

Machine Learning Worksheet 1 Solutions

A1. (B) $O(n)$

A2. (B) Logistic Regression

A3. (B) Gradient Descent

A4. (C) Lasso

A5. (D) All of the above

A6. (A) True

A7. (A) Scaling cost function by half makes gradient descent converge faster.

A8. (B) Correlation

A9. (A) and (B)

A10. (A) and (C)

A11. (C) and (D)

Q12. Which Linear Regression training algorithm can we use if we have a training set with millions of features?

Ans.: 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 needs to load the entire dataset into memory in order to take 1 step of gradient descent.

Batch would be ok with the limitation that we have enough memory to load all the data.

Q13. Which algorithms will not suffer or might suffer, if the features in training set have very different scales?

Ans.: 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, the data should be scaled before training the model.

Note that the Normal Equation will work just fine without scaling.