# WORKSHEET

MACHINE LEARNING – WORKSHEET 3

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

1. Give short description each of Linear, RBF, Polynomial kernels used in SVM.

Ans 1 The kernel functions are used to map the original dataset (linear/nonlinear) into a higher dimensional space with

view to making it linear dataset. Usually linear and polynomial kernels are less time consuming and provides less

accuracy than the rbf or Gaussian kernels. The k cross validation is used to divide the training set into k distinct

subsets. Then every subset is used for training and others k-1 are used for validation in the entire training phase.

This is done for the better training of the classification task.

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

Ans A residual sum of squares (RSS) is a statistical technique used to measure the amount in a data set that is

not explained by a regression model. Regression is a measurement that helps determine the strength of the

relationship between a dependent variable and a series of other changing variables or independent variables.

Whereas R Squared is a goodness-of-fit measure for linear regression models. This statistic indicates the

percentage of the variance in the dependent variable that the explain collectively

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 Total SS (TSS or SST) tells you how much variation there is in the Total SS = Σ (Yi – mean of

Y)2. The **sum of squares total**, denoted **SST**, is the squared differences between the

observed dependent variable and its **mean**.

TSS=SSR+SSE

1. What is Gini –impurity index?

Ans: - The Gini impurity measure is one of the methods used in decision tree algorithms to decide the optimal split

from a root node, and subsequent splits. Gini Impurity tells us what is the probability of misclassifying an observation.

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

Ans: - Over-fitting is the phenomenon in which the learning system tightly fits the given training data so much that it

would be inaccurate in predicting the outcomes of the untrained data. 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.

1. What is an ensemble technique in machine learning?

Ans: - Ensemblemethods are techniques that create multiple models and then combine them to produce improved results. Ensemblemethods usually produces more accurate solutions than a single model would. This has been the case in a number of machinelearning competitions, where the winning solutions used ensemblemethods.

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

Ans: -Bagging is a way to decrease the variance inthe prediction by generating additional data for training from

dataset using combinations with repetitions to produce multi-sets of the original data. Boosting is an

iterative technique which adjusts the weight of an observation based on the last classification

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

Ans: -Out-of-bag (OOB) error, also called out-of-bag estimate, is a method of measuring the

prediction error of randomforests, boosted decision trees, and other machine learning models utilizing bootstrap

aggregating (bagging) to sub-sample data samples used for training.

1. What is K-fold cross-validation?

Ans: -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: - In machinelearning, hyperparameteroptimization or tuning is the problem of choosing a set of

optimal hyperparameters for a learning algorithm. A hyperparameter is a parameter whose value is used to control

the learning process. By contrast, the values of other parameters (typically node weights) are learned.

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

Ans: -When the learningrate is too large, gradientdescentcan inadvertently increase rather than decrease

the training errors.When the learningrate is too small, training is not only slower, but may become permanently

stuck with a hightraining error.

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 is tension between the error introduced by the bias and the variance.

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

Ans: - This is a form of regression, that constrains/ regularizes or shrinks the coefficient estimates towards zero. In

other words*,*this technique discourages learning a more complex or flexible model, so as to avoid the risk of

overfitting.

1. Differentiate between Adaboost and Gradient Boosting

Ans: -Ada boost is more about 'voting weights' and gradientboosting is more

adding gradient optimization'. Adaboost doesn't overfit because it is more about 'organizing people to vote' than

'voting'. In fact, if you have a gradientboosting model, you can use it in ada boost along with other models.

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

Ans: - Logistic regression is known and used as a linear classifier. It is used to come up with a hyperplane in feature space to

separate observations that belong to a class from all the other observations that do not belong to that class. The

decision boundary is thus linear. Robust and efficient implementations are readily available (e.g. scikit-learn) to use

logistic regression as a linear classifier.