**MACHINE LEARNING**

**ANSWERS**

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

**Theory answers**

12.Batch gradient descent, stochastic gradient descent or mini-batch gradient descent. It doest not need to load the entire data set into memory
for taking 1st step of gradient descent. Batch gradient descent is used when it have enough memory to load all data. But Normal equations method
cannot be used because computational complexity grows very quickly with number of features.

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. Features scaling is required for the various gradient descent algorithms. feature scaling will help gradient descent coverge quicker.
The cost function will have the shape of an elongated bowl, so the Gradient Descent Algorithms will take a long to converge. To solve this you should
scale the data before training the model.