Machine learning worksheet-3 answer

1.ans) svm algorithm is implemented in practice using a kernel, a kernel transforms an input data space into the required form, the main objective is to segregate the given dataset in the best possible way, the distance between the either nearest points are known as the margin, the objective is to select a hyperplane with the maximum possible margin between support vectors in the given dataset

\*linear kernel: a linear kernel can be used as normal dot product any two given observations, the product between two vectors is the sum of the multiplication of each pair of input values

\*polynomial kernel: a polynomial kernel is a more generalized form of the linear kernel, the polynomial kernel can distinguish curved or nonlinear input space

\*rbf: rbf is known as radial basis function kernel, it can map an input space in infinite dimensional space

3.ans) In regression analysis, the three main types of sum of squares are the total sum of squares, explain sum of squares, residual sum of squares

\*total sum of squares

The total sum of squares is a variation of the values of a dependent variable from the sample mean of the dependent variable, essentially the total sum of squares quantifies the total variation in a sample

\*explain sum of squares

Explain sum of squares also known as regression sum of squares, a higher regression sum of squares indicates that the model does not fit the data well, it represents modeled data

\*residual sum of squares

Residual sum of squares essentially measures the variation of modeling errors, generally a lower residual sum of squares indicates the regression model can be better explain the data, while higher residual sum of squares indicates the model poorly explain the data

Tss=ssr+sse

Tss=total sum of squared, ssr=sum of square regressions, sse=sum of square residual

4.ans) gini-index or gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen, impurity means if the elements belong to a single class, then it is called pure, the degree of gini index varies between 0 and 1, where 0 denotes that all elements belong to a certain class or if there exists only one class, and 1 denotes that the elements are randomly distributed across various classes, a gini index of 0.5 denotes equally distributed elements into some classes

5.ans) the unregularized decision tree prone to overfitting, especially when a tree is a 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.ans) ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms alone, it consists of two methods they are bagging and boosting

7.ans) bagging is a way to decrease the variance in the prediction by generating additional data for training from dataset using combination 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

9.ans) 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 modelling problem because it is easy to understand ,easy to implement and results in skill estimates that generally have a lower bias than other methods, 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

10.ans) hypertuning is the problem of choosing a set of optimal hyperparameters for a learning algorithm, hypertuning control the overall behaviour of a machine learning model, the main aim is to find an optimal combination of hyperparameters that minimizes a predefined loss function to give better results

11.ans) when the learning rate is too large, gradient descent can inadvertently increase rather than decrease the training error, when the learning rate is too small, training is not only slower, but may become permanently stuck with a high training error

12.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

13.ans) regularisation is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting, commonly used regularisation techniques are lasso regression, ridge regression, elastic net regression

14.ans) 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.ans) we can only use Logistic Regression for linear classifier because it is used to come up with a hyperplane in feature space to separate observations that belong to class from all the other observation that do not belong to that class, the decision boundary is thus linear