**Deep-Learning-Worksheet-2 Answer**

1. C
2. A
3. C
4. B
5. C
6. B
7. D
8. D
9. A,B and D
10. B and D
11. Linear Function is appropriate to be used when output is the regression problem.
12. When we use high learning rate, we don’t reach to the global minima at all, the values jumps here and there but it don’t reach the global minima.

While on using very small learning rates, it takes so much time to reach the global minimum as it takes very minor steps on each iterations.

1. For three inputs the number of combinations of 0 and 1 considering x1,x2 and x3 is:

(000,100,010,110,001,101,011,111) Total=8

And for four inputs,the number of combination is:

(0000,1000,0100,1100,0010,0110,1110,0001,1001,0101,1101,0011,1011,0111,1111)

Total =16

So, We see the pattern goes as 2^3,2^4 ,So the final formula is 2^n.

1. The gradients coming from the deeper layers have to go through continuous matrix multiplications because of the the chain rule, and as they approach the earlier layers, if they have small values (<1), they shrink exponentially until they vanish and make it impossible for the model to learn , this is the **vanishing gradient problem**. While on the other hand if they have large values (>1) they get larger and eventually blow up and crash the model, this is the **exploding gradient problem.**
2. The definition terms of each term is showing the difference among itself.Here is the definition

* **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 50)