WORKSHEET- 7 MACHINE LEARNING

Ans1. D) All of the above

Ans2. A) Random forest

Ans3. B) The regularization will decrease

Ans4. C) both A & B

Ans5. A) It's an ensemble of weak learners.

Ans6. C) Both of them

Ans7. B) Bias will decrease, Variance increase

Ans8. B) model is overfitting

Ans9.)

$$Gini = 1 - \sum_j p_j^2$$

Gini = 1 - (0.40^2 + 0.60^2)

= 1 - (0.16+0.36)

= 1 – (0.52)

= 0.48

$$Entropy = -\sum_j p_j \log_2 p_j$$

Entropy = - [0.4 * log2(0.4) + 0.6 * log2(0.6)]

= - [0.4 * -1.32192809489 + 0.6 * -0.736965594166]

= 0.97

Ans10) Random forests overcome several problems with decision trees, including:

Reduction in overfitting: by averaging several trees, there is a significantly lower

riskof overfitting.

Less variance: By using multiple trees, you reduce the chance of stumbling across a classifier that doesn't perform well because of the relationship between the train and test data.

As a consequence, in almost all cases, random forests are more accurate than decisiontrees.

Ans11) Since the features have different scales, there is a chance that higher weightage is given to features with higher magnitude. This will impact the performance of the machine learning algorithm and obviously, We do not want our algorithm to be biased towards one feature. We scale our data before employing a distance based algorithm so that all the features contribute equally to the result.

Normalization is a scaling technique in which values are shifted and rescaled sothat they end up ranging between 0 and 1. It is also known as **Min-Max scaling**.

Standardization is another scaling technique where the values are centred around the mean with a unit standard deviation. This means that the mean of the attribute becomes zero and the resultant distribution has a unit standard deviation.

Ans12) Machine learning algorithms like linear regression, logistic regression, neural network, etc. that use gradient descent as an optimization technique require data to be scaled. The presence of feature value X in the formula will affect the step size of the gradient descent. The difference in ranges of features will cause different step sizes for each feature. To ensure that the gradient descent moves smoothly towards the minima and that the steps for gradient descent are updated at the same rate for all the features,we scale the data before feeding it to the model.

Ans13) When the class distribution is slightly skewed, accuracy can still be a useful metric. When the skew in the class distributions are severe, accuracy can become an

unreliable measure of model performance. The reason for this unreliability is centered around the average machine learning practitioner and the intuitions for classification accuracy. Typically, classification predictive modeling is practiced with small datasets where the class distribution is equal or very close to equal. Therefore, most

practitionersdevelop an intuition that large accuracy score (or conversely small error rate scores) aregood, and values above 90 percent are great. Achieving 90 percent classificationaccuracy, or even 99 percent classification accuracy, may be trivial on an imbalanced classification problem.This means that intuitions for classification accuracy developed on balanced class distributions will be applied and will be wrong, misleading the practitioner into thinking that a model has good or even excellent performance when it, in fact, does not.

Ans13) The F-score, also called the F1-score, is a measure of a model's accuracy on a dataset. It is used to evaluate binary classification systems, which classify examples into 'Positive' or 'Negative'. The F-score is a way of combining the precision and recallof the model, and it is defined as the harmonic mean of the model's precision and recall.The F-score is commonly used for evaluating information retrieval systems such as search engines, and also for many kinds of machine learning models, in particular in natural language processing. It is possible to adjust the F-score to give more importance to precision over recall, or vice-versa. Common adjusted F-scores are the F0.5-score and the F2-score, as well as the standard F1-score.The formula for the standard F1-score is the harmonic mean of the precision and recall. A perfect model has an F-score of 1.

$$F_1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}$$

Ans15) **Fit()** - In the fit() method, where we use the required formula and perform the calculation on the feature values of input data and fit this calculation to the transformer. For applying the fit() method we have to use .fit() in front of the transformer object.

Transform() – For changing the data we probably do transform, in the transform() method, where we apply the calculations that we have calculated in fit() to every datapoint in feature F. We have to use .transform() in front of a fit object because we transform the fit calculations.