

## Assignment 9

Tuesday, 17. January 2023

21:38

Prathvish Mitharc - F028692

Subrat kishore Dutta - F028022.

### Exercise 9.1

a) i] The CNN architecture typically include convolution layers, pooling layers and fully connected layers. The feature map which is generated after applying convolution have important details about image like edges or shapes. The pooling layer usually reduce the dimensionality. And the final fully connected layer helps in prediction. Because of this kind of architecture the CNN can extract the features of images accurately, thus it is used in image recognition task.

ii] CNN are mainly used for image recognition, but they can also be used for NLP. In NLP we can use 1D convolutional filters which will help in extracting the features present in the text.

b) i]  $\text{Image}(I) = H * W * C$

$\text{Kernel}(K) = N * M$

$\text{Stride} = T * S$

Considering the number of kernels = 10

$$O_{o,ij} = \sum_{c=0}^{C-1} \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} I[C, x+T*i, y+S*j] \cdot K[C, x, y]$$

$$\begin{aligned} \text{ii] Output dimension} &= \frac{(N - F + 2P)}{S} + 1 \\ &= \frac{(32 - 5 + 2 \cdot 2)}{1} + 1 \\ &= \frac{(27 + 4)}{1} + 1 = 32 \end{aligned}$$

Where  $N = 32$  (Image dimension)

$F = 5$  (Filter size)

$P = 2$  (padding)

$S = 1$  (Stride)

$\therefore$  Output dimension =  $32 \times 32 \times 10$   
↳ No. of filters

iii] - Each filter has  $5 \times 5 \times 3 + 1 = 76$  parameters  
since there are 10 filters totally there are  
 $76 \times 10 = 760$  parameters

- The layer of the same size =  $32 \times 32 \times 3 \times 10$   
= 30720 parameters.

Wednesday, 11 January 2023

12:09 PM

### Exercise 9.1

(c)

given the effective receptive field  $R_k$  is given as

$$R_k = R_{k-1} + (K-1) \prod_{i=1}^{k-1} S_i$$

where  $k$  is the depth of the layer,  $K$  is the kernel size and  $S$  is the stride

1. Considering the total number of layers as 5 with kernel size  $K=5$  and stride  $S=1$ .

- layer 1st the receptive field will be  $R_0 = 5$ .
- layer 2nd the receptive field will be  $R_1 = 5 + 4 \cdot 1 = 9$
- layer 3rd the receptive field will be  $R_2 = 9 + 4 \cdot 1 = 13$
- layer 4th the receptive field will be  $R_3 = 13 + 4 \cdot 1 = 17$
- layer 5th the receptive field will be  $R_4 = 17 + 4 \cdot 1 = 21$

2. Adding pooling layer reduces the spatial resolution of the feature maps, making the receptive field larger. This is because pooling incorporates the output from multiple units into a single unit. This makes the receptive field of each of the unit larger as each unit aggregates information from a larger area.