Ques 1. What is the *vanishing gradient problem* in deep learning and how can it be mediated?

Ans 1. The vanishing gradient problem comes in the picture when the neural network model is train with gradient descent based optimization method.

In neural network model, we do back-propagation (move backward in the network) and calculate the gradient loss function with respect to the weights. As it moves backward, the gradient will tend to become smaller and smaller at each iteration. The small gradient is always a problem, as smaller the gradient, the neurons in the earlier layers will take more time to learn. Thus, the training process takes long time to complete and it will also affect the prediction accuracy. This problem is referred as the vanishing gradient problem.

We can avoid the vanishing gradient problem by changing the activation function from sigmoidal or tanh to RELU function.

Ques 2. What is the result when convolving the array A with the filter B without padding?

```
A= array([[0, 0, 0, 2],
          [2, 1, 0, 0],
          [2, 1, 0, 2],
          [2, 2, 2, 1]])
```

```
B= array([[2, 0, 1],
          [0, 0, 0],
          [2, 0, 1]])
```

Ans. 2 Filtered Matrix = array([[4,6],

[10,7]])

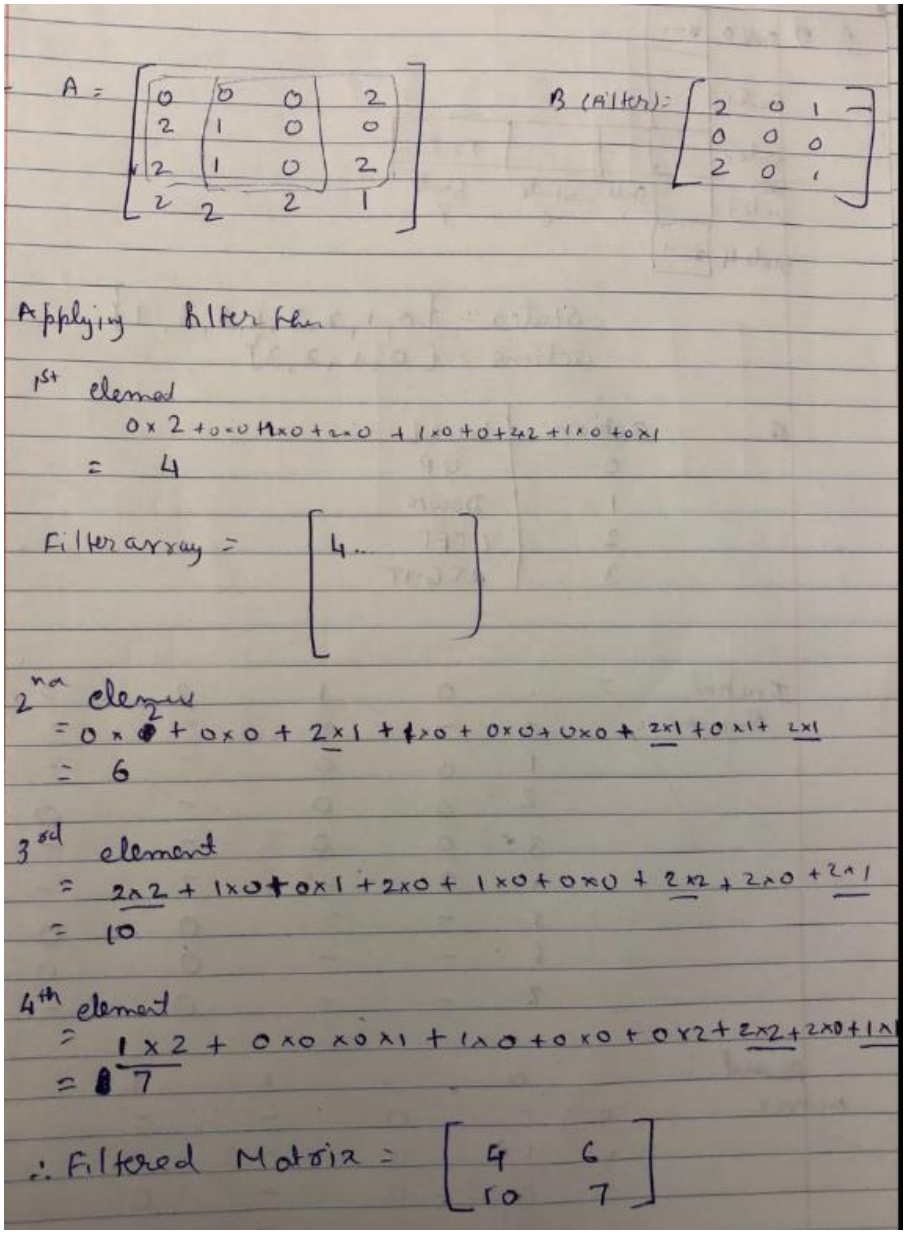


Fig1. Solution for Ans 2.

Ques. 3 What is a pooling operation in convolutional neural networks and why is this operation important?

Ans 3. The pooling operation is just taking the average or the maximum of the responses in a certain area of the filtered images which we have obtained through convolution. Thus, the pooling operation compress down the image by only considering the average or the maximal feature of the convuled image.

