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

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

1. Which of the following can be used to fit non-linear data?

C) Polynomial Regression

1. Which of the following can be used to optimize the cost function of Linear Regression?

B) Gradient Descent

1. Which of the following method does not have closed form solution for its coefficients?

C) Lasso

1. Which gradient descent algorithm always gives optimal solution?
   1. Stochastic Gradient Descent
2. Generalization error measures how well a model performs on training data.
   1. True
3. The cost function of linear regression can be given as 𝐽(𝑤 , 𝑤 ) = 1

∑𝑚

(𝑤

+ 𝑤 𝑥(𝑖) − 𝑦(𝑖))2.

The half term at start is due to:

0 1 2𝑚

𝑖=1 0 1

* 1. scaling cost function by half makes gradient descent converge faster.
  2. presence of half makes it easy to do grid search.
  3. it does not matter whether half is there or not.
  4. None of the above.

1. Which of the following will have symmetric relation between dependent variable and independent variable?
   1. Regression B) Correlation

C) Both of them

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

1. Which of the following is true about Normal Equation used to compute the coefficient of the Linear Regression?
   1. It becomes slow when number of features are very large.
   2. We need to iterate.
2. 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?
   1. Linear Regression will have high bias and low variance.
   2. Polynomial with degree 5 will have high bias and low variance
3. Which of the following sentence is false regarding regression?
   1. It discovers causal relationship.

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

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

Ans: We can try to use gradient descent, stochastic gradient descent.

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

Ans: The normal equations methods does not require normalizing the features and it remain unaffected. Feature scaling is required for the various gradient descent algorithms and will also help to converge quicker.