MACHINE LEARNING WORKSHEET-1

1. The computational complexity of linear regression is:
B) $O(n)$

2. Which of the following can be used to fit non-linear data?
C) Polynomial Regression

3. Which of the following can be used to optimize the cost function of Linear Regression?
B) Gradient Descent

4. Which of the following method does not have a closed-form solution for its coefficients?
C) Lasso

Which gradient descent algorithm always gives optimal solution?
D) All of the above

6. Generalization error measures how well a model performs on training data.
B) False

7. The cost function of linear regression can be given as $J(w0,w1) = 1\ 2m \sum (w0 + w1x(i) - y(i))2\ m\ i=1$. The half term at start is due to:

C) it does not matter whether half is there or not

8. Which of the following will have symmetric relation between a dependent variable and an independent variable?
C) Both of them

In Q9 to Q11, more than one options are correct, Choose all the correct options:

9. 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 are very large.

10. Which of the following statement/s are true if we generated data with the help of polynomial features with 5 degrees of freedom which perfectly fits the data?
A) Linear Regression will have high bias and low variance.
C) Polynomial with degree 5 will have low bias and high variance.

11. Which of the following sentence is false regarding regression?
C) It discovers causal relationship.
D) No inference can be made from regression line.

Q12 and Q13 are subjective answer-type questions. Answer them briefly.

12. Which Linear Regression training algorithm can we use if we have a training set with millions of features?

- Gradient descent should be used when having a large set of features. As we saw that when we use the Normal Equation for Linear Regression, It suffers from an increase in the number of features. It becomes computationally heavy.

13. Which algorithms will not suffer or might suffer if the features in the training set have very different scales?

- The normal equation has a closed-form solution that uses the mean of the features to compute the coefficients and intercepts. This doesn't change with the change in scale.
  Gradient descent on the other hand is affected by the change in scale as it might take a lot of time to converge at the global minimum.