Ans. 1  B) O(n)

Ans. 2 A) Lasso Regression

Ans. 3 B) Gradient Regression

Ans 4 C) Lasso

Ans. 5 D) All of the above

Ans. 6 A) True

Ans. 7 ) None of the above

Ans. 8 C) Both of them

Ans. 9 A) We don't have to choose the learning rate.
B) It becomes slow when number of features are very large.

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

Ans. 11 C) It discovers causal relationship.

Ans. 12 If I have a training set with millions of features I can use Stochastic Gradient Descent or Mini-batch Gradient Descent, and perhaps Batch Gradient Descent if the training set fits in memory. But I cannot use the Normal Equation because the computational complexity grows quickly with the number of features.

Ans. 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.