Machine Learning:

Question 1 to 8:

1. C) $O(n2)$
2. C) Polynomial Regression
3. B) Gradient Descent
4. B) Ridge
5. C) Batch Gradient Descent
6. B) False
7. C) it does not matter whether half is there or not.
8. B) Correlation

Question 9 to 11:

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 ; C) Polynomial with deree 5 will have low bias and high variance.

11. C) It discovers causal relationship ; D) No inference can be made from regression line.

Question 12 & 13:

12. Gradient descent, Stochastic gradient descent, or mini-batch gradient descent. SGD AND MBGD would work the best because neither of them need to load the entire dataset into memory in order to take one step of gradient descent. Batch would be ok with the caveat that you have enough 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.