

MACHINE LEARNING

WORKSHEET 1

1. B) $O(n)$
2. B) Logistic Regression
3. B) Gradient Descent
4. C) Lasso
5. A) Stochastic Gradient Descent
6. A) True
7. A) scaling cost function by half makes gradient descent converge faster.
8. B) Correlation
9. A) We don't have to choose the learning rate
C) It become slow when number of features are very large.

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

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

11. C) It discovers causal relationship

12. we can use batch 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 1 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.

THANKYOU.....

