Adib Atkari, 40216815 ; A3 COMP 433

Question 1:

```
1 1 1 1 0
0 0 1 1 1        1 4 1
0 0 0 0 0        0 2 1
1 1 0 1 0        2 4 2
1 1 0 1 1
```

a) slide = 1

Output = $\frac{5-3}{2} + 1 = 3$

```
1 1 1 1 4 1      1+4+1+1 = 7.
0 0 1 0 2 1
0 0 0 2 4 2
```

```
1 1 1    1+4+1+2+1 = 9.
0 1 1
0 0 0
```

```
1 1 0    1+4+2+1 = 8.
1 1 1
0 0 0
```

7 9 8

```
0 0 1    1+2+4 = 7
0 0 0
1 1 0
```

```
0 1 1    4+1+2+2 = 9.
0 0 0
1 0 1
```

7 9 10

```
1 1 1    1+4+1+4 = 10.
0 0 0
0 1 0
```

0 0 0
1 1 0 $\quad 2+2+4=8$
1 1 0

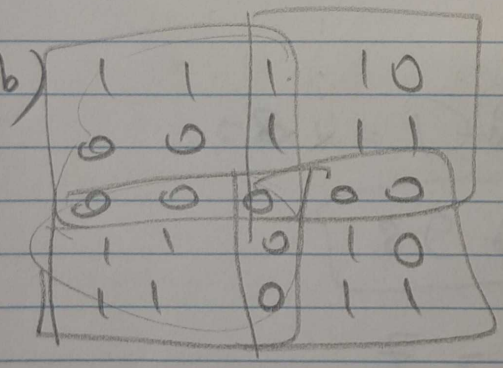
0 0 0
1 0 1 $\quad 1+2+2=5$
1 0 1

0 0 0
0 1 0 $\quad 2+4+2=8.$
0 1 1

output Q a

7	9	8
7	9	10
8	5	8

b)



1 4 1
0 2 1
2 4 2

output $= \frac{5-3}{2} + 1$
$= 2$

1 1 1
0 0 1 $\quad 1+4+1+1=7$
0 0 0

1 1 0
1 1 1 $\quad 1+4+2+1=8$
0 0 0

0 0 0
1 1 0
1 1 0

2+2+4=8

0 0 0
0 1 0
0 1 1

2+4+2=8

output Q6

8	8
8	8

c) 1 pixel Padding, stride=1

output size = $\frac{7-3}{1}+1=5$

0 0 0 0 0 0 0
0 1 1 1 1 0 0
0 0 0 1 1 1 0
0 0 0 0 0 0 0
0 1 1 0 1 0 0
0 1 1 0 1 1 0
0 0 0 0 0 0 0

