MACHINE LEARNING

Question 1: B) O (n)

Question 2: B) Logistic Regression

Question 3: B) Gradient Descent

Question 4: C) Lasso

Question 5: B) Mini-Batch Gradient Descent

Question 6: B) False

Question 7: A) scaling cost function by half makes gradient descent converge faster

Question 8: B) Correlation

Question 9: A) We don't have to choose the learning rate.

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

C) We need to iterate.

Question 10: A) Linear Regression will have high bias and low variance

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

Question 11: C) It discovers causal relationship.

D) No inference can be made from regression line.

Question 12:

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.

Question 13:

If the features in the 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 one should scale the data before training the model. Note that the Normal Equation will work just fine without scaling.