**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.

**Answer 1:-**The main difference between these kernels, for example if linear kernel is giving us good accuracy for one class and rbf is giving for other class, what factors they depend upon and information we can get from it.

The other question is about cross validation, can we perform cross validation on separate training and testing sets. Like 5 fold cross validation. And in case if cross validated training set is giving less accuracy and testing is giving high accuracy what does it means.

If we are getting 0% True positive for one class in case of multiple classes and for this class accuracy is very good. So it means our results are wrong. Why this scenario occurred in a system.

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??

**Answer 2: -** The residual sum of squares measures the amount of error remaining between the regression function and the data set. A smaller residual sum of squares figure represents a regression function. Residual sum of squares–also known as the sum of squared residuals–essentially determines how well a regression model explains or represents the data in the model.

R-squared (R2) is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a [regression](https://www.investopedia.com/terms/r/regression.asp)model. Whereas correlation explains the strength of the relationship between an independent and dependent variable, R-squared explains to what extent the variance of one variable explains the variance of the second variable. So, if the R2 of a model is 0.50, then approximately half of the observed variation can be explained by the model's inputs.

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.

Answer 3:- There are three terms we must define. The **sum of squares total**, the **sum of squares regression**, and the **sum of squares error**.

The **sum of squares total**, denoted **SST**, is the squared differences between the observed *dependent variable* and its **mean**. You can think of this as the dispersion of the observed variables around the [**mean**](https://365datascience.com/measures-central-tendency/) – [much like the **variance** in descriptive statistics](https://365datascience.com/coefficient-variation-variance-standard-deviation/)

The second term is the **sum of squares due to regression**, or **SSR**. It is the sum of the differences between the *predicted* value and the **mean** of the *dependent variable*. Think of it as a measure that describes how well our line fits the [data](https://365datascience.com/numerical-categorical-data/).

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**.

The last term is the **sum of squares error**, or **SSE**. The error is the difference between the *observed* value and the *predicted* value

We usually want to [minimize the error](https://365datascience.com/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.

Mathematically, **SST** = **SSR** + **SSE**.

1. What is Gini –impurity index?

**Answer 4:** - Gini index or Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen. ... A Gini Index of 0.5 denotes equally distributed elements into some classes

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

**Answer 5: -** Overfitting in machine learning can single-handedly ruin your models. ... or it's not properly regularized), it can end up “memorizing the noise” instead of ... does much better on the training set than on the test set, then we're likely overfitting. ... is a constrained model (i.e. you could limit the max depth of each decision tree).

1. What is an ensemble technique in machine learning?

**Answer 6: -** 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?

**Answer 7:-**  Bagging is a way to decrease the variance in the 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?

**Answer 8**:- 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

1. What is K-fold cross-validation?

**Answer 9: -** 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

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

**Answer 10:** - In **machine learning**, **hyperparameter optimization** 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?

**Answer 11**:- Deep learning neural networks are trained using the stochastic gradient descent optimization algorithm.

The learning rate is a hyperparameter that controls how much to change the model in response to the estimated error each time the model weights are updated. Choosing the learning rate is challenging as a value too small may result in a long training process that could get stuck, whereas a value too large may result in learning a sub-optimal set of weights too fast or an unstable training process.

The learning rate may be the most important hyperparameter when configuring your neural network. Therefore it is vital to know how to investigate the effects of the learning rate on model performance and to build an intuition about the dynamics of the learning rate on model behavior.

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

**Answer 12**: - **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?

**Answer 13:-** Regularisation is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting. ... The commonly used regularisation techniques are : L1 regularisation

14. Differentiate between Adaboost and Gradient Boosting

**Answer 14**: - **Adaboost** is more about 'voting weights' and **gradient boosting** is more about 'adding **gradient** optimization'. **Adaboost** doesn't overfit because it is more about 'organizing people to vote' than 'voting'. In fact, if you have a **gradient boosting** model, you can use it in **adaboost** along with other models

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

Answer 15:- **Logistic regression** has traditionally been **used** to come up with a hyperplane that separates the feature space into classes. But **if we** suspect that the decision boundary **is nonlinear we** may get better results by attempting some **nonlinear** functional forms for the **logit** function