MACHINE LEARNING

ASSIGNMENT – 5

Q1. R-squared or Residual Sum of Squares (RSS) which one of these two is a better measure of goodness of fit model in regression and why?

ANS1. Residual sum of square is a statistical tools which is used to measure the amount of variance in the data set that is not explained by the regression model itself. Instead it measures the amount of variance in the residuals or errors.The smaller the value of RSS the better line fits to the data.
But RSS is a scale variant that is the value depends on the scale of the target variables.
and R-square is scale invariant statistical model which better explain the linear regression model.
So R square provides a better measure of fit as comparison to Residual sum of square.

R2 = (TSS – RSS)/TSS
Where R2 = rsquare,TSS= total sum of square, RSS = Residual sum of square

Q2 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.

Ans2. TSS is Total sum of square.

Total variation in target variable is the sum of squares of the difference between the actual values and their mean.

$$TSS = \sum (y_i - \bar{y})^2$$

TSS or Total sum of squares gives the total variation in Y. We observe that it is very similar to the variance of Y. While the variance is the average of the squared sums of difference between actual values and data points, TSS is the total of the squared sums.

RSS is Residual sum of square. It is the square of the distance of the actual point from the regression line.RSS gives us variation in the target value which is not captured by the model.

Explained sum of squares (ESS) is basically the square of the variation of the actual value from the mean in regression model.

R2 = (TSS – RSS)/TSS

Q3. What is the need of regularization in machine learning?

Ans3. When we use regression model to train in some data there is a good chance that the model will overfit the given training data set. Regularization sort out this problem of overfitting by restricting the degree of freedom of a given equation i.e simply reducing the number of degree of a polynomials Functions by reducing their corresponding weight. To regularize the model ,a shrinkage penalty is added to the cost function.
Three methods are generally used for regularization
a) Lasso
b) Ridge
c)Elastic net

Q 4. What is Gini–impurity index?

Ans. Gini impurity is a function which determines how best the decision tree was split. The Gini Index varies between 0 and 1, where 0 represents purity of the classification and 1 denotes random distribution of elements among various classes. The Gini index works best for the categorical data and provides result in terms of success or failure and hence performs only in binary split.

Q5. Are unregularized decision-trees prone to overfitting? If yes, why?

Ans5 yes unregularized decision tree are prone to overfitting. In this case tree are free to grow deep and hence learn nearly all the training data set even from noises but fail to predict correct prediction in the test datset.

Q6. . What is an ensemble technique in machine learning?
Ans. **Ensemble technique is a machine learning method through which several model combined together with the base model to produce an effective final model. Hence ensemble learning helps to improve machine learning results by combining several model .Baiscally there are two types of ensemble technique. Base model can be any model like Decision tree classifier.**
a) Bagging
b) Boosting

Q7. What is the difference between Bagging and Boosting techniques?

Ans. Difference between Bagging and Boosting are:

Bagging : it is a homogeneous weak learners model that learn from each other independently in parallel and combine them for determining model average.
a) the simplest way of combining predictions that belongs to the same type.
b) Aim to decrease variance not bias.
c) Each model receives equal weight
d) Each model is built independently
e) Bagging tries to solve the overfitting model.
f) if the classifier is unstable (high variance) then we can apply bagging.
g) in this base model is trained parralelly.
h) Also known as Bootstarp aggregating

hi Random forest apply this method to enhance accuracy.

Boosting: it's also a homogeneous model but works different from bagging. In this model learners learn sequentially and adeptively to improve model prediction of a learning algorithms.

