

MACHINE LEARNING

1. A
2. C
3. B
4. C
5. D
6. A
7. C
8. B
9. A & B
10. A & D
11. C & D

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

Ans: We can use both Batch Gradient Descent, Stochastic Gradient Descent or Mini-Batch Gradient Descent, But Batch Gradient Descent, Stochastic Gradient Descent would work best because neither of them needs to load the entire dataset into memory in order to take 1 step of gradient descent.

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

Ans: The normal equation method does not require normalizing the features, so it remains unaffected by features in the training set having very different scales. Feature's Scaling is required for various gradient descent algorithms. Feature scaling will help gradient descent coverage quicker. so the Gradient Descent Algorithms will take a long to coverage. To solve this, you should scale the data before training the model

