Machine learning

1 A

2 B

3 D

4 A

5 B

6 A & D

7 B & C

8 C

9 A & B

10 The adjusted R-squared compares the explanatory power of regression models that contain different numbers of predictors. The adjusted R-squared increases only if the new term improves the model more than would be expected by chance. It decreases when a predictor improves the model by less than expected by chance.

11 Ridge and Lasso regression uses two different penalty functions. Ridge uses l2 where as lasso go with l1. In ridge regression, the penalty is the sum of the squares of the coefficients and for the Lasso, it’s the sum of the absolute values of the coefficients. It’s a shrinkage towards zero using an absolute value (l1 penalty) rather than a sum of squares(l2 penalty).As we know that ridge regression can’t zero coefficients. Here, you either select all the coefficients or none of them whereas LASSO does both parameter shrinkage and variable selection automatically because it zero out the co-efficients of collinear variables. Here it helps to select the variable(s) out of given n variables while performing lasso regression.

12 Variance inflation factor (VIF) is a measure of the amount of multicollinearity in a set of multiple regression variables. Mathematically, the VIF for a regression model variable is equal to the ratio of the overall model variance to the variance of a model that includes only that single independent variable. VIF above 10 indicates high correlation and is cause for concern. Some authors suggest a more conservative level of 2.5 or above. Sometimes a high VIF is no cause for concern at all. For example, you can get a high VIF by including products or powers from other variables in your regression, like x and x2.

13 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 can help the gradient descent converge more quickly towards the minima. Feature scaling is essential for machine learning algorithms that calculate distances between data. Therefore, the range of all features should be normalized so that each feature contributes approximately proportionately to the final distance. 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 can help the gradient descent converge more quickly towards the minima.

14 Five metrics give us some hints about the goodness-of-fit of our model. The first twometrics, the Mean Absolute Error and the Root Mean Squared Error. The most common metric for regression tasks is MSE. It has a convex shape. It is the average of the squared difference between the predicted and actual value.

15 Sensitivity = TP /TP + FN

1000/1000+1200=0.45

Specificity = TN/TN + FP

50/50+250 = 0.16

Accuracy = TP + TN/TP+TN+FP+FN

1000+50/1000+50+250+1200=0.42

Recall = TP/TP+FN

1000/1000+1200= 0.45

Precision = TP/TP+FP

1000/1000+250=0.8

SQL

1 A

2 A, C & D

3 C

4 C

5 B

6 B

7 A

8 C

9 D

10 ASC

Statistic

1 D

2 C

3 A

4 C

5 A

6 D

7 C

8 B

9 B

10 Dot plots include ALL values from the data set, with one dot for each occurrence of an observed value from the set. A histogram or box plot will deal more efficiently with large data sets. Dot plots show all values in the set. The median, however, is not readily seen, as it is in the box plot.

11 A metric scale measures quantitative characteristics or variables. The term metric scale summarizes interval scales, ratio scales and absolute scales. Quantitative numerical values allow for most of statistical calculations and comparisons. It is an interval scale.

12 Create a null hypothesis.

Create an alternative hypothesis.

Determine the significance level.

Decide on the type of test you'll use.

Perform a power analysis to find out your sample size.

Calculate the standard deviation.

Use the standard error formula.

13  To count the number of Heads when tossing 5 coins over and over. This is an example of the [Binomial distribution](https://en.wikipedia.org/wiki/Binomial_distribution).