Machine learning assignment

1. 𝑂(𝑛 3 )
2. Ridge Regression
3. Gradient Descent
4. Lasso
5. All the above
6. True
7. scaling cost function by half makes gradient descent converge faster.
8. Both of them
9. a)We don’t have to choose the learning rate.

b)It becomes slow when number of features are very large.

10) a)Linear Regression will have high bias and low variance.

b) Polynomial with degree 5 will have low bias and high variance.

11)c) It discovers causal relationship.

d) No inference can be made from regression line.

12)What Linear Regression training algorithm can you use if you have a training set with millions of features?

Ans) 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?

Ans) 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.