1. D
2. C
3. B
4. C
5. A
6. B
7. C
8. B
9. A,B
10. A,C
11. C,D
12.

    Batch gradient descent, stochastic gradient descent, or mini-batch gradient descent can be used. Stochastic Gradient descent and mini-batch gradient descent 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. We can also use Batch gradient descent, but it needs large memory as we need to do each step of gradient descent on all training examples.

    The normal equations method is not be a good choice because it is computationally inefficient. The main cause of the computational complexity comes from inverse operation on an (n x n) matrix with n features.

13.

Gradient descent algorithm for linear regression is affected by features having different scales and it is a good approach to scale the data for gradient descent as the algorithm converges faster if we have scaled data.

Solving linear regression using Normal equations doesn’t require the data to be scaled.