MACHINE LEARNING WORKSHEET 6

In Q1 to Q5, only one option is correct, Choose the correct option

1. A

2. B

3. A

4. A

5. B

In Q6 to Q9, more than one options are correct, Choose all the correct options:

6. A,D

7. B,C

8. C

9. A,B

Q10 to Q15 are subjective answer type questions, Answer them briefly.

10. Explain how does the adjusted R-squared penalize the presence of unnecessary predictors in the model?

A. R^2 explains how well the model predicts the observation. Adjusted R^2 is the one that takes into account more observations (or degrees of freedom). So,
Adjusted R^2 predicts the model better? Then why is this less than R^2? It appears it should often be more.

11. Differentiate between Ridge and Lasso Regression.

A. 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. Limitation of Lasso Regression: Lasso sometimes struggles with some types of data.

12. What is VIF? What is the suitable value of a VIF for a feature to be included in a regression modelling?

A. The Variance Inflation Factor (VIF) measures the severity of multicollinearity in Regression analysis It is a statistical concept that indicates the increase in the variance of a regression coefficient as a result of collinearity

13. Why do we need to scale the data before feeding it to the train the model

A. **We scale** down the images **before feeding it** into the network in order to reduce the number of parameters. When the number of parameters are high, **we** tend to increase the **requirement** of computation power. **Scaling** down images **does** decreases the detail and the **scale** size is purely dependent on the target of our **model**.

14. What are the different metrics which are used to check the goodness of fit in linear regression

A. The **goodness of fit** of a statistical model describes how well it fits a set of observations. Measures of goodness of fit typically summarize the discrepancy between observed values and the values expected under the model in question. Such measures can be used in statistical hypothesis testing, e.g. to test of normality of residuals, to test whether two samples are drawn from identical distributions , or whether outcome frequencies follow a specified distribution . In the analysis of variance, one of the components into which the variance is partitioned may be a lack of fit sum of suares.

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

A. **SENSITIVITY:**

TP/TP+FN=0.8

SPECIFICITY:

TN/TN+FP=0.96

PRECISION:

TP/TP+FN=0.95

RECALL:

TP/TP+FN=0.8

ACCURACY:

TP+TN/TP+TN+FP+FN=0.88