Machine Learning Worksheet 6

1. C)
2. C)
3. C)
4. D) Either recall or F1 score should be used
5. B)
6. A) and D)
7. A) and D)
8. A) and C)
9. A) and B)
10. The adjusted R-squared is a modified version of R-squared that has been adjusted for the number of predictors in the model. 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. The difference between ridge and lasso regression is that it tends to make coefficients to absolute zero as compared to Ridge which never sets the value of coefficient to absolute zero.
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.
13. Most of the times, the dataset will contain features highly varying in magnitudes, units and range. But since, most of the machine learning algorithms use Euclidean distance between two data points in their computations, this is a problem.
    If left alone, these algorithms only take in the magnitude of features neglecting the units. The results would vary greatly between different units, 5kg and 5000gms. The features with high magnitudes will weigh in a lot more in the distance calculations than features with low magnitudes. To suppress this effect, we need to bring all features to the same level of magnitudes. This can be achieved by scaling.

14. Following are the main evaluation metrics for Linear Regression's goodness of fit evaluation:
    i) Mean Squared Error
    ii) Mean Absolute Error
    iii) R Squared

15. Sensitivity: 0.8
    Specificity: 0.96
    Precision: 0.95
    Recall: 0.8

Accuracy: 0.88