

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

- The computational complexity of linear regression is:
 - A) $O(n^2)$
 - B) $O(n)$ answer
 - C) $O(n^3)$
 - D) $O(n^4)$
- Which of the following can be used to fit non-linear data?
 - A) Lasso Regression
 - B) Logistic Regression
 - C) Polynomial Regression
 - D) Ridge Regression
- Which of the following can be used to optimize the cost function of Linear Regression?
 - A) Entropy
 - B) Gradient Descent (answer)
 - C) Pasting
 - D) None of the above.
- Which of the following method does not have closed form solution for its coefficients?
 - A) extrapolation
 - B) Ridge
 - C) Lasso (Answer)
 - D) Elastic Nets
- Which gradient descent algorithm always gives optimal solution?
 - A) Stochastic Gradient Descent (Answer)
 - B) Mini-Batch Gradient Descent
 - C) Batch Gradient Descent
 - D) All of the above
- Generalization error measures how well a model performs on training data.
 - A) True
 - B) False
- The cost function of linear regression can be given as $J(\theta_0, \theta_1) = \frac{1}{2} \sum_{i=1}^n (y_i - \theta_0 - \theta_1 x_i)^2$

\sum

(

$+ (y_i - \theta_0 - \theta_1 x_i)^2$.

The half term at start is due to:

$$= \frac{1}{2} \sum_{i=1}^n (y_i - \theta_0 - \theta_1 x_i)^2$$

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

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

- Which of the following is true about Normal Equation used to compute the coefficient of the LinearRegression?
 - We don't have to choose the learning rate.
 - It becomes slow when number of features are very large.
 - We need to iterate.
 - It does not make use of dependent variable.
- 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?
 - Linear Regression will have high bias and low variance.
 - Linear Regression will have low bias and high variance.
 - Polynomial with degree 5 will have low bias and high variance.
 - Polynomial with degree 5 will have high bias and low variance.
- Which of the following sentence is false regarding regression?
 - It relates inputs to outputs.
 - It is used for prediction.
 - It discovers causal relationship.
 - No inference can be made from regression line.

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

- Which Linear Regression training algorithm can we use if we have a training set with millions of features? **ans- SGDRegression**
- Which algorithms will not suffer or might suffer, if the features in training set have very different scales?
 - **answer:- The normal equations method does not require normalizing the features, so it remains unaffected by features in the training set having very different scales.**
 -
 - **Feature scaling is required for the various gradient descent algorithms. Feature scaling will help gradient descent converge quicker.**