SOLUTIONS MACHINE LEARNING WORKSHEET-6

1) C
2) B
3) C
4) C
5) B
6) A and D
7) B
8) D
9) A

10) The adjusted R-squared is a modification of the R-squared metric, which is a statistical measure of how well the regression model fits the data. While R-squared explains the percentage of variation in the dependent variable that is explained by the independent variables in the model, the adjusted R-squared provides a more accurate measure of model fit by penalizing the presence of unnecessary predictors.

    The adjusted R-squared takes into account the number of predictors in the model and adjusts the R-squared value based on the degrees of freedom. The degrees of freedom are calculated as the total number of observations minus the number of predictors in the model. The formula for the adjusted R-squared is:

    Adjusted R-squared = 1 - [(1 - R-squared) * (n - 1) / (n - k - 1)]

    where n is the number of observations and k is the number of predictors in the model.

    As k, the number of predictors in the model, increases, the denominator of the adjusted R-squared formula increases, resulting in a smaller adjusted R-squared value. This means that adding unnecessary predictors to the model will decrease the adjusted R-squared, indicating that the model fit is worse. In other words, the adjusted R-squared penalizes the presence of unnecessary predictors in the model by reducing the R-squared value.

11) Ridge and Lasso regression are two commonly used regularization techniques in linear regression. Both of these techniques are used to handle overfitting by introducing a penalty term to the cost function. The penalty term limits the magnitude of the coefficients of the predictors. **The key difference between Ridge and Lasso regression lies in the type of penalty used.**

Ridge Regression:

- Ridge regression adds a penalty term to the cost function, which is the sum of squared coefficients multiplied by a hyperparameter λ (lambda).
- The objective function of Ridge Regression is:
- minimize (sum of squared residuals) + (lambda * sum of squared coefficients)
- The L2-norm regularization used by Ridge regression shrinks the coefficient values towards zero, but never completely eliminates any of them.
- Ridge regression is used when there is a possibility of multicollinearity in the data, where the predictor variables are highly correlated. By reducing the coefficients of these variables, Ridge regression reduces the impact of multicollinearity on the model.

Lasso Regression:

- Lasso regression also adds a penalty term to the cost function, which is the sum of absolute coefficients multiplied by a hyperparameter λ (lambda).
- The objective function of Lasso Regression is:
  minimize (sum of squared residuals) + (lambda * sum of absolute coefficients)
- The L1-norm regularization used by Lasso regression can completely eliminate some coefficients, resulting in a sparse model where some predictors are given zero coefficients.
- Lasso regression is used when the goal is to select the most important predictors out of a large set of potential predictors. The zero coefficients in the model can be interpreted as the predictors that were not selected by the algorithm.
- In summary, Ridge regression is used to reduce the impact of multicollinearity in the data, while Lasso regression is used for feature selection by producing a sparse model with some coefficients set to zero.

12) **VIF stands for Variance Inflation Factor**. It is a measure of the extent to which a predictor variable is correlated with other predictor variables in a regression model. VIF quantifies the degree to which the variance of the estimated regression coefficient is increased due to the presence of correlated predictors.
The VIF value ranges from 1 upwards, with a value of 1 indicating no correlation between the predictor variable and the other variables, and higher values indicating more severe collinearity. As a general rule of thumb, a VIF value of less than 5 indicates low collinearity, while values above 10 indicate high collinearity. When building a regression model, it is generally desirable to include only those predictor variables that have low VIF values, as high VIF values can lead to unreliable and unstable estimates of the regression coefficients.

13) Scaling the data before feeding it to the model is an important preprocessing step in many machine learning algorithms. There are several reasons why scaling is necessary:
    1. **Preventing bias**: If the input features have different scales, then the features with larger scales can dominate the learning process, leading to bias in the model. Scaling ensures that each feature contributes equally to the learning process.
    2. **Accelerating convergence**: Scaling can speed up the learning process by making the optimization algorithm converge faster.
    3. **Improving accuracy**: Scaling can improve the accuracy of the model by reducing the noise and increasing the signal-to-noise ratio.
    4. **Handling outliers**: Scaling can also help in handling outliers, as the outliers in the unscaled features can dominate the learning process.
    5. **Ensuring compatibility**: Some machine learning algorithms, such as K-nearest neighbors (KNN) and support vector machines (SVM), require scaling to ensure compatibility between different features.

Overall, scaling the data can improve the performance and accuracy of the machine learning model, and is an important step in many machine learning workflows.

14) There are several metrics that can be used to check the goodness of fit in linear regression. Here are some of the most commonly used metrics:
    1. **R-squared (R^2)**: This metric is the proportion of the variance in the dependent variable that is explained by the independent variable(s). R-squared ranges from 0 to 1, with higher values indicating a better fit.
    2. **Mean Squared Error (MSE)**: This metric measures the average squared difference between the predicted and actual values. Lower values of MSE indicate a better fit.
    3. **Root Mean Squared Error (RMSE)**: This metric is the square root of the MSE and is often used to provide a more easily interpretable estimate of the average difference between the predicted and actual values.
    4. **Mean Absolute Error (MAE)**: This metric measures the average absolute difference between the predicted and actual values. Like MSE and RMSE, lower values of MAE indicate a better fit.
    5. **Adjusted R-squared**: This metric is similar to R-squared but takes into account the number of independent variables in the model. It penalizes the addition of unnecessary independent variables and is often used to compare models with different numbers of independent variables.

15) Using the standard definitions:

True Positive (TP): predicted positive and actually positive
False Positive (FP): predicted positive but actually negative

True Negative (TN): predicted negative and actually negative

False Negative (FN): predicted negative but actually positive

The confusion matrix can be represented as follows:

Actual/Predicted	*True*	*False*
True	1000 (TP)	50 (FP)
False	250 (FN)	120 (TN)

Sensitivity (also known as recall or true positive rate): TP / (TP + FN) = 1000 / (1000 + 250) = 0.8 or 80%
Specificity: TN / (TN + FP) = 120 / (120 + 50) = 0.706 or 70.6%
Precision: TP / (TP + FP) = 1000 / (1000 + 50) = 0.952 or 95.2%
Recall (same as sensitivity): TP / (TP + FN) = 1000 / (1000 + 250) = 0.8 or 80%
Accuracy: (TP + TN) / (TP + TN + FP + FN) = (1000 + 120) / (1000 + 120 + 50 + 250) = 0.8 or 80%

Therefore, the sensitivity (recall) is 0.8 or 80%, the specificity is 0.706 or 70.6%, the precision is 0.952 or 95.2%, the recall (same as sensitivity) is 0.8 or 80%, and the accuracy is 0.8 or 80%.