MACHINE LEARNING WORKSHEET 5

Q1.

Residual sum of squares is the better measures goodness of fit model in regression.

Reason:

High r square value or low square value does not show reliability of model or chosen the right regression. It might be high for bad model and vice versa.

But

RSS shows the actual distance between the actual data and predicted data. The distance is actual residual or error or loss.

The lower the sum of squared residual, the better the regression model is at explaining the data.

R SQUARED=1-(RSS/TSS)

Q2.

TSS=sum of the square the deviation(vertical distance) of actual value from the mean (given values).

RSS=Sum of the square the deviation(vertical distance) of the actual value from the predicted value.

ESS=sum of square the deviation(vertical distance) of the predicted value from the mean(given values)

TSS=RSS+ESS

Q3.
Need of regularization in machine learning are:

1. Reduce the model complexity by minimizing the independent variables or features.
2. Reduce the computation amount by shrinking the magnitude of the regression coefficient.

Q4.
Gini-impurity index measures the purity of nodes split to form decision tree.

Q5.
Yes, because when we do not regularize the decision tree means we have a lot of independent variables or features which affects the output/dependent variables. It means our model is working good for train data set but it fails for test or unseen data set.

Q6.
Ensemble technique:
Is a machine learning technique that combines several base models in order to produce one optimal predictive model.
Two types: Bagging(parallel styles) and Boosting (sequential style)

Q7.
Bagging:
Gives equal weight to each model, it means the values which are failed to be classified by the model it is not passed to the next model(models are in parallel system) and
It reduces variance not bias.

Where as

Boosting:
The new model are given new subset of train data set along the with data which are failed to be classified by previous model by increasing their weight(models are in sequential system) and
It reduces bias not variance.

Q8.
Sometimes it is not possible to select all sample from the train data set for our models so the unselected sample is knowns “out of bag samples” so
We use this sample as test datto check the performance of our final model and the error we get while checking it is knowns as “out of bag error”.

Q9.
K fold cross validation: in this we split our train data set in “k” number of group or folds and we use “k-1” folds to train the model and the remain fold it is used to test the mode and repeat the same with each fold and the last we take the average of models to make final optimal model.

Q10.
The process of choosing optimal parameter for algorithm to train your model -hyperparameter

It is done:
Maximize model performance
Minimize predefine loss function
Better results with fewer errors

Example svm.Svc()

Parameter.	value
Kernel.	rbf, linear, poly
C	integer
Gamma.	Float

Q11.
Learning rate is how aggressively or slowly we want to change step to get minima in gradient descent algorithm to make the model best fit, so if we have a

large learning rate it might chance that it skip the minima.
New value=old value -step size
Where step size= learning rate *slope.

Q12.
In logistic regression at core we use linear equation on which sigmoid is applied that convert the output ranges between 0 and 1 and based on threshold value we classify into 0 and 1 but
For non-linear data, we have continuous output.

Q13.
Ada boosting:
Weight of records get highlighted.
Stumps (two leaf nodes).
Tree not grown fully.

Where as
Gradient boosting:
Optimizing the loss.
Leaf nodes (8-32).
Tree is bigger.

Q14.
Bias -variance trade off :
Finding a balance of prediction error between test and train data

Q15. Parameter for SVM(support vector machine)=kernel and its values are

1. Linear= when the data is linearly separable in 2D(linear), 3D(plane) and above 3D(Hyperplane).
2. RBF/Gaussian kernel=when the data is not linearly separable.
   We make it linearly separable by mapping the data in infinite dimensional space.
3. Poly =when the data is not linearly separable .
   We make it linearly separable by mapping the data in higher dimensional space.