

Machine Learning – Worksheet 1(Answers)

Q1) option (B)

Q2) option (A)

Q3) option (B)

Q4) option (C)

Q5) option (A)

Q6) option (A)

Q7) option (D)

Q8) option (B)

Q9) option (A) and option (B)

Q10) option (A) and option (C)

Q11) option (C)

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

Ans: 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 one step of gradient descent. We can also use Batch Gradient descent, because it is also 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 $(n \times n)$ matrix.

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

Ans: The normal equations method does not require normalization of features, so it remains unaffected if the features in the training set have very different scales. Feature scaling is required for the various gradient descent algorithms. Feature scaling will help gradient descent converge quicker.