This is operation is important as it helps to obtain the limited nodes which can represent the class labels. These labels represent the compressed summary of the image.

Ques 4. What is a gated recurrent network? Name an example of such a neural network.

Ans 4. The basic recurrent network has some form of the memory of the previous states. This memory starts fading exponentially, this phenomenon can be referred as form of short-term memory of the network. The short-term memory may be not good for some applications. If this memory somehow can be stored and only gets in the picture when required. Hence, it is useful to gate some of these information (memory) until they are useful later in the process. This type of recurrent network which can gate the information for later (also known as long term memory) is known as gated recurrent network. The first network which has taken long-term memory in use is called LSTM which stands for Long Short-Term Memory.

Ques 5. In reinforcement learning, what is a policy?

Ans 5. The policy specifies what action need to be taken by the agent. In other word, it defines the behaviour of the agent in the environment.

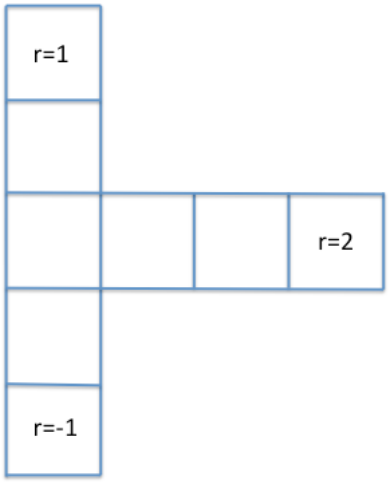
Ques. 6 Explain the difference between the SARSA and Q-Learning algorithm.

Ans 6. In q-learning algorithm, we have the knowledge of transition function and reward function. While in SARSA, we do not have knowledge of the transition function and reward function. Hence, in SARSA the agent has to explore the environment by acting in it and in Q-Learning the agent has not to explore the environment, sit here and calculate the optimal solution.

