

notes ref. inst. Deep Learning - Homework 2 all work (1/2) ED

Forward 1:
Input 1:

Part 2

Forward

backward

Input 2:

01. Channel 1:

$$\begin{matrix} 2 & 3 & 4 & 5 & 6 \\ 14 & 15 & 16 & 1 & 1 \\ 6 & 7 & 8 & 0 & 0 \\ 8 & 10 & 12 & 1 & 2 \\ 22 & 24 & 26 & 3 & 3 \end{matrix}$$

Channel 2:

$$\begin{matrix} 22 & 23 & 24 & 12 & 14 \\ 14 & 15 & 16 & 0 & 0 \\ 6 & 7 & 8 & 0 & 0 \\ 8 & 10 & 12 & 1 & 2 \\ 22 & 24 & 26 & 5 & 5 \end{matrix}$$

Output

For setting 1, we have $(5+0-3)+1 \Rightarrow 3 \times 2$ output

For setting 2, we have $(5+0-3)+1 \Rightarrow 2 \times 2$ output

Setting 1: Channel 1: (3×2) & Channel 2: (3×2)

$$\begin{aligned} \text{Out}_1 &= 22 + 25 + 28 = \text{softmax}(42, 14, 15, 16) \\ \text{Out}_2 &= 22 + 28 + 32 = 36 \Rightarrow \text{softmax}(6, 17, 8) \\ \text{Out}_3 &= 28 + 36 + 41 = 46 \Rightarrow \text{softmax}(8, 10, 12) \\ \text{Out}_4 &= 22 + 26 + 3 = 41 \Rightarrow \text{softmax}(14, 16) \end{aligned}$$

Setting 2: Channel 1:

Channel 2:

$$\begin{aligned} \text{Out}_1 &= 22 + 28 + 36 = \text{softmax}(14, 16, 12) \\ \text{Out}_2 &= 22 + 28 + 46 = \text{softmax}(14, 16, 12) \\ \text{Out}_3 &= 28 + 36 + 41 = 46 \Rightarrow \text{softmax}(14, 16, 12) \\ \text{Out}_4 &= 22 + 26 + 3 = 41 \Rightarrow \text{softmax}(14, 16, 12) \end{aligned}$$

Q2 Given output = $[0.5, 0.3, 0.2]^T$ and target

applying softmax to these values we get

$$f(x)_i = \frac{e^{x_i}}{\sum_{k=1}^K e^{x_k}}$$

$$\Rightarrow [0.190, 0.424, 0.3843]$$

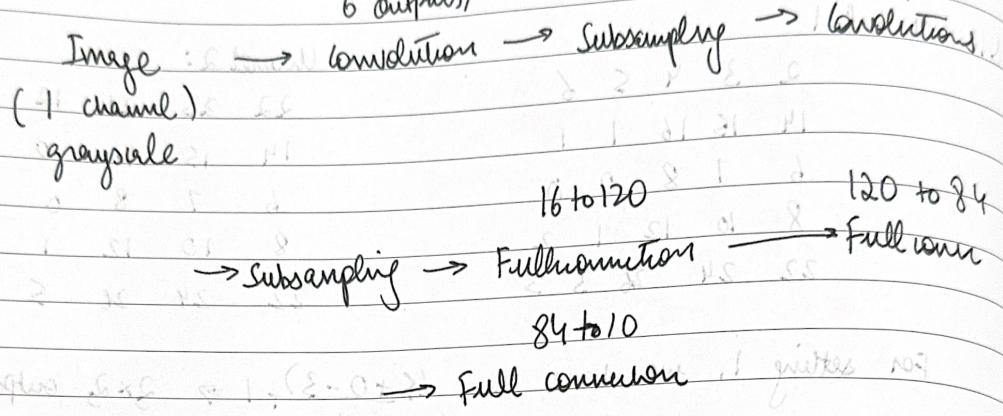
$$\text{Now, } L = -\log(0.424)$$

$$\Rightarrow -0.85$$

Q3 Let's draw the model in the form below for easier calculations

6 outputs

16 inputs / outputs



Now, we know subsampling layers don't have trainable parameters so we left with

$$\text{low layer 1 parameters} \Rightarrow (5 \times 5 \times 1 + 1) \times 6 = 156$$

$$\text{low layer 2 parameters} \Rightarrow (5 \times 5 \times 6 + 1) \times 16 = 2416 \text{ (approx.)}$$

$$(CS) \text{ full-connected layer 1} \Rightarrow (5 \times 5 \times 16 + 1) \times 120 = 48120$$

$$\text{fully-connected layer 2} \Rightarrow (120 \times 84) + 84 = 10164$$

$$\text{" " layer 3} \Rightarrow (84 \times 10) + 10 \Rightarrow 850$$

In addition to these, we have 2 bias (6×2) trainable parameters in the 1st fully connected layer, whereas the second fully connected layer has 850 trainable parameters. Adding everything up we get: 61,750 trainable parameters approximately.

61,750 trainable parameters approximately.

[888.0 , 900.0 , 911.0]

(PVC) sol = 1, with

28.0 - &