1 4 1
0 2 1
2 4 2

0 0 0
0 1 1
0 0 0

2+1=3

0 0 0
1 1 1
0 1 1

2+1+4+2=9

0 0 0
1 1 1
0 0 1

2+1+0+2=5

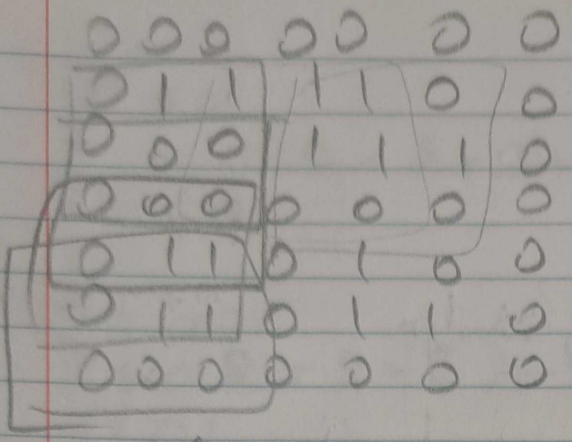
0 0 0
1 0 0
1 1 0

2+4=6

0 0 0
1 1 0
1 1 1

2+2+4+2=10

3, 5, 9, 10, 6



1 4 1
0 2 1
2 4 2

$\Sigma A+1 - 5$

1 1 1
0 0 1
0 0 0

$1+4+1+1 = 7.$

1 1 1
0 1 1
0 0 0

$2+1+1+4+1 = 9.$

1 1 0
1 1 1
0 0 0

$1+4+2+1 = 8.$

5, 7, 9, 8, 3

1 0 0
1 1 0
0 0 0

$1+2 = 3$

0 0 0
0 0 0
0 1 1

0+4=6

0 0 1
0 0 0
1 1 0

1+2+4=
7

0 1 1
0 0 0
1 0 1

9+1+2+2
=9.

1 1 1
0 0 0
0 1 0

1+4+1+4
=10

1 1 0
0 0 0
1 0 0

1+4+2=7

[6, 7, 9, 10, 7]

0 0 0
0 1 1
0 1 1

2+1+4+2
=9

0 0 0
1 1 0
1 1 0

0+2+2+4=8

0 0 0
1 0 0
1 1 0

2+4=6

0 0 0
1 0 1
1 0 1

1+2+2=5

0 0 0
0 1 0
0 1 1

2+4+2
=8

[9, 8, 5, 8, 6]

0 1 1 0 1 1 0 0 0	1 1 0 1 1 0 0 0 0	1 0 1 1 0 1 0 0 0	0 1 0 0 1 1 0 0 0
=8	1+4+2=7	5	4+2+1 =7

1 0 0
0 1 0
0 0 0

1+0+2=5

[8, 7, 3, 7, 3]

Max output Qc

3	5	9	10	6
5	7	9	8	3
6	7	9	10	7
9	8	5	8	6
8	7	3	7	5

d) Max Pooling Size = 2
stride = 2

1 1 1 1 0
0 0 1 1 1
0 0 0 0 0
1 1 0 1 0
1 1 0 1 1

output size =

$\frac{N + 2P - F}{S} + 1$

$= 2.5$

We Need to use $\lfloor 2.5 \rfloor$ aka 2x2

Output ouser Qd

1	1
1	1

e)
$$\begin{matrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 \end{matrix} \Rightarrow \begin{matrix} 1 & - & 1 & - & 0 \\ - & - & - & - & - \\ 0 & - & 0 & - & 0 \\ - & - & - & - & - \\ 1 & - & 0 & - & 1 \end{matrix} \quad \begin{matrix} 1 & 4 & 1 \\ 0 & 2 & 1 \\ 2 & 4 & 2 \end{matrix}$$

$1+4+2+2 = 9$

output $Qe = \boxed{9}$; 1×1

Question 2:

$$O = \left\lfloor \frac{I + 2P - K}{S} \right\rfloor + 1$$

We calculate for W & H at Same time

Conv0 1: $O = \left\lfloor \frac{64 + 2\times1 - 3}{1} \right\rfloor + 1 = 64$

Dimension: $64 \times 64 \times 32$

Conv0 2: $O = \left\lfloor \frac{64 + 0 - 7}{1} \right\rfloor + 1 = 58$

Dim: $58 \times 58 \times 64$

Poolip1: $O = \left\lfloor \frac{58 - 2}{2} \right\rfloor + 1 = 29$

Dim: $29 \times 29 \times 64$

Conv03: $O = \left\lfloor \frac{29+2(1)-3}{1} \right\rfloor + 1 = 29$

Channels: 128

Dim: $29 \times 29 \times 128$

Conv04: $O = \left\lfloor \frac{29+0-7}{1} \right\rfloor + 1 = 23$

Dim: $23 \times 23 \times 256$

Pool_p2: $O = \left\lfloor \frac{23-2}{2} \right\rfloor + 1 = 11$

Dim: $11 \times 11 \times 256$

m of Flatten: $11 \times 11 \times 256 = 30,976$

Dim of Dense = 10

Question 5:

$$\begin{matrix} -7 & 0 & 1 \\ 2 & 3 & 4 \\ 5 & 6 & 7 \\ 8 & 9 & 10 \end{matrix} \qquad \begin{matrix} 2 & 1 \\ 0 & -1 \end{matrix}$$

