1. C
2. A
3. C
4. B
5. A
6. B
7. D
8. B
9. A,B
10. A,B
11. Softmax Function
12. The learning rate controls how quickly the model is adapted to the problem. Smaller learning rates require more training epochs given the smaller changes made to the weights each update, whereas larger learning rates result in rapid changes and require fewer training epochs.

A learning rate that is too large can cause the model to converge too quickly to a suboptimal solution, whereas a learning rate that is too small can cause the process to get stuck.

1. The formula for the number of binary input patterns is:

2^n , where n in the number of inputs

8 = 2^3 , 16 = 2^4 and 32 = 2^5

1. In a network of n hidden layers, n derivatives will be multiplied together. If the derivatives are large then the gradient will increase exponentially as we propagate down the model until they eventually explode, and this is what we call the problem of exploding gradient. Alternatively, if the derivatives are small then the gradient will decrease exponentially as we propagate through the model until it eventually vanishes, and this is the vanishing gradient problem.
2. **Epoch** – Represents one iteration over the entire dataset (everything put into the training model).

**Batch** – Refers to when we cannot pass the entire dataset into the neural network at once, so we divide the dataset into several batches.

**Iteration** – if we have 10,000 images as data and a batch size of 200. then an epoch should run 50 iterations (10,000 divided by 5