Answers

1. B

2. B

3. B

4. B

5. B

6. B

7. A

8. A

9. A, B

10. C, A

11. C

12. We could use gradient descent algorithms for a training set with millions of features and especially stochastic gradient descent (SGD) for faster computation.

13. Gradient descent algorithms will suffer with features of the training set with different scales where scaling of features will be required to eliminate the problem but the analytical solution using normal equations method will not get affected.