**MACHINE-LEARNING-WORKSHEET-3**

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

Linear Kernel is used when the data is Linearly separable, that is, it can be separated using a single Line. It is one of the most common kernels to be used. It is mostly used when there are a Large number of Features in a particular Data Set.

1.Training a SVM with a Linear Kernel is Faster than with any other Kernel.

1. When training a SVM with a Linear Kernel, only the optimisation of the C Regularisation parameter is required. On the other hand, when training with other kernels, there is a need to optimise the γ parameter which means that performing a grid search will usually take more time.

Kernels in SVM classification refer to the function that is responsible for defining the decision boundaries between the classes. Apart from the classic linear kernel which assumes that the different classes are separated by a straight line, a RBF (radial basis function) kernel is used when the boundaries are hypothesized to be curve-shaped.

RBF kernel uses two main parameters, gamma and C that are related to:

**1.the decision region (how spread the region is), and 2.the penalty for misclassifying a data point respectively.**

Usually, a grid search is needed to conlcude to the values of these two parameters, especially for gamma which is very sensitive to changes. Logarithmic scale values are usually used as a search area for gamma. Grid-searching C, on the other side, usually means trying power of 10 values (0.001, 0.01, … 10, 100 etc).

Although pythoh-targeted, the following link might be useful to understand the use of RBF kernel in SVM classification:

SVC Parameters When Using RBF Kernel

**2. 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??**

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 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.

In statistics, the residual sum of squares (RSS), also known as the sum of squared residuals (SSR) or the sum of squared estimate of errors (SSE), is the sum of the squares of residuals (deviations predicted from actual empirical values of data). It is a measure of the discrepancy between the data and an estimation model. A small RSS indicates a tight fit of the model to the data. It is used as an optimality criterion in parameter selection and model selection.

In general, total sum of squares = explained sum of squares + residual sum of squares.

The smaller the residual sum of squares, the better your model fits your data; The greater the residual sum of squares, the poorer your model fits your data. A value of zero means your model is a perfect fit.

R-squared is a better measure of goodness of fit of model in regression.

Because taken alone, the RSS isn't so informative.

**3. 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.**

the total sum of squares ( TSS ), which measures how much variation there is in the observed data.

The residual sum of squares(RSS), which measures the variation in the error between the observed data and modelled values.

sum of squares (ESS), alternatively known as the model sum of squares or sum of squares due to regression

In general, total sum of squares = explained sum of squares + residual sum of squares.

**4. What is Gini –impurity index?**

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.

Gini Impurity is a measurement of the likelihood of an incorrect classification of a new instance of a random variable, if that new instance were randomly classified according to the distribution of class labels from the data set.

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

Yes,because Decision trees are prone to overfitting, especially when a tree is particularly deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions. This small sample could lead to unsound conclusions.

**6. What is an ensemble technique in machine learning?**

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.

**7. What is the difference between Bagging and Boosting techniques?**

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. If an observation was classified incorrectly, it tries to increase the weight of this observation. Boosting in general builds strong predictive models.

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

Out-of-bag (OOB) error, also called out-of-bag estimate, is a method of measuring the prediction error of random forests, boosted decision trees, and other machine learning models utilizing bootstrap aggregating (bagging) to sub-sample data samples used for training.

**9. What is K-fold cross-validation?**

Cross-validation is a statistical method used to estimate the skill of machine learning models.

It is commonly used in applied machine learning to compare and select a model for a given predictive modeling problem because it is easy to understand, easy to implement, and results in skill estimates that generally have a lower bias than other methods.

In this tutorial, you will discover a gentle introduction to the k-fold cross-validation procedure for estimating the skill of machine learning models.

After completing this tutorial, you will know:

That k-fold cross validation is a procedure used to estimate the skill of the model on new data. There are common tactics that you can use to select the value of k for your dataset. There are commonly used variations on cross-validation such as stratified and repeated that are available in scikit-learn.

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

Tuning is the process of maximizing a model's performance without overfitting or creating too high of a variance. In machine learning, this is accomplished by selecting appropriate “hyperparameters.” Hyperparameters can be thought of as the “dials” or “knobs” of a machine learning model.

Hyperparameters are important because they directly control the behaviour of the training algorithm and have a significant impact on the performance of the model is being trained. “A good choice of hyperparameters can really make an algorithm shine”.

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

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.

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

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.

In statistics and machine learning, the bias–variance tradeoff is the property of a set of predictive models whereby models with a lower bias in parameter estimation have a higher variance of the parameter estimates across samples, and vice versa. The bias–variance dilemma or bias–variance problem is the conflict in trying to simultaneously minimize these two sources of error that prevent supervised learning algorithms from generalizing beyond their training set.

The bias error is an error from erroneous assumptions in the learning algorithm. High bias can cause an algorithm to miss the relevant relations between features and target outputs (underfitting). The variance is an error from sensitivity to small fluctuations in the training set. High variance can cause an algorithm to model the random noise in the training data, rather than the intended outputs (overfitting).

**13. What is the need of regularization in machine learning?**

The word regularize means to make things regular or acceptable. This is exactly why we use it for. Regularizations are techniques used to reduce the error by fitting a function appropriately on the given training set and avoid overfitting.

**14. Differentiate between Adaboost and Gradient Boosting ?**

Adaboost is more about ‘voting weights’ and Gradient boosting is more about ‘adding gradient optimization’.

Adaboost increases the accuracy by giving more weightage to the target which is misclassified by the model. At each iteration, Adaptive boosting algorithm changes the sample distribution by modifying the weights attached to each of the instances. It increases the weights of the wrongly predicted instances and decreases the ones of the correctly predicted instances.

Gradient boosting calculates the gradient (derivative) of the Loss Function with respect to the prediction (instead of the features). Gradient boosting increases the accuracy by minimizing the Loss Function (error which is difference of actual and predicted value) and having this loss as target for the next iteration.

Gradient boosting algorithm builds first weak learner and calculates the Loss Function. It then builds a second learner to predict the loss after the first step. The step continues for third learner and then for fourth learner and so on until a certain threshold is reached.

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

Logistic regression is a generalized linear model (GLM). This is because it contains the 3 usual components which are needed to be classified as a GLM:

1) A random component (exponential family)

2) Linear Predictor

3) Link Function.

In [ ]: