

STATISTICS WORKSHEET

Q1 to Q10 are MCQs with only one correct answer. Choose the correct option.

- 1) D
- 2) C
- 3) C
- 4) B
- 5) C
- 6) B
- 7) A
- 8) A
- 9) B
- 10) A

MACHINE LEARNING

Q1 to Q15 are subjective answer type questions, Answer them briefly.

ANSWERS:

- 1) ANS: I think 'R SQUARED' is a better measure of goodness of a fit model in regression because, it is a statistical measure in regression model that determines the proportion of variable that can be explained by the independent variable.
- 2) ANS: The Total Sum of Squares tells how much variation there is in the dependent variable.
The Explained SS tells how much the variation in the dependent variable your model explained.
The Residual sum of Squares tell how much of the dependent variable's variation your model did not explain.
- 3) ANS: Regularization refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.
- 4) ANS: It calculate the amount of probability of a specific feature that is classified incorrectly when selected randomly.
- 5) ANS: The condition when the model completely fits the training data but fails to generalize the testing unseen data.

- 6) ANS: Ensemble method is a machine learning techniques that combines several base models in order to produce one optimal predictive model.
- 7) ANS: Bagging is a technique for reducing prediction variance by producing additional data for training from a dataset by combining repetitions with combinations to create multisets of the original data. Boosting is an iterative strategy for adjusting an observation's weight based on the previous classification.
- 8) ANS: The average error for each calculated using predictions from the trees that do not contain in their respective bootstrap sample.
- 9) ANS: When the dataset is split into a k number of folds and is used to evaluate the model's ability when given new data.
- 10) ANS: Finding a set of optimal hyperparameter values for a learning algorithm while applying this optimized algorithm to any dataset.
- 11) ANS: I think it causes the model to converge too quickly to a suboptimal solution.
- 12) ANS: Non-linear problems can't be solved with logistic regression because it has a linear decision surface.
- 13) Ans: AdaBoost is the first designed boosting algorithm with a particular loss function

Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem.

- 14) ANS: The property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters.
- 15) ANS: Linear Kernel is used when the data is linearly separable .

RBF Kernel is used in various kernelized learning algorithms.

Polynomials Kernel is used to represents the similarity of vectors.

