

## Machine Learning Solutions

1. D
2. C
3. B
4. C
5. D
6. B
7. C
8. C
9. A, B
10. A, D
11. C, D

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

- Linear Regression is not but finding the relationship between Independent & Dependent variables.
- There are three algorithms namely: Batch Gradient descent, Stochastic gradient descent (SGD), & Mini-Batch gradient Descent.
- The Batch Gradient descent sums up all the training samples at each step which makes it pretty time consuming and does not make it a good fit for large data sets.
- Where as the SGD takes a single sample at each iteration and sums up all of them at the end of each step which makes it ideal for using for large datasets since it can even compute faster.
- The Mini-Batch Gradient descent is also similar to SGD which uses 'n' samples instead of 1- sample at each iteration which makes this too ideal for large datasets.

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

- The Gradient descent suffers from features of different scales because it will take a longer time to reach the global max. Feature scaling will help global descent coverage quicker.
- The normal equations method does not require normalizing the features, so it remains unaffected by features in training set having different scales.