Q1-In which of the following you can say that the model is overfitting?
Answer- c)High R-squared value for train-set and Low R-squared value for test-set.

Q2- . Which among the following is a disadvantage of decision trees?
Answer-B) Decision trees are highly prone to overfitting.

Q3- Which of the following is an ensemble technique?
Answer-C) Random Forest

Q4- Suppose you are building a classification model for detection of a fatal disease where detection of the disease is most important. In this case which of the following metrics you would focus on?
Answer- B) Sensitivity

Q5- The value of AUC (Area under Curve) value for ROC curve of model A is 0.70 and of model B is 0.85. Which of these two models is doing better job in classification?
Answer- B) Model B

In Q6 to Q9, more than one options are correct, Choose all the correct options

Q6- Which of the following are the regularization technique in Linear Regression?
Answer- A Ridge) & D) Lasso

Q7- Which of the following is not an example of boosting technique?
Answer- B) Decision Tree & C) Random Forest

Q8- Which of the techniques are used for regularization of Decision Trees?
Answer- A) Pruning & C) Restricting the max depth of the tree

Q10 to Q15 are subjective answer type questions, Answer them briefly.

Q10-Explain how does the adjusted R-squared penalize the presence of unnecessary predictors in the model?
Answer- As we increase the number of independent variables in our equation, the R2 increases as well. But that doesn't mean that the new independentvariables have any correlation with the output variable. In other words, even with the addition of new features in our model, it is not necessary that our model will yield better results but R2 value will increase. **To rectify this problem, we use Adjusted R2 value which penalises excessive use of such features which do not correlate with the output data**.
Let's understand this with an example:
We can see that R2 always increases with an increase in the number of independent variables. Thus, it doesn't give a better picture and so we need Adjusted R2 value to keep this in check. Mathematically, it is calculated as:

R2 ADJUSTED=(1-R2)(n-1)/ N-P-1

Where,
R2=Sample R Squared,
P =Number of Predictors
N=Total Sample Size

In the equation above, when p = 0, we can see that adjusted R2 becomes equal to R2. Thus, adjusted R2 will always be less than or equal to R2,and it penalises the excess of independent variables which do not affect the dependent variable

Q11- Differentiate between Ridge and Lasso Regression.
Answer -**Ridge (L2 Form) regression shrinks the coefficients for those predictors which contribute very less in the model but have huge weights, very close to zero. But it never makes them exactly zero**. Thus, the final model will still

contain all those predictors, though with less weights. This doesn't help in interpreting the model very well. This is where Lasso **LASSO(Least Absolute Shrinkage and Selection Operator) Regression (L1 Form) regression differs with Ridge regression. In Lasso, the L1 penalty does reduce some coefficients exactly to zero when we use a sufficiently large tuning parameter λ. So, in addition to regularizing, lasso also performs feature selection.**

Q12- What is VIF? What is the suitable value of a VIF for a feature to be included in a regression modeling?
Answer- **Variance inflation factor (VIF) is a measure of the amount of multicollinearity in a set of multiple regression variables**.It measures how much the behavior (variance) of an independent variable is influenced, or inflated, by its interaction/correlation with the other independent variables. Variance inflation factors allow a quick measure of how much a variable is contributing to the standard error in the regression.
Small VIF values, VIF < 3, indicate low correlation among variables under ideal conditions. The default VIF cutoff value is 5; only variables with a VIF less than 5 will be included in the model. However, note that many sources say that a VIF of less than 10 is acceptable

---

Q13- Why do we need to scale the data before feeding it to the train the model?
Answer- If an algorithm uses gradient descent, then 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. Having features on a similar scale will help the gradient descent converge more quickly towards the minima.
Distance-based algorithms like KNN, K-means, and SVM are most affected by the range of features. This is because behind the scenes they are using distances between data points to determine their similarity and hence perform the task at hand. Therefore, we scale our data before employing a distance-based algorithm so that all the features contribute equally to the result.

Q14- What are the different metrics which are used to check the goodness of fit in linear regression?
Answer- There are three error metrics that are commonly used for evaluating and reporting the performance of a regression model

- **Mean absolute error (MAE) :** Represents average error
- **Mean squared error (MSE) :** Similar to MAE but noise is exaggerated and larger errors are "punished". It is harder to interpret than MAE as it's not in base units, however, it is generally more popular.
- **Root mean squared error (RMSE):** Most popular metric, similar to MSE, however, the result is square rooted to make it more interpretable as it's in base units. It is recommended that RMSE be used as the primary metric to interpret your model

---

Q15-From the following confusion matrix calculate sensitivity, specificity, precision, recall and accuracy.

Actual/Predicted	True	False
True	1000	50
False	250	1200

TP=1000,
FP=50
FN=250
TN=1200

1. **Sensitivity / Recall =(TP/(TP+FN))**
   = 1000/(1000+250)
   = 0.8

2. **Specificity =(TN / TN + FP)**
   =1200/(1200+50)
   =0.96

3. **Precision = TP / TP + FP**

= 1000/(1000+50)

= 0.95

4. **Accuracy = (TP + TN) / (TP + TN + FP + FN)**

=(1000+1200)/(1000+50+250+1200)

= 2200/2500

= 0.88