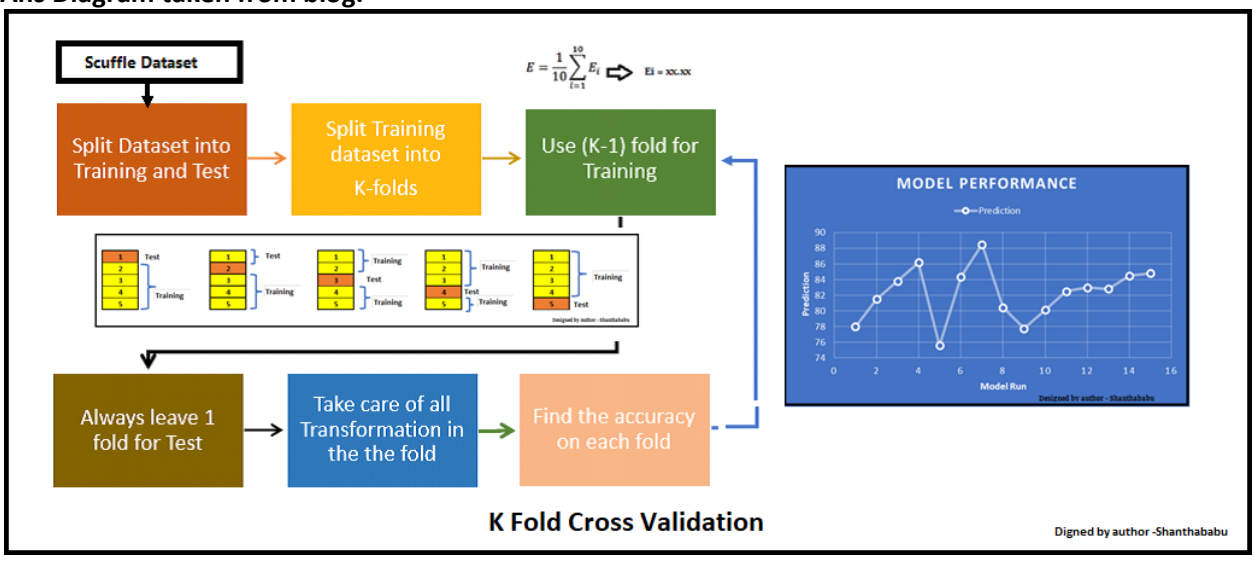
a) A way of combining predictions that belongs to the that belongs to the different types.
b) aim to decrease bias and not variance.
c) models are weighted according to there performance.
d) New models are influenced by the previous model.
e) Boosting tries to reduce bias.
f) if the classifier are simple and stable then apply boosting
g) in this base model is trained sequentially.
h) Ada boost method uses boosting technique.

Q8. What is out-of-bag error in random forests?

Ans. The prediction error on each of the bootstrap samples, which is basically a ensemble technique followed by random forest method, is known as OOB error. multiple trees are built in random forest by using Bootstrap aggregating method and the resulting predictions are averaged . during bootstrap process random resamples of variables and records are taken due to which various prediction error in each step occurs which is known as OOB error.

Q9. What is K-fold cross-validation?

Ans Diagram taken from blog.



K fold cross validation method is used in machine learning model to find out how well our machine learning model can predict the outcome of the unseen data. It is a method which is easy to comprehend and apply for mainly less number of features and also gives evaluation which is less

biased. The data is split into k-number of smaller samples and hence named as a k fold cross validation.

Steps to use K fold cross validation:

First, shuffle the dataset and split into k number of subsamples.
In the first iteration, the first subset is used as the test data while all the other subsets are considered as the training data.
Train the model with the training data and evaluate it using the test subset. Keep the evaluation score or error rate, and get rid of the model.
Now, in the next iteration, select a different subset as the test data set, and make everything else (including the test set we used in the previous iteration) part of the training data.
Re-train the model with the training data and test it using the new test data set, keep the evaluation score and discard the model.
Continue iterating the above k times. Each data subsamples will be used in each iteration until all data is considered. You will end up with a k number of evaluation scores.
The total error rate is the average of all these individual evaluation scores.

Q10. What is hyper parameter tuning in machine learning and why it is done?

Ans. Hyperparemetre is those parameter who is explicitely defined by the user to control the learning process. These parameter is selected and set by the user to tune the model so that it enhances the model performance more than its base model. Hence they are external to the model and there value cannot be changed during the training process.

Q11. What issues can occur if we have a large learning rate in Gradient Descent?

Ans Large learning rate in Gradient descent puts the model at risk to overshoot the minima and the model is not able to converge to global minima position which is the optimal position in model for predicting the result accurately.

Q12. **Differentiate between Adaboost and Gradient Boosting**.

Ans. **Ada boost**

a) Its an additive model where shortcoming of the previous model are identified by high weight data points.

b) Here the trees are usually grown as a decision stumps.

c) each classifier has different weight assigned to the final prediction based on performance.

d) It gives weight to both classifier and observations thus capturing maximum variance within data.

Gradient Boost

a) An additive model where shortcoming of the previous model are identified by the gradient.

b) the trees are generally grown to the the greater depth.

c) all classifier are weighted equally and their predictive capacity is limited with learning rate to increase accuracy.