P=0
S=1
then Relu then 2×2 avg Pooling

Conv0 → Relu → Pooling

$-2 + (-3) = -5$

$1 - 4 = -3$

$2 \times 2 + 3 \times 1 - 6 = 1$

$3 \times 2 + 4 \times 1 - 7 = 3$

$5 \times 2 + 6 \times 1 + (-9) = 7$

$6 \times 2 + 7 \times 1 - 10 = 9$

output

-5	-3
1	3
7	9

Relu Operation: $\max(0, x)$

0	0
1	3
7	9

2×2 Pooling Avg $(0+0+1+3)/4 = 1$

output of all Ops = [1]

Question 7:

In a transformer, the self attention mechanism calculates the output of x_i as follows:

1) Query: $q_i = W^Q x_i \in \mathbb{R}^{d_K}$

Key: $k_i = W^K x_i \in \mathbb{R}^{d_K}$

Value: $v_i = W^V x_i \in \mathbb{R}^{d_V}$

q_1 is Query for x_1, k_j are all keys for all inputs, & v_j are the values containg the actual input data.

2) attention score: q_1 is compared with all keys to measure Relevance

$$\text{Score}(x_1, x_j) = \frac{q_1 \cdot k_j}{\sqrt{d_K}}$$

We Need to maximize $\text{score}(x_1, x_i)$

3) Softmax to get attention weights

$$\alpha_{ij} = \frac{\exp(\text{score}(x_i, x_j))}{\sum_l \exp(\text{score}(x_i, x_l))}$$

this gives how much x_i attends to each position in the sentence.

4) Weighted Sum of Values: $z_i = \sum_j \alpha_{ij} v_j$

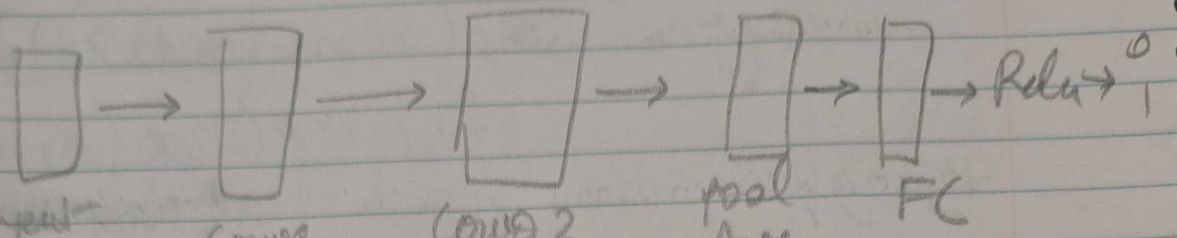
this aggregates info from all tokens based on Relevance to x_i.

Query → look for Relevant Info in Sequence of Keys.

then Gather Info (values) weighted by importance to produce a Context aware Representation z_i.

(12)

Question 3:



a) $H_{out} = \left\lfloor \frac{32+2\times1-3}{1} \right\rfloor + 1 = 32$

$W_{out} = 32$

Depth = 16

total Neuros = $32 \times 32 \times 16 = 16,384$ Neurons

Neuros = Height x width x Nb of Filters

b) Weights per filter = $3 \times 3 \times 3 = 27$

(3×3: filter size; ×3: RGB)

total W = $27 \times 16 = 432$

total B = 16

total Params = 448 params.

c) $H_{out} = \left\lfloor \frac{32 + 2\times0 - 3}{1} \right\rfloor + 1 = 30$

$W_{out} = 30$

Depth = 16 (filters)

Nb of Neurons: 14,900 Neurons.

d) $\underbrace{3\times3}_{\text{Filter size}} \times \underbrace{16}_{\text{Input depth}} = 144 \rightarrow$ weight / Filter

total W = 144 × 16 = 2304

total B = 16 (one per Filter)

total pa = 2320

e) $H_{out} = \left\lfloor \frac{(30-2)}{2} + 1 \right\rfloor = 15 = W_{out}$

Depth = 16

After Relu, No Chges

total features = 15 × 15 × 16 = 3600 feat

f) - Fewer Parameters
- More Non linearity
- better Feature Hierarchies
- Improved Gradients Flow.

