

MACHINE LEARNING WORKSHEET

1. $O(n)$
2. Ridge Regression
3. Entropy
4. Lasso
5. All of the above
6. False
7. scaling cost function by half makes gradient descent converge faster.
8. Regression
9. We need to iterate.
 - It does not make use of dependent variable.
10. Linear Regression will have high bias and low variance.
 - Linear Regression will have low bias and high variance.
11. It relates inputs to outputs.
 - It is used for prediction.
 - It discovers causal relationship.
- 12.

Linear Regression is a **machine learning** algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting

Yes 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.

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 $(n \times n)$ matrix.

- 13.

MACHINE LEARNING WORKSHEET

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.