads/2015/04/ss2.jpg)

**ESS:** The Explained SS tells you how much of the variation in the dependent variable your model explained.  
Explained SS = Σ(Y-Hat – mean of Y)2.

**RSS:** The residual sum of squares tells you how much of the dependent variable’s variation your model **did not explain**. It is the sum of the squared differences between the actual Y and the predicted Y:  
Residual Sum of Squares = Σ e2

1. Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.
2. Gini Impurity is a measurement used to build Decision Trees to determine how the features of a dataset should split nodes to form the tree.
3. 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.
4. Bagging is a technique for reducing prediction variance by producing additional data for training from a dataset by combining repetitions with combinations to create multi-sets of the original data. Boosting is an iterative strategy for adjusting an observation's weight based on the previous classification.
5. The out-of-bag (OOB) error is the average error for each calculated using predictions from the trees that do not contain in their respective bootstrap sample. This allows the RandomForestClassifier to be fit and validated whilst being trained.
6. 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. As such, the procedure is often called k-fold cross-validation. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross-validation.

Cross-validation is primarily used in applied machine learning to estimate the skill of a machine learning model on unseen data. That is, to use a limited sample in order to estimate how the model is expected to perform in general when used to make predictions on data not used during the training of the model.

1. Hyperparameter tuning consists of finding a set of optimal hyperparameter values for a learning algorithm while applying this optimized algorithm to any data set. That combination of hyperparameters maximizes the model's performance, minimizing a predefined loss function to produce better results with fewer errors.
2. In order for Gradient Descent to work, we must set the learning rate to an appropriate value. This parameter determines how fast or slow we will move towards the optimal weights. If the learning rate is very large we will skip the optimal solution.
3. Logistic Regression has traditionally been used as a linear classifier, i.e. when the classes can be separated in the feature space by linear boundaries. That can be remedied however if we happen to have a better idea as to the shape of the decision boundary.
4. AdaBoost is the first designed boosting algorithm with a particular loss function. On the other hand, Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem. This makes Gradient Boosting more flexible than AdaBoost.
5. In statistics and machine learning, the bias–variance tradeoff is the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters.

#### linear Kernel:

**F(x, xj) = sum( x.xj)**

Here, **x, xj** represents the data you’re trying to classify.

#### Polynomial Kernel:

**F(x, xj) = (x.xj+1)^d**

Here ‘.’ shows the **dot product** of both the values, and **d** denotes the degree.

F(x, xj) representing the **decision boundary** to separate the given classes.

### Gaussian Radial Basis Function (RBF):

 It is one of the most preferred and used kernel functions in svm. It is usually chosen for non-linear data. It helps to make proper separation when there is no prior knowledge of data.

**F(x, xj) = exp(-gamma \* ||x - xj||^2)**

The value of gamma varies from **0 to 1**. You have to manually provide the value of gamma in the code. The most preferred value for **gamma is 0.1**.