1. D) All of the above
2. A) Random forest
3. A) The regularization will increase
4. A) It regularizes the decision tree by limiting the maximum depth up to which a tree can be grown.
5. A) It's an ensemble of weak learners. C) In case of classification problems, the prediction is made by taking the mode of the class labels predicted by the component trees.
6. C) Both of them
7. B) Bias will decrease, Variance increase
8. B) model is overfitting
9. Gini index = 1- (p(A)2 + p(B)2 ) = 1-((0.4)2+(0.6)2 ) = 0.48 Entropy = -(p(A) log2(p(A)) + p(B) log2(p(B))) = -(0.4*log2(0.4) + 0.6*log2(0.6))) = 0.97
10. The main advantage of random forest over Decision Trees is that random forest alleviates the problem of over-fitting. The decision trees are highly likely to overfit the training data if we do not regularize them by restricting the max depth of the tree, max number of leaf nodes etc. Random forest is an ensemble technique in which each tree is trained on a bootstrapped sample which is taken from the training dataset and the prediction is made by taking average of the prediction made by each tree. So, random forests are very less likely to overfit and they generalize well on unseen data as compared to decision trees.
11. Most of the times, a dataset will contain features highly varying in magnitudes, units and range. But since, most of the machine learning algorithms use Euclidean distance between two data points in their computations, this is a problem because, the results would vary greatly between different units, 5 km and 5000 m. The features with high magnitudes will weight a lot more in the distance calculations than features with low magnitudes. To avoid this effect, we use scaling which transforms all features to the same level of magnitudes. The two major scaling techniques used now a days are standardization and normalization.
12. The advantage which scaling provides in optimization using gradient descent algorithm is that if the data is scaled, the gradient descent algorithm reached the optimal solution quickly as compared to the data which is not scaled.
13. consider the case of a highly imbalanced dataset in which we have 90% data points belonging to label 1 and only 10% belonging to label 0. In that case even if we naively assign label 1 to all the data points we will get 90% accuracy because 90% data-points are of label 1 so, accuracy here will not be a good metric to evaluate the performance of a model. Even a very naive model like giving label 1 without even looking
14. f-score metric is the metric which to measure the performance of a classification model. F-score (F1 score) is the harmonic mean of precision and recall: F-Measure = (2 * Precision * Recall) / (Precision + Recall)
15. fit() : fit() is used for training. It calculates the parameters(weights) to be applied on different features in order to transform them as required by the model. transform() : transform uses the parameters(weights) learned by fit() and applies it to respective features of dataset in order to change them as per requirement of the model. fit_transform() : fit_transform() first fits() the data and then transforms the same dataset

on which fit is performed. Basically, it does both the tasks on the same dataset. fit() is generally used on training data in supervised learning where there is no need to change the training data, fit_transform is used on training data in unsupervised learning where training data must also be modified and transform() is used only on test data because we need to change the test data as per the parameters(weights) learned by training.