MACHINE LEARNING ANSWERS

In Q1 to Q8, only one option is correct, Choose the correct option:

1. The computational complexity of linear regression is:
Answer : B) 0(n)

2. Which of the following can be used to fit non-linear data?
Answer: B) Logistic Regression

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

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

5. Which gradient descent algorithm always gives optimal solution?
Answer: D) All of the above

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

7. The cost function of linear regression can be given as j(0, 1)= $\frac{1}{2m}\sum_{i=1}^{m}(w_0 + w_1x^{(i)} - y^{(i)})^2$
Answer: D) None of the above

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

9. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
Answer: A) we don't have to choose the learning rate,B) It becomes slow when number of the features is very large ,C) No need to iterate

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?
Answer : D) polynomial with degree 5 will have high bias and low variance

11. Which of the following sentence is false regarding regression?
Answer : C) It discover casual relationship.

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?

Answer : If you have a training set with millions of features you can use Stochastic Gradient Descent or Mini-batch Gradient Descent, and perhaps Batch Gradient Descent if the training set fits in memory. But you cannot use the Normal Equation because the computational complexity grows quickly (more than quadratically) with the number of features.

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

Answer : If the features in your training set have very different scales, the cost function will have the shape of an elongated bowl, so the Gradient Descent algorithms will take a long time to converge. To solve this you should scale the data before training the model. Note that the Normal Equation will work just fine without scaling.