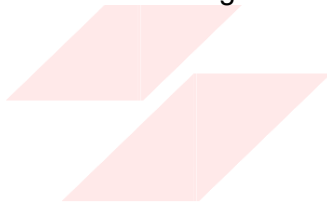# MACHINE LEARNING

Answer Keys

1. C) $O(n^2)$
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. (It is used to cancel out square during calculating its derivative.
8. B) Correlation
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 degree 5 will have low bias and high variance.
11. C) It discovers causal relationship.
    C) No inference can be made from regression line.
12. If we have a training set with millions of features, we can use Stochastic Gradient Descent or Mini-batch Gradient Descent, and also Batch Gradient Descent if the training set fits in memory. But we cannot use Normal Equation or the SVD approach as the computational complexity grows quickly with the number of features.
13. If the features in the training set have very different scales, the cost function will have the shape of elongated bowl, so the Gradient Descent algorithms will take long time to converge. To solve this, we should scale the data before training the model. The Normal Equation or the SVD approach will not suffer as the weights will be balanced off accordingly.