MACHINE LEARNING- Solved

1. C) High R-squared value for train-set and Low R-squared value for test-set.
2. B) Decision trees are highly prone to overfitting.
3. C) Random Forest
4. B) Sensitivity
5. B) Model B
6. The regularization techniques in linear regression are A)Ridge and D)Lasso
7. C) Random Forest & B) Decision Tree
8. A) Pruning & C) Restricting the max depth of the tree
9. B) A tree in the ensemble focuses more on the data points on which the previous tree was not
10. The adjusted R-squared penalizes the inclusion of unnecessary predictors in the model by adjusting the R-squared value for the number of predictors. As the number of predictors in a model increases, the R-squared value may also increase even if the predictors are not truly related to the outcome variable. The adjusted R-squared, on the other hand, will decrease if additional predictors do not improve the model significantly, as it takes into account the number of predictors used in the model.
    In other words, the adjusted R-squared penalizes the addition of predictors that do not improve the fit of the model beyond what would be expected by chance. It provides a more accurate estimate of the true explanatory power of the model by taking into account the number of predictors used, and can be used to compare models with different numbers of predictors.
11. Ridge regression and Lasso regression are both types of linear regression with regularization, which is a technique used to prevent overfitting and improve the generalization of the model. The main difference between Ridge and Lasso regression lies in the type of regularization used.
    Ridge regression adds a penalty term to the sum of squared coefficients in the regression equation, which is called L2 regularization. The penalty term is proportional to the square of the coefficients, and it shrinks their values towards zero. This means that Ridge regression can reduce the impact of less important predictors on the model, but it cannot completely eliminate them. As a result, Ridge regression tends to perform well when all predictors in the model are relevant.
    Lasso regression, on the other hand, also adds a penalty term to the sum of squared coefficients, but it uses L1 regularization instead of L2. Lasso regression shrinks the coefficients as well, but it has the added effect of driving some of them all the way to zero, effectively removing them from the model. This makes Lasso regression useful when there are many predictors in the model, some of which may be irrelevant or redundant.
    In summary, Ridge regression is good when all predictors are important, but their coefficients may need to be shrunk to reduce overfitting. Lasso regression is useful when there are many predictors, and some of them may not be useful or may be redundant, so it can eliminate them from the model.
12. VIF (Variance Inflation Factor) is a measure of multicollinearity in a multiple regression model. It quantifies the severity of multicollinearity in the set of predictor variables by

examining how much the variance of the estimated regression coefficient for each predictor is increased due to multicollinearity.

The VIF for each predictor is computed as the ratio of the variance of the model's coefficient estimate for that predictor, divided by the variance of that predictor's sampling distribution assuming that the other predictors are uncorrelated with it. A VIF value of 1 indicates no multicollinearity, whereas higher values indicate increasing levels of multicollinearity.

There is no specific threshold value of VIF that would necessarily require the removal of a variable from the regression model. However, some common rules of thumb suggest that a VIF value of 5 or higher is indicative of significant multicollinearity, and values above 10 are considered very high. In general, variables with higher VIF values may need to be examined more closely to determine whether they are contributing to the model in a meaningful way or whether they may be redundant with other variables.

13. Scaling the data is an important pre-processing step in many machine learning algorithms because it can improve the performance and stability of the model. Here are some of the reasons why scaling is necessary:
    a. Different units: The values of features can be measured in different units, such as weight in pounds and height in inches, which can lead to issues in the distance calculations. Scaling transforms the values of all features to the same unit, making the distance calculations more reliable.
    b. Outliers: Features that contain outliers can affect the performance of many machine learning algorithms. Scaling can help reduce the influence of outliers by shrinking the range of the data.
    c. Gradient Descent: Gradient descent is an iterative optimization algorithm that is used to minimize the cost function of a machine learning model. Gradient descent works faster and converges more quickly when the data is scaled.
    d. Regularization: Many regularization techniques, such as L1 and L2 regularization, involve adding a penalty term to the cost function of the model. The magnitude of this penalty can be influenced by the scale of the features, so scaling can help to ensure that the regularization is applied fairly to all features.

    The suitable value of a VIF for a feature to be included in a regression modeling depends on the specific use case and the threshold set by the practitioner or the team. However, a common rule of thumb is that a VIF of 10 or higher indicates that the feature may be highly collinear with other features and should be further investigated or removed from the model.
14. There are several metrics that can be used to check the goodness of fit in linear regression. Here are a few common ones:
    a. R-squared (coefficient of determination): This metric measures the proportion of variance in the dependent variable that is explained by the independent variables. R-squared ranges from 0 to 1, with higher values indicating a better fit.
    b. Mean squared error (MSE): This metric measures the average squared difference between the predicted values and the actual values of the dependent variable. Lower values of MSE indicate a better fit.
    c. Root mean squared error (RMSE): This is the square root of the MSE and is measured in the same units as the dependent variable. RMSE is a commonly used metric for evaluating the performance of regression models.

d. Mean absolute error (MAE): This metric measures the average absolute difference between the predicted values and the actual values of the dependent variable. Lower values of MAE indicate a better fit.
e. Residual plots: These plots can be used to visualize the pattern of residuals (the differences between the predicted values and the actual values of the dependent variable) and identify any patterns or outliers. A good fit is indicated when the residuals are randomly distributed around 0, with no obvious patterns or outliers.

It's important to note that different metrics may be more appropriate for different types of problems or data, and no single metric is perfect for all situations. It's also a good idea to use multiple metrics to evaluate the performance of a model.

15. Based on this confusion matrix, we can calculate the following metrics:
    a. Sensitivity: the proportion of actual positives that are correctly identified by the model
    b. Sensitivity = true positives / (true positives + false negatives) = 1000 / (1000 + 250) = 0.8
    c. Specificity: the proportion of actual negatives that are correctly identified by the model
    d. Specificity = true negatives / (true negatives + false positives) = 1200 / (1200 + 50) = 0.96
    e. Precision: the proportion of predicted positives that are actually positive
    f. Precision = true positives / (true positives + false positives) = 1000 / (1000 + 50) = 0.9524
    g. Recall: the proportion of actual positives that are correctly identified by the model
    h. Recall = true positives / (true positives + false negatives) = 1000 / (1000 + 250) = 0.8
    i. Accuracy: the proportion of all predictions that are correct
    j. Accuracy = (true positives + true negatives) / (true positives + true negatives + false positives + false negatives) = (1000 + 1200) / (1000 + 1200 + 250 + 50) = 0.896

    Therefore, based on the given confusion matrix,

    sensitivity is 0.8,

    specificity is 0.96,

    precision is 0.9524,

    recall is 0.8,

    and accuracy is 0.896.