**Machine Learning**

1. C
2. C
3. B
4. C
5. D
6. C
7. C
8. C
9. C
10. A,B
11. C,D
12. We can use Batch Gradient Descent, Stochastic Gradient Descent, Mini-Batch Gradient Descent. MBGD and SGD will work best among them because in these two algorithms neither each of them need to load the whole dataset into the memory in order to take gradient descent. And Batch Gradient Descent will also fine if you have more memory to load all the data.
13. 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.