MACHINE LEARNING

1) c
2) b
3) c
4) b
5) b
6) a,d
7) b
8) a
9) b
10)

R-squared is a commonly used metric to measure the goodness of fit of a regression model. However, one limitation of R-squared is that it does not take into account the number of predictors used in the model. As the number of predictors increases, R-squared tends to increase even if the additional predictors do not contribute significantly to the model.

Adjusted R-squared is a modified version of R-squared that takes into account the number of predictors used in the model. It penalizes the presence of unnecessary predictors in the model by adjusting for the number of predictors used. Adjusted R-squared is given by the following formula:

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

where n is the sample size, k is the number of predictors, and R-squared is the regular R-squared value.

The adjusted R-squared value ranges from negative infinity to 1, with higher values indicating a better fit of the model. The adjusted R-squared penalizes the presence of unnecessary predictors by subtracting a penalty term from the regular R-squared value. The penalty term is proportional to the number of predictors used in the model, and it becomes larger as the number of predictors increases.

11)Ridge and Lasso regression are two popular regularization techniques used in linear regression to prevent overfitting of the model. Both techniques introduce a penalty term in the cost function of the regression model, which penalizes the coefficients of the predictors in the model. However, there are some key differences between Ridge and Lasso regression:

1. Penalty term: Ridge regression adds the L2 norm of the coefficient vector as the penalty term, while Lasso regression adds the L1 norm of the coefficient vector as the penalty term.
2. Impact on coefficients: Ridge regression shrinks the coefficients towards zero, but never actually sets them to zero. This means that all predictors remain in the model, but their impact is reduced. Lasso regression, on the other hand, has the ability to set some coefficients to zero, effectively performing variable selection and eliminating some predictors from the model altogether.
3. Solution uniqueness: Ridge regression always has a unique solution, while Lasso regression may not have a unique solution in some cases. This is because the L1 norm used in Lasso regression can cause some coefficients to become exactly zero, leading to multiple solutions that produce the same minimum error.
4. Bias-variance tradeoff: Ridge regression trades off between reducing variance (overfitting) and increasing bias (underfitting), while Lasso regression trades off between reducing variance and selecting the most important predictors. This means that Ridge regression is more suitable when all predictors are potentially useful, while Lasso regression is more suitable when some predictors are more important than others.

12) VIF stands for Variance Inflation Factor. It is a measure of the multicollinearity among the predictors (independent variables) in a regression model. Multicollinearity occurs when two or more predictors in the model are highly correlated with each other, which can lead to unstable and unreliable estimates of the coefficients and standard errors.

The VIF for a predictor is the ratio of the variance of its coefficient estimate when it is included in the model to the variance of its coefficient estimate when it is excluded from the model, after controlling for the other predictors. A VIF value of 1 indicates no multicollinearity, while higher values indicate increasing levels of multicollinearity.

In general, a VIF value of 1-2 is considered acceptable, indicating little or no multicollinearity. VIF values above 5 or 10 are often considered problematic, indicating

high levels of multicollinearity that can lead to unreliable coefficient estimates and inflated standard errors.

13) Scaling the data before training a model is important for several reasons:

1. Prevents bias: Some machine learning algorithms (such as k-NN, SVM, and neural networks) are sensitive to the scale of the input features. If the features are on different scales, then the algorithm may give more importance to the features with higher scales, leading to biased results.
2. Improves convergence: Scaling the data can help some optimization algorithms (such as gradient descent) converge faster to the optimal solution. When the features are on different scales, the optimization algorithm may take longer to converge or get stuck in a local minimum.
3. Helps with regularization: Some regularization techniques (such as L1 and L2 regularization) are sensitive to the scale of the input features. If the features are not scaled, then the regularization penalty may not be distributed evenly among the features, leading to suboptimal results.
4. Reduces computational cost: Scaling the data can reduce the computational cost of some machine learning algorithms. When the features are on different scales, some algorithms may require more iterations or computations to reach the optimal solution.

14) In linear regression, there are several metrics that can be used to check the goodness of fit of the model. Some of the most commonly used metrics are:

1. R-squared (R2): This metric measures the proportion of variance in the dependent variable that is explained by the independent variables in the model. It ranges from 0 to 1, with higher values indicating a better fit.
2. Adjusted R-squared: This metric is similar to R-squared but takes into account the number of independent variables in the model. It penalizes models that include unnecessary variables that do not improve the fit.
3. Mean Squared Error (MSE): This metric measures the average squared difference between the actual and predicted values of the dependent variable. Lower values indicate a better fit.
4. Root Mean Squared Error (RMSE): This metric is the square root of the MSE and is in the same units as the dependent variable. It is often used to interpret the errors in the same units as the dependent variable.
5. Mean Absolute Error (MAE): This metric measures the average absolute difference between the actual and predicted values of the dependent variable. It is less sensitive to outliers than the MSE.

6. Residual plots: These plots can be used to visualize the distribution of the residuals (the difference between the actual and predicted values of the dependent variable). A good model should have residuals that are randomly distributed around zero and have a constant variance

15) Predicted

| True | False |

Actual -------|--------|-------|

True | 1000 | 250 |

False | 50 | 1200 |

We can use the following formulas to calculate the different metrics:

- Sensitivity (True Positive Rate): TP / (TP + FN)
- Specificity (True Negative Rate): TN / (TN + FP)
- Precision: TP / (TP + FP)
- Recall: TP / (TP + FN)
- Accuracy: (TP + TN) / (TP + TN + FP + FN)

where TP = True Positives, TN = True Negatives, FP = False Positives, and FN = False Negatives.

Using these formulas, we get:

- Sensitivity = 1000 / (1000 + 50) = 0.9524 or 95.24%
- Specificity = 1200 / (1200 + 250) = 0.8276 or 82.76%
- Precision = 1000 / (1000 + 250) = 0.8 or 80%
- Recall = 1000 / (1000 + 50) = 0.9524 or 95.24%
- Accuracy = (1000 + 1200) / (1000 + 1200 + 250 + 50) = 0.8933 or 89.33%

Statistics

1)d

2)c

3)a

4)c

5)c

6)a

7)c

8)b

9)b