Question 4:

1) flatten the CNN output into a 1D vector.

2) Apply logistic Regression: this will produce a prob for 2 classes

use if $\hat{y} \geq 0.5 \rightarrow$ class 1

else class 0.

Dimensions: $15 \times 15 \times 16 = 3600$ features

Weights: 3600

Bias: 1

total 3601 parameters

Question 6:

a) $h_t = w_{in} x_t + w_{hh} h_{t-1}$.

b) $x = [0.3, 0.4, 0.2]$, $h_0 = 0.5$, $w_{in} = w_{hh} = w_{hy} = 0.2$.

$h_1 = 0.2 \times 0.3 + 0.2 \times 0.5 = 0.16$

$h_2 = 0.2 \times 0.4 + 0.2 \times 0.16 = 0.112$

$h_3 = 0.2 \times 0.2 + 0.2 \times 0.112 = 0.0624$

$y = w_{hy} h_3 = 0.2 \times 0.0624 = 0.01248$

c) $l = \frac{1}{2}(y - y_r)^2$ let $e = y - y_r$

$$\frac{\partial l}{\partial w_{hy}} = (y - y_r) h_3 = e h_3.$$

$$h_3 = w_{in}(x_3 + w_{hh} x_2 + w_{hh}^2 x_1) + w_{hh}^3 h_0$$

$$\frac{\partial h_3}{\partial w_{in}} = x_3 + w_{hh} x_2 + w_{hh}^2 x_1 = \sum_{t=1}^{3} w_{hh}^{3-t} x_t$$

$$\frac{\partial l}{\partial w_{in}} = (y - y_r) w_{hy} \sum_{t=1}^{3} w_{hh}^{3-t} x_t$$

$$\frac{\partial h_3}{\partial w_{hh}} = w_{in}(x_2 + 2 w_{hh} x_1) + 3 w_{hh}^2 h_0$$

$$\frac{\partial l}{\partial w_{hh}} = (y - y_r) w_{hy} (w_{in}(x_2 + 2 w_{hh} x_1) + 3 w_{hh}^2 h_0)$$

d) $y_1 = 0.4$, $\eta = 0.15$.

$h_3 = 0.0624$, $y = 0.01248$, $e = -0.38752$.

$$\frac{\partial l}{\partial w_{hy}} = -0.024 18 = e.h_3$$

$$\frac{\partial l}{\partial w_{ih}} = e.w_{hy}(x_3 + w_{hh}x_2 + w_{hh}^2 x_1)$$

$$= -0.38752 \times 0.2(0.2 + 0.2 \times 0.4 + 0.2^2 \times 0.3)$$

$$= -0.02963$$

$$\frac{\partial l}{\partial w_{hh}} = e.w_{hy}(w_{ih}(x_2 + 2w_{hh}x_1) + 3w_{hh}^2 h_0)$$

$$= -0.38752 \times 0.2(0.2(0.4 + 0.2 \times 2.0.3) + 3 \times 0.2^2 \times 0.5)$$

$$= -0.027606$$

$$w_k^{New} = w_k^{old} - \eta \frac{\partial L}{\partial w_k}$$

$$w_{ih}^{New} = 0.2 - 0.15(-0.0226\ldots)$$

$$= 0.20339$$

$$w_{hh}^{New} = 0.901906$$

$$w_{hy}^{New} = 0.203627$$