Worksheet 6

MACHINE LEARNING

In Q1 to Q5, only one option is correct, Choose the correct option:

1. In which of the following you can say that the model is overfitting?

   Ans-> (c)High R-squared value for train-set and Low R-squared value for test-set.

2 Which among the following is a disadvantage of decision trees?

Ans- > B) Decision trees are highly prone to overfitting

3 Which of the following is an ensemble technique?

Ans-> C) Random Forest

4. Suppose you are building a classification model for detection of a fatal disease where detection of the disease is most important. In this case which of the following metrics you would focus on?

Ans-> A) Accuracy

5. The value of AUC (Area under Curve) value for ROC curve of model A is 0.70 and of model B is 0.85. Which of these two models is doing better job in classification?

Ans-> B) Model B

In Q6 to Q9, more than one options are correct, Choose all the correct options:

6. Which of the following are the regularization technique in Linear Regression??

Ans-> A) Ridge ,D) Lasso

7. Which of the following is not an example of boosting technique?

Ans- > A) Adaboos , D) XGBoost

8. Which of the techniques are used for regularization of Decision Trees?

Ans-> A) Pruning , C) Restricting the max depth of the tree .

9. Which of the following statements is true regarding the Adaboost technique?

Ans -> A) We initialize the probabilities of the distribution as 1/n, where n is the number of data-points , B) A tree in the ensemble focuses more on the data points on which the previous tree was not performing well .

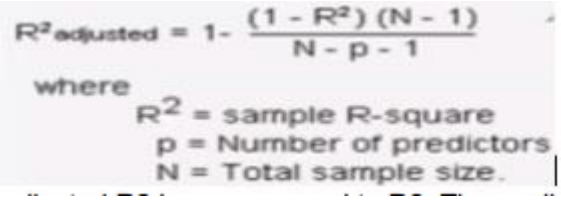
Q10 to Q15 are subjective answer type questions, Answer them briefly.

10-> Explain how does the adjusted R-squared penalize the presence of unnecessary predictors in the model?

Ans-> If you use $R2$ for multiple linear Regression means you have feature 1 , feature 2, feature 3 ,feature 4 _ _ __ and so on. so aswell do when you increases that feature to predict something what happens. as an you add more feature you tepm to get more $R2$ score . Not always but sometime there is a chance and we dont ignore it . why we take a chance.if we add more and more feature $R2$ trying to increase , bu this adjusted $R2$ will try to punish it , it

will try to pull down . it is going to punish this $R2$, its penalise it . this reason for we use got a balance 1 , This reason for we are using adjusting $R2$.

It is a metric to evaluate how well is our model . As we increase the number of independent variables in our equation ,the $R2$ Increases as well But that does'nt mean that the new independent variables have any correlation with the output variable . In other word even with the addition of new features in our model. it is not necessary that our model ,will yield better result but $R2$ value will increase . To rectify this problem, we use the adjusted $R2$ which penalises excessive use of such feature we do not coreelate with the output data .Lets understand this with an example ..
We can see that $R2$ always increase with an increasae in the number of independent variables Thus it doesnot give better picture and so we nees adjusted $R2$ value to keep this check mathematically it is calculated as.



where $R2$ = Sample R- sqaure

N - Number of record , Number of observation , Number of Rows .

P- Number of columns , Number of Feature , Independent variables.
In the equation above , when p==0 , can see the adjusted $R2$ will always be less that or equal to $R2$, and it penalise the excess of independent variables which do not affect the dependent variables .

11 -> Differentiate between Ridge and Lasso Regression

Ans->

LASSO-> If we use a LASSO or L1 Form , so this will internally try to understand relatioship between feature and labels . it come out with a relationship which feature have realtioship with target , if any of the feature does not have any kind of relationship it will make it as if there is are do not exist . it will neglect the unwanted feature . means it will give zero importance .

2 -> it acts like feature selection

3-> regulariation = λ * Σ|βj|

4-> We also called L1

5-> LASSO regression penalizes the model based on the sum of magnitude of the coeffici ent The regularization terms is given by regularization

Ridge -> This also similer to L1 form but what is ridge regression or L1 form does it will give importance but very very samll , very litter importance suppose say all important feature give 100% importance and unwanted feature only give 0.01% importance it will give very little importance if they dont have any relatioship.

2 -> Thats is major difference between LASSO and Ridge . it does not act like a feature selection tools.

3-> regularization = λ * Σ| β2j |

4-> This is also called L2

5-> Ridge Regression penalizes the model based on the sum of square of manitude of the coefficient .

12.What is VIF? What is the suitable value of a VIF for a feature to be included in a regression modelling?

Ans -> VIF (Variance inflation Factor) -> There is an chance as as when you add more and more feature , unrelevant features or features vs features relation if it has . so model might be overfitted model there is a possibility and i am not saying 100% but there is a chance . so far those reason i need to check relationship between feature vs feature and technically in machine learning world we called it as multicollinearity Problem.

There are various ways 1 of the way is Variance inflation Factor

the VIF for a regression model variable is equal to the **ratio** of the overall model variance to the variance of a model that includes only that single independent variable. This ratio is calculated for each independent variable.

13.Why do we need to scale the data before feeding it to the train the model?

Ans-> Standard Scaler - We are going to do our columns to unit less , and then we are going to standadize this data then we called a standard scaler .

Suppose we have a different different unit in features then we have to remove all unit and standardize the data..

Example -- > I have Four Friends and all friends are equally knowledgeble and one friends in very close to me .then i cannot listen him only remaining 3 are eqaully knowlegeble . i have to listen to all my four friends . i cannot biased .i have to listen all my friends .

Biased- We always try to listen one guy.

Similarly you have to give importance to all the features equal amount of impotance . we should give for that we are going to bring down this number into a similar scaler for that reason we are apply standard scaler

fit_transform()--> Use Some Formula and it will transform to the its standardized data. means it will going to transfer to new number .(Z distrubition formula is used)

Q 14 -> What are the different metrics which are used to check the goodness of fit in linear regression?

Ans- > Mean absolute error (MAE) : Represent average error(You are going to take error from every single data points , and take average)

Mean Sqaured error (MSE) :Similarly to MAE but noise is exaggerated and larger error are "punished", it is harder to interpret than MAE as it not in base units , however it is generally more popular. (There are some outliers in data . it can ignore all outliers and it will go with majority)

Root Mean Sqaured Error : Most popular metric , similar to MSE , how ever , the result is square rooted to make it more . it interpretable as it's in base units . it is recommended that RMSE be used as primary metric to interpret your model .

14. From the following confusion matrix calculate sensitivity, specificity, precision, recall and accuracy

Actual/Predicted	True	False
True	1000 TP	50 FP
False	250 FN	1200 TN

Sol -> Precision -> TP/(TP + FP)

1000/(1000 +50)= 0.95

Recall -> TP/(TP + FN)

1000/(1000+250) = 0.8

Accuracy -> (TP+TN)/(TP+TN+FP+FN)

(1000+1200)/(1000+1200+50+250) = 0.88

Specificity -> TN/(TN+FP)

1200/(1200+50) = 0.96