Machine Learning

Homework 5

Karan Sunil Kumbhar
Id. - 12140860
BTech CSE
2025

CS550 Machine Learning

November 30, 2023

Problem 1

Suppose images are 224x224x3 and we use 64 convolutional filters that are 3x3.

(a) How many responses will be computed for this layer of a CNN (stride=1, no padding)?

(b) How much zero padding is necessary to produce an output of size equal to the input?

(c) Repeat 'a' for the case when the stride is 3.

Solution.

(a) To calculate the number of responses (feature maps) produced by a convolutional layer in a CNN, we use the formula:

$$\text{Number of Responses} = \left(\frac{\text{Input Size} - \text{Filter Size}}{\text{Stride}} + 1\right)^2 \times \text{Number of Filters}$$

Given the information:

$$\begin{aligned} \text{Input Size} &: 224 \times 224 \times 3 \text{ (width, height, channels)} \\ \text{Filter Size} &: 3 \times 3 \\ \text{Stride} &: 1 \\ \text{Number of Filters} &: 64 \end{aligned}$$

Substitute these values into the formula:

$$\begin{aligned} \text{Number of Responses} &= \left(\frac{224 - 3}{1} + 1\right)^2 \times 64 \\ &= 222 \times 222 \times 64 \\ &= 49284 \times 64 \\ &= 3,154,176 \end{aligned}$$

Therefore, the number of responses (feature maps) computed for this layer of the CNN is $3,154,176$.

(b) To determine the zero padding necessary to produce an output of the same size as the input, we use the formula:

$$\text{Output Size} = \frac{\text{Input Size} - \text{Filter Size} + 2 \times \text{Padding}}{\text{Stride}} + 1$$

Given the information:

$$\begin{aligned} \text{Input Size} &: 224 \times 224 \text{ (width, height)} \\ \text{Filter Size} &: 3 \times 3 \\ \text{Stride} &: 1 \\ \text{Output Size} &: 224 \times 224 \end{aligned}$$

Substitute these values into the formula and solve for Padding:

$$\begin{aligned} 224 &= \frac{224 - 3 + 2 \times \text{Padding}}{1} + 1 \\ 224 &= 221 + 2 \times \text{Padding} + 1 \end{aligned}$$

Solving for Padding:

$$\begin{aligned} 2 \times \text{Padding} &= 224 - 221 - 1 \\ 2 \times \text{Padding} &= 2 \\ \text{Padding} &= \frac{2}{2} \\ &= 1 \end{aligned}$$

Therefore, the necessary zero padding to produce an output of the same size as the input is 1 **in each dimension.**

(c) To calculate the number of responses when the stride is 3, we use the formula:

$$\text{Number of Responses} = \left(\frac{\text{Input Size} - \text{Filter Size}}{\text{Stride}} + 1\right)^2 \times \text{Number of Filters}$$

Given the information:

$$\begin{aligned}
\text{Input Size} &: 224 \times 224 \text{ (width, height)} \\
\text{Filter Size} &: 3 \times 3 \\
\text{Stride} &: 3 \\
\text{Number of Filters} &: 64
\end{aligned}$$

Substitute these values into the formula:

$$\begin{aligned}
\text{Number of Responses} &= \left(\frac{224-3}{3} + 1\right)^2 \times 64 \\
&= (73.67 + 1) \times 64
\end{aligned}$$

here we can see dimension is fractional so here we can do 2 thing eighter discard last 2 columns and 2 rows or another way add proper padding

- discarding extra dimension(taking $73.67 \approx 73$), we get

$$\begin{aligned}
\text{Number of Responses} &= (73 + 1) \times 64 \\
&= 74 \times 74 \times 64 \\
&= 5476 \times 64 \\
&= 350464
\end{aligned}$$

- Adding proper padding (taking $73.67 \approx 74$), we get

$$\begin{aligned}
\text{Number of Responses} &= (74 + 1) \times 64 \\
&= 75 \times 75 \times 64 \\
&= 5625 \times 64 \\
&= 360000
\end{aligned}$$

Therefore, the number of responses computed for this layer of the CNN with a stride of 3 is $350,464$(when discard extra dimensions) or 360000 (adding proper padding).

Problem 2

Assume that the inputs are single bits 0 (white) and 1 (black). Consider a 3x3 filter, whose weights are w_{ij}, for $0 \le i \le 2$ and $0 \le j \le 2$ and whose bias is b. Suggest weights and bias so that the output of this filter will detect the following simple features.

(a) Vertical boundary, where the left column is 0, and the other two columns are 1.

(b) A diagonal boundary, where only the triangle of three pixels in the upper right corner are 1

(c) A corner, in which the 2x2 square in the lower right is 0 and the other pixels are 1.

Solution.

