MACHINE LEARNING WORKSHEET \_ 3

**Ans 1** –

**Linear Kernel:**

As the name suggests the linear kernel is used when we have a binary classification linear problem.

Formula:

K ( Xn , Xi ) = ( Xn , Xi )

**RBF Karnel:**

RBF Kernel is used when we have a binary classification non-linear problem.

Formula:

K ( Xn , Xi ) = exp (- Gamma ||Xn – Xi ||^2 +C )

**Polynomial Kernel:**

Polynomial Kernel is used when we have a binary classification non-linear problem.

Formula:

K ( Xn , Xi ) = ( Gamma ( Xn , Xi ) + r)^d

Where,

C = Cost,

Xn = Support vector data,

and r = coefficient , d = degree

**Ans 2** –

R –squared is a better measure of goodness-of –fit for linear regression.

The R-squared represents the proportion of the variance in data which is explained by the model. So the closer the R-squared to 1 , the better the fit.

The RSS is the sum of the squared distances between the actual values and the predicted values. So the value of RSS largely depends on the scale of the response variable. So RSS is difficult to interpret and not very informative.

**Ans 3 –**

RSS represents the sum squares of difference between the actual value of target variable and predicted value of target variable by the model.

RSS = Sum (Yi – Yhat )^2

ESS represents the sum of squares of difference between the predicted value of target variable and the mean of actual value of target variable.

ESS = Sum (Yhat – Ymean)^2

TSS represents the sum of squares of difference between the actual value of target variable and the mean of actual values of target variable.

TSS = Sum (Yi – Ymean)^2

Where,

Yi = actual observed value of Y

**Ans 4 –**

Gini impurity is the probability of a datapoint being classified incorrectly.

G ( Gini Impurity ) = SUM (i = 1 to C) = P(i) \* (1-P(i))

Where,

C = total classes,

P(i) = Probability of picking a datapoint with class i

**Ans 5 –**

Yes , unregularised decision trees are more prone to over-fitting.

Regularisation in decision trees is done by pruning the tree. Pruning is done by limiting the max depth of the tree. If we don’t do this the tree can continue to fit till each data point is a different leaf in a tree. By limiting the max depth of the tree, we stop the nodes splitting beyond a point and thus prevent over-fitting.

**Ans 6-**

Ensemble technique in machine learning is a technique that combines several base models in order to produce one optimal predictive model.

**Ans 7-**

**Bagging :** Bagging is also known as bootstrap aggregating sits on top of the majority voting principle. The samples are bootstrapped each time when the model is trained. When the samples are chosen, they are used to train and validate the predictions. The samples are then replaced back into the training set. The samples are selected at random. This technique is known as bagging. To sum up, base classifiers such as decision trees are fitted on random subsets of the original training set. Subsequently, the individual predictions are aggregated (voting or averaging etc.). The final results are then used as predictions. It reduces the variance of a black box estimator. Due to this the chances of overfitting is ruled out.

**Boosting:** The concept of Adaptive Boost revolves around correcting previous classifier mistakes. Each classifier gets trained on the sample set and learns to predict. The misclassification errors are then fed into the next classifier in the chain and are used to correct the mistakes until the final model predicts accurate results. When a weak-classifier misclassifies a training sample, the algorithm then uses these very samples to improve the performance of the ensemble.

**Ans 8** –

The out of bag estimate for the generalisation error is the error rate of the out-of-bag classifier on the training set in a random forest classifier. The study of this error gives empirical evidence to show that the out-of-bag estimate is as accurate as using a test set of same size as the training set,

**Ans 9 -**

The K-fold cross validation is a procedure to estimate the skill of the model on new data. It shuffles the dataset randomly and split the dataset into K groups.

**Ans 10** -

Hyperparameter tuning is the process of choosing a set of optimal hyperparameter for a learning algorithm. It is done because the choice of appropriate hyperparameters plays crucial role in the success of a model.

**Ans 11-**

A large learning rate allows the model to learn faster but we can’t arrive on a sub-optimal final set of weights using a large learning rate.

**Ans 12 –**

If a model is too simple it can have high bias and low variance. If a model is too complex it can have high variance and low bias. The bias-variance trade-off means finding a good balance between bias and variance.

**Ans 13**-

Over-fitting is a problem that occurs when a machine learning model performs well on training data but doesn’t perform well on new or unseen data. To overcome over-fitting regularisation is used.

**Ans 14 -**

ADABOOST:

* Adaboost is a special case with a particular loss function.
* Less flexible.
* Shortcoming of existing weak learners are identified by high-weight datapoints.

GRADIENT BOOSTING:

* Gradient boosting is a generic algorithm to find approximate solutions to the addictive modelling problems.
* More flexible.
* Shortcomings of existing weak learners are identified by gradients.