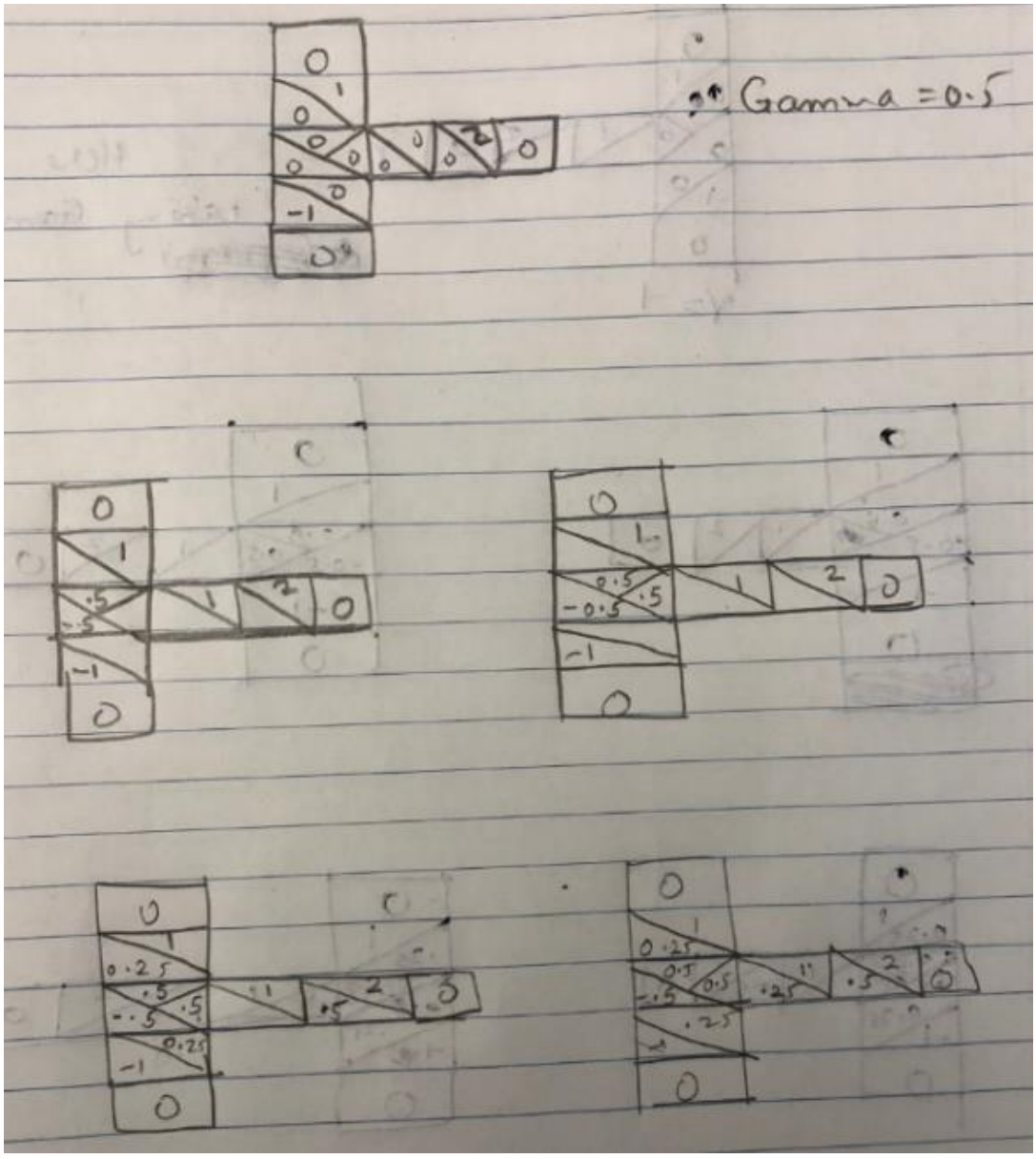
Ques. 7 Briefly explain `dropout' and why it is used in deep networks.

Ans 7. Dropout is a method in which some of the parameters of the model are set to zero during the output generation in the training process. It is used in deep networks as it prevents overfitting and helps to detect the underlying pattern of each class in the data by switching off some hidden nodes during training process. It is also used as regularization technique.

Ques 8. What are the optimal Q values for the T-maze below, assuming that we value diminishing returns with y=0.5?



Ans. 8



Solution for Ans 9.

References:

[1] Anish Singh Walia. 2017. The Vanishing Gradient Problem – Anish Singh Walia – Medium. (June 2017). Retrieved April 9, 2018 from https://medium.com/@anishsingh20/the-vanishing-gradient-problem-48ae7f501257

[2] Amar Budhiraja. 2016. Learning Less to Learn Better - Dropout in (Deep) Machine learning. (December 2016). Retrieved April 9, 2018 from https://medium.com/@amarbudhiraja/https-medium-com-amarbudhiraja-learning-less-to-learn-better-dropout-in-deep-machine-learning-74334da4bfc5