(a) To detect a vertical boundary where the left column is 0 and the other two columns are 1, we can use the following filter:

$$\text{Weight Matrix:} \quad \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$$

$$\text{Bias:} \quad -5$$

When this filter is applied to an input matrix representing the pattern, it should produce a high value(in this case high value is 6), indicating the detection of the vertical boundary. For all other pattern it will produce less value than 6. And when we add bias -5 to it will give value as 1 or less than 1. And then we will apply Relu activation function to get desired output(pixel value 1 correspoding to required pattern and others values are zeros).

(b) To detect a diagonal boundary where only the triangle of three pixels in the upper right corner is 1, we can use the following filter:

$$\text{Weight Matrix:} \quad \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

$$\text{Bias:} \quad -2$$

When this filter is applied to an input matrix representing the pattern, it should produce a high value as 3, indicating the detection of the diagonal boundary in the upper right corner. In all other cases it is less than 3. And then when we add bias term to it will product values one or less than one. And after applying Relu activation function it will produce one corresponding to required pattern and other values are zero.

(c) To detect a corner where the 2x2 square in the lower right is 0, and the other pixels are 1, you can use the following filter:

$$\text{Weight Matrix:} \quad \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

$$\text{Bias:} \quad -4$$

When this filter is applied to an input matrix representing the pattern, it should produce a high value, indicating the detection of the corner with highest value as 5 and after adding bias it will produce 1 or less than one. And then applying Relu activation function we will get 1 values corresponding to corner and zero correspoding to lower right square. We also can apply logical not to output after applying Relu function to get high values correspoding to right lower square

Problem 3

In this exercise, you are asked to design the input weights for one or more nodes of the hidden state of an RNN. The input is a sequence of bits, 0 or 1 only. Note that you can use other nodes to help with the node requested. Also note that you can apply a transformation to the output of the node so a "yes" answer has one value and a "no" answer has another.

a) A node to signal when the input is 1 and the previous input is 0

b) A node to signal when the last three inputs have all been 1

c) A node to signal when the input is the same as the previous input

Solution.

a) We can create state table as follows

ht-1	xt	ht
0	0	1
1	0	0
2	0	2
0	1	0
1	1	2
2	1	2

Table 1: State machine table

The transition rule is given by:

$$h_{t+1} = \begin{cases} 1 & \text{if } x_t + h_{t-1} = 0 \\ 0 & \text{if } x_t + h_{t-1} = 1 \\ 2 & \text{if } x_t + h_{t-1} \geq 2 \end{cases}$$

b) We can create state table as follows

ht-1	xt	ht
0	0	0
1	0	0
2	0	0
3	0	3
0	1	1
1	1	2
2	1	3
3	1	3

Table 2: State machine table

The transition rule is given by:

$$h_{t+1} = \begin{cases} 3 & \text{if } h_t + 1 \geq 3 \\ (h_t + 1) \cdot (x_t) & \text{if } h_t + 1 < 3 \end{cases}$$

Problem 4

Let A be a matrix of 4-dimensional data points

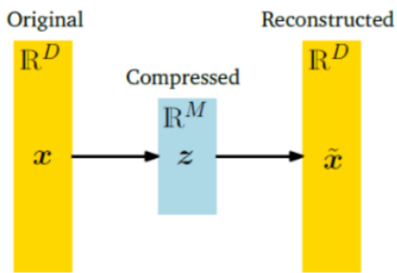
$$A = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 9 \\ 4 & 16 \end{bmatrix}$$

a) Compute Eigenpairs for $A^T A$?

b) What do you expect the eigenvalues of AA^T to be?

c) Find the eigenvectors of AA^T, using the eigenvalues from part (c).

d) Write the 1-dimensional and 2-dimensional encodings (z) of columns of A using PCA.



Solution.

a) To compute the eigenpairs for $A^T A$, where A is a matrix of 4-dimensional data points given by

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 9 \\ 4 & 16 \end{bmatrix}$$

we follow these steps:

1. **Calculate A^TA:**

$$A^TA = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 4 & 9 & 16 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 9 \\ 4 & 16 \end{bmatrix}$$
$$= \begin{bmatrix} 1+4+9+16 & 1+8+27+64 \\ 1+8+27+64 & 1+16+81+256 \end{bmatrix}$$
$$= \begin{bmatrix} 30 & 100 \\ 100 & 354 \end{bmatrix}$$

2. **Find Eigenvalues (λ):** The characteristic equation is given by:

$$|A^TA - \lambda I| = \begin{vmatrix} 30-\lambda & 100 \\ 100 & 354-\lambda \end{vmatrix}$$
$$= (30-\lambda)(354-\lambda) - 100 \times 100 \quad = 0$$
$$0 = 30 \times 354 + \lambda^2 - (354+30) \times \lambda - 10000$$
$$0 = \lambda^2 - 384\lambda + 620$$

Solving this quadratic equation will give us the eigenvalues.

$$(\lambda^2 - 384\lambda + 620) = 0$$

This equation can be solved using the quadratic formula:

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

where $a = 1$, $b = -384$, and $c = 620$. Substituting these values:

$$\lambda = \frac{384 \pm \sqrt{(-384