**MACHINE LEARNING – WORKSHEET 4**

**(Solutions)**

**ANS 1)**

1. GridSearchCV()

**ANS 2)**

A) Random forest

**ANS 3)**

1. The regularization will decrease

**ANS 4)**

1. **Both A** (It regularizes the decision tree by limiting the maximum depth up to which a tree can be grown.) **&** **B** (It denotes the number of children a node can have.)

**ANS 5)**

1. It's an ensemble of weak learners.

**ANS 6)**

C) Both of them

**ANS 7)**

B) Bias will decrease, Variance increase

**ANS 8)**

1. model is overfitting

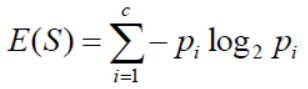
**ANS 9)**

**Gini index** (a criterion to minimize the probability of misclassification):

Formula:  ****

**Gini Index**= 1-[(40/100)^2+(60/100)^2]= **0.48**

**Entropy** (a way to measure impurity):

Formula: 

*Where, E(s) is Entropy.*

**Entropy** = -[(40/100) \* log2 (40/100)] –[(60/100) \*log2 (60/100)]= **0.9709505944546686**

**ANS 10)**

## Random forests overcome several problems with decision trees, including:

* Reduction in overfitting: by averaging several trees, there is a significantly lower risk of 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.
* Ability to handle multiple input features without need for feature deletion
* High predictive accuracy. **Decision trees** do not have same predictive accuracy
* Efficient on large datasets

As a consequence, in almost all cases, **random forests** are more accurate than **decision trees**.

**ANS 11)**

Most of the times, the dataset will contain features highly varying in magnitudes, units and range.

**Need of scaling all the numerical features:**

* Scaling is required to rescale the data and it’s used when we want features to be compared on the same scale for our algorithm. And, when all features are in the same scale, it also helps algorithms to understand the relative relationship better.
* Scaling is helpful in Distance-based algorithms and also in faster convergence

Two techniques used for scaling:

* Standard Scaler
* MinMax Scaler

**ANS 12)**

Feature scaling helps in causing **Gradient** **Descent** to converge much faster as standardizing all the variables on to the same scale. This is because θ will descend quickly on small ranges and slowly on large ranges, and so will oscillate inefficiently down to the optimum when the variables are very uneven.

**ANS 13)**

In case of a highly imbalanced dataset for a classification problem, accuracy **is not a good** metric to measure the performance of the model.

**Classification accuracy** is the **number of correct predictions divided by the total number of predictions.**

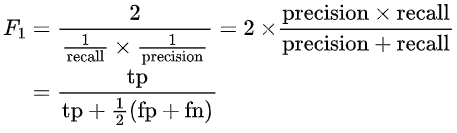
Accuracy can be misleading. For example, in a problem where there is a large class imbalance, a model can predict the value of the majority class for all predictions and achieve a high classification accuracy. So, further performance measures are needed such as **Confusion matrix and ROC-AUC Score**

**ANS 14)**

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](https://deepai.org/machine-learning-glossary-and-terms/classifier) examples into ‘positive’ or ‘negative’.

The F-score is a way of combining the [precision and recall](https://deepai.org/machine-learning-glossary-and-terms/precision-and-recall) of the model, and it is defined as the [harmonic mean](https://deepai.org/machine-learning-glossary-and-terms/harmonic-mean) of the model’s precision and recall.

## F-score Formula:

The formula for the standard F1-score is the harmonic mean of the precision and recall

Where,

**Precision**= Precision is the fraction of true positive examples among the examples that the model classified as positive. In other words, the number of true positives divided by the number of false positives plus true positives.

**Recall**= Recall, also known as sensitivity, is the fraction of examples classified as positive, among the total number of positive examples. In other words, the number of true positives divided by the number of true positives plus false negatives.

**tp**=  The number of true positives classified by the model.

**fn**= The number of false negatives classified by the model.

**fp**= The number of false positives classified by the model.

**ANS 15)**

**Fit:**

"fit" computes the mean and std (standard deviation) to be used for later scaling.

**Transform:**

"transform" uses a previously computed mean and std (standard deviation) to autoscale the data (subtract mean from all values and then divide it by std (standard deviation)).

**Fit Transform:**

"fit\_transform" does both these process at the same time.