**Worksheet 6 Machine Learning**

1) C

2) B

3) C

4) C

5) B

6) A, D

7) C

8) A

9) A

10) The adjusted R-squared compensates for the addition of variables and **only increases if the new predictor enhances the model above what would be obtained by probability**. Conversely, it will decrease when a predictor improves the model less than what is predicted by chance.

11) Ridge regression puts a similar constraint on the coefficients by introducing a penalty factor. However, **while lasso regression takes the magnitude of the coefficients, ridge regression takes the square**. Ridge regression is also referred to as L2 Regularization.

12) What is VIF?

* A variance inflation factor (VIF) provides a measure of multicollinearity among the independent variables in a multiple regression model.
* Detecting multicollinearity is important because while multicollinearity does not reduce the explanatory power of the model, it does reduce the statistical significance of the independent variables.
* A large VIF on an independent variable indicates a highly collinear relationship to the other variables that should be considered or adjusted for in the structure of the model and selection of independent variables.

What is the suitable value of a VIF for a feature to be included in a regression modelling? Most research papers consider a VIF (Variance Inflation Factor) > 10 as an indicator of multicollinearity, but some choose a more conservative threshold of **5 or even 2.5**

**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.**

**14) There are three error metrics that are commonly used for evaluating and reporting the performance of a regression model; they are: Mean Squared Error (MSE). Root Mean Squared Error (RMSE). Mean Absolute Error (MAE)**

**15) Sensitivity** 0.9524 TPR = TP / (TP + FN)

**Specificity** 0.1724 SPC = TN / (FP + TN)

**Precision** 0.4545 PPV = TP / (TP + FP)

**Accuracy** 0.5000 ACC = (TP + TN) / (P + N)