Machine learning

1) D) All of the above
2) A) Random forest
3) D) kernel will be changed to linear
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
6) C) Both of them
7) B) Bias will decrease, Variance increase
8) B) model is overfitting
9)
10) Random forests consist of multiple single trees each based on a random sample of the training data. They are typically more accurate than single decision trees. It reduces overfitting in decision trees and helps to improve the accuracy. It is flexible to both classification and regression problems.
11) If not scaled the features with higher value range start dominating when calculating distances. It helps in speeding up the calculations in an algorithm. Two techniques are Min-Max scaler and Standard scaler.
12) It helps in faster convergence due to stable gradients produced by the standardized variables. It dampens the oscillation when descending.
13) Accuracy is not a proper measure, since it does not distinguish between the numbers of correctly classified examples of different classes.
14) F-score is a measure of a test's accuracy. The F-score is a way of combining the precision and recall of the model, and it is defined as the harmonic mean of the model's precision and recall.

$$F_1 = 2 \cdot \frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} = \frac{\mathrm{TP}}{\mathrm{TP} + \frac{1}{2}(\mathrm{FP} + \mathrm{FN})}$$

15) fit() function calculates the values of these parameters.
transform() function applies the values of the parameters on the actual data and gives the normalized value.
fit_transform() function performs both in the same step