**MACHINE LEARNING WORKSHEET**

***In Q1 to Q11, only one option is correct, choose the correct option:***

1. Which of the following methods do we use to find the best fit line for data in Linear Regression?

A) Least Square Error

B) Maximum Likelihood

C) Logarithmic Loss

D) Both A and B

Answer: A

2. Which of the following statement is true about outliers in linear regression?

A) Linear regression is sensitive to outliers

B) linear regression is not sensitive to outliers

C) Can’t say

D) none of these

Answer: A

3. A line falls from left to right if a slope is \_\_\_\_\_\_?

A) Positive

B) Negative

C) Zero

D) Undefined

Answer: B

4. Which of the following will have symmetric relation between dependent variable and independent variable?

A) Regression

B) Correlation

C) Both of them

D) None of these

Answer: B

5. Which of the following is the reason for over fitting condition?

A) High bias and high variance

B) Low bias and low variance

C) Low bias and high variance

D) none of these

Answer: C

6. If output involves label then that model is called as:

A) Descriptive model

B) Predictive modal

C) Reinforcement learning

D) All of the above

Answer: B

7. Lasso and Ridge regression techniques belong to \_\_\_\_\_\_\_\_\_?

A) Cross validation

B) Removing outliers

C) SMOTE

D) Regularization

Answer: D

8. To overcome with imbalance dataset which technique can be used

A) Cross validation

B) Regularization

C) Kernel

D) SMOTE

Answer: D

9. The AUC Receiver Operator Characteristic (AUCROC) curve is an evaluation metric for binary classification problems. It uses \_\_\_\_\_ to make graph?

A) TPR and FPR

B) Sensitivity and precision

C) Sensitivity and Specificity

D) Recall and precision

Answer: A

10. In AUC Receiver Operator Characteristic (AUCROC) curve for the better model area under the curve should be less.

A) True

B) False

Answer: B

11. Pick the feature extraction from below:

A) Construction bag of words from a email

B) Apply PCA to project high dimensional data

C) Removing stop words

D) Forward selection

Answer: B

***In Q12, more than one options are correct, choose all the correct options:***

12. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?

A) We don’t have to choose the learning rate.

B) It becomes slow when number of features is very large.

C) We need to iterate.

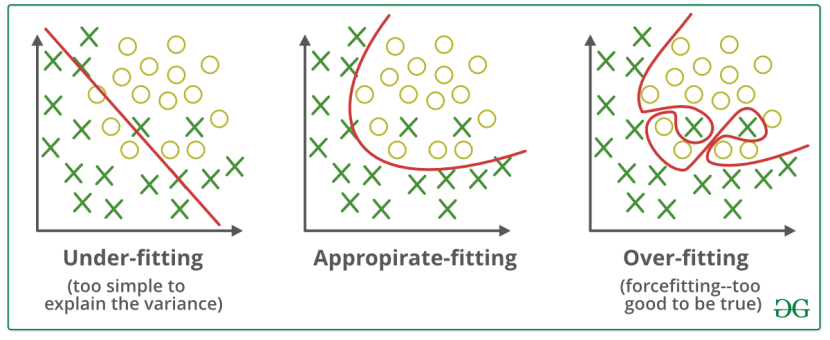
D) It does not make use of dependent variable.

Answer: A,B,C

***Q13 and Q15 are subjective answer type questions, Answer them briefly.***

13. Explain the term regularization?

* Regularization is most important concept of Machine learning.
* It is a form of regression, that regularizes or shrinks the coefficient estimates towards zero. In other words, this technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting.
* Regularization is a technique used to reduce the error by fitting a function appropriately on the given training set and avoid overfitting.
* Below given, is an example of overfitting and under fitting.



* And we have to appropriate fit/Best fit to our dataset where Regularization is used.
* In the context of machine learning, the term ‘Regularization’ refers to a set of techniques that help the machine to learn more than just memorize.

14. Which particular algorithms are used for regularization?

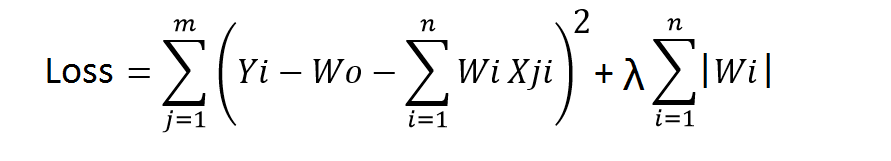
The Regularization techniques/Algorithms are :

1. L1 regularization

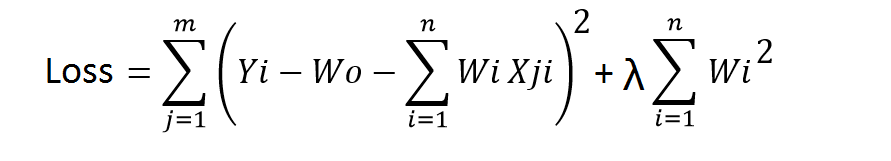
2. L2 regularization

3. Dropout regularization

* Most common/Particular used is L1 and L2 techniques:-
* L1 LASSO(Least Absolute Shrinkage and Selection Operator)  Regression: Below is a L1 equation



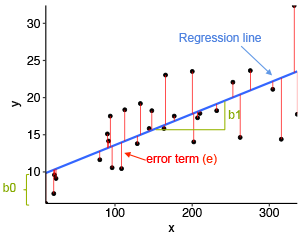
* This technique uses absolute weight values for normalization. λ is again a tuning parameter and behaves in the same as it does when using ridge regression.
* As loss function only considers absolute weights, optimization algorithms penalize higher weight values.
* In ridge regression, loss function along with the optimization algorithm brings parameters near to zero but not actually zero, while lasso eliminates less important features and sets respective weight values to zero. Thus, lasso also performs feature selection along with regularization.
* L2 Ridge Regression.



* As seen above, the original loss function is modified by adding normalised weights. Here normalised weights are in the form of squares.
* When using this technique, we add the sum of weight’s square to a loss function and thus create a new loss function.
* You may have noticed parameters λ along with normalised weights. λ is the parameter that needs to be tuned using a cross-validation dataset. When you use λ=0, it returns the residual sum of square as loss function which you chose initially. For a very high value of λ, loss will ignore core loss function and minimise weight’s square and will end up taking the parameters’ value as zero.
* Both Technique highly used to remove errors and overfitting datasets.

15. Explain the term error present in linear regression equation?

* A Linear Regression model’s main aim is to find the best fit linear line and the optimal values of intercept and coefficients such that the error is minimized.
* Error is the difference between the actual value and Predicted value and the goal is to reduce this difference.



In the above diagram,

* x is our dependent variable which is plotted on the x-axis and y is the dependent variable which is plotted on the y-axis.
* Black dots are the data points i.e the actual values.
* bo is the intercept which is 10 and b1 is the slope of the x variable.
* The blue line is the best fit line predicted by the model i.e the predicted values lie on the blue line.

The vertical distance between the data point and the regression line is known as error or residual. Each data point has one residual and the sum of all the differences is known as the Sum of Residuals/Errors.