Machine Learning Worksheet

- **A1.B):-** 0(n)
- **A2.B):-** Logistic Regression
- **A3.B):-** Gradient Descent
- **A4.C):-** Lasso
- **A5.A):-** Stochastic Gradient Descent
- **A6. A):-** True
- **A7.A):-** Scaling cost function by half makes gradient descent coverage faster.
- **A8.B):-** Correlation
- **A9.A):-** We don't have to choose the learning rate
  B):- It become slow when number of features are very large.
- **A10. B):-** Linear regression will have high bias and low variance.
  C):- polynomial with degree 5 will have low bias and high variance.
- **A11.C):-** It discovers casual relationship.

- **A12**. We could use batch gradient descent, Stochastic gradient descent, or Mini-batch gradient descent. Stochastic gradient descent and Mini-batch gradient descent would work the best because neither of them need to load the entire dataset into memory in order to take 1 step of gradient descent. Batch would be ok with the caveat that we have enough memory to load all the data.

  The normal equations method would not be a good choice because it is computationally inefficient. The main cause of the computational complexity comes from inverse operation on an (n x n) matrix.

- **A13.** 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.