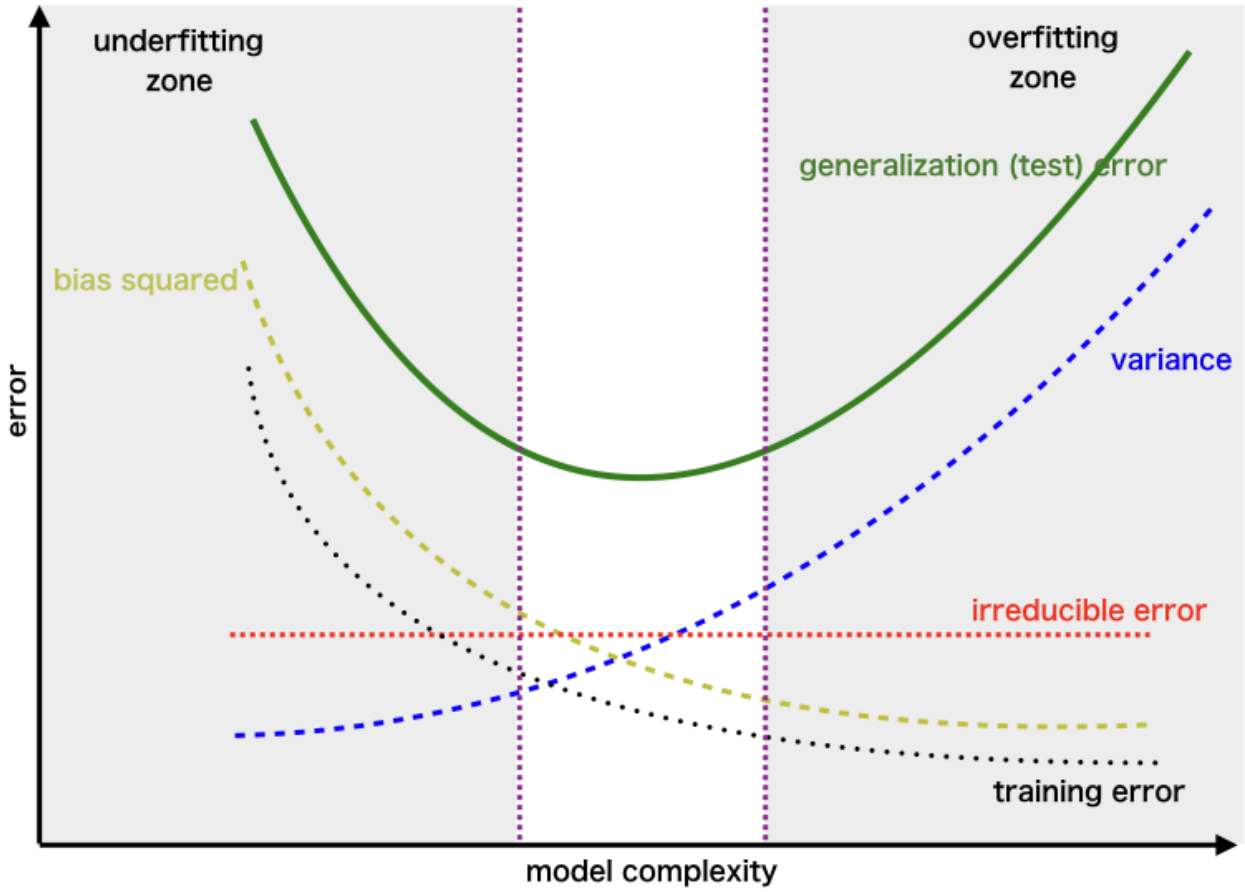
d) it builds trees on previous classifier residuals thus capturing variance in the data.

Q13. What is bias-variance trade off in machine learning?

Bias and variance are two problem that always arises during training the machine learning models.

The bias is the difference in the actual value to the predicted value in the machine learning model.so high bias provides large inaccurate result in both training and testing and training dataset. It majorly creates a underfitting problem. Wheras Variance is the variability of model prediction for agiven data point which tells us spread of our data .The model with high variance has a very complex fit and provides a overfitting probem.

So finally we have to find a sweetspot where both bias and variance are tradeoff to provide most optimistic and accurate result. This is called bias –Variance trade off.



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

Ans.

Linear kernel: Linear kernel is used when the data is linearly separable that is it can be separated by single line. It is mostly used when the data set has large number of features for example in Test classification method.

Polynomial kernel: It represents the similarity of vectors in the training set of data in a feature space over a polynomial of the original variables in the kernel

RBF: Radial basis kernel is a kernel function that is used in machine learning to find a lon linear classifier or regression line.

Q15 . Can we use Logistic Regression for classification of Non-Linear Data? If not, why?

Ans. Non linear problem can not be solved by logistic regression method because it has a linear base for making decision.