1. A
2. B
3. C
4. A
5. D
6. A, D
7. B, C
8. A, C
9. B
10. The R-square test is used to determine the fit in regression. R-squared near 1 is a good fit for the model. The adjusted R-squared is a modified version of R-squared that adjusts for the number of predictors in the regression model. Because of the way it's calculated, adjusted r-squared can be used to compare the fit of regression models with different numbers of predictor variables. It penalises adding unnecessary features and allows a comparison of a regression model with a different number of predictors.
11.

Ridge Regression	Lasso Regression
	Lasso regression stands for Least Absolute Shrinkage and Selection Operator
add a penalty term which is equal to the square of the coefficient	adds penalty term to the cost function. This term is the absolute sum of the coefficients
Ridge never sets the value of coefficient to absolute zero.	lasso regression is that it tends to make coefficients to absolute zero
Ridge regression decreases the complexity of a model	Lasso sometimes struggles with some types of data.
Ridge model is not good for feature reduction	If there are two or more highly collinear variables then LASSO regression select one of them randomly which is not good for the interpretation of data

12. VIF stands for Variance Inflation Factor. VIF is another commonly used tool to detect whether multicollinearity exists in a regression model. It measures how much the variance of the estimated regression coefficient is inflated due to collinearity.
13. The pre-processing step of data scaling is advised when using machine learning. When your data has input values with different scales, standardiation can be helpful and even necessary in some machine learning algorithms.
14. the different metrics which are used to check the goodness of fit in linear regression
    - Mean Squared Error:
      The most common metric for regression tasks is MSE. It is the average of the squared difference between the predicted and actual value. MSE penalizes large errors.
    - Mean Absolute Error:
      This is simply the average of the absolute difference between the target value and the value predicted by the model. MAE does not penalize large errors.

- Root Mean Squared Error:
  This is the square root of the average of the squared difference of the predicted and actual value. We are aware that residuals represent the distance between the points and the regression line.

- R-squared
  It measures the strength of the relationship between your model and the dependent variable. If the data points are very close to the regression line, then the model accounts for a good amount of variance, thus resulting in a high R² value.

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

Actual/Predicted	True	False
True	1000	50
False	250	1200

Accuracy:
It's the ratio between the number of correct predictions and the total number of predictions.
Accuracy = (TP+TN)/(TP+TN+FP+FN) or Correct Prediction/ total Prediction
= (1000+1200)/(1000+1200+250+50)
= 0.88

Precision:
how many predictions are actually positive out of all the total positive predicted. Precision is a useful measure in cases where false positives are a greater concern than false negatives. Whenever False Positive is much more important use Precision.
Precision = TP/(TP+FP)
= 1000/(1000+250)
= 0.8

Recall:
how many observations of positive class are actually predicted as positive. It is also known as Sensitivity. Recall is defined as the ratio of the total number of correctly classified positive classes divide by the total number of positive classes. Recall is a useful measure when the false negative is more important than the false positive.
Whenever False Negative is much more important use Recall.
Recall = TP/(TP+FN)
= 1000/(1000+50)
= 0.95

F1-Score:
F1-Score is used when the False Negatives and False Positives are important. F1-Score is a better metric for Imbalanced Data.
F1-Score = 2*((Recall*Precision)/ (Recall+ Precision))
= 2*((0.95*0.8)/ (0.95+0.8))
= 2*(0.76/1.75)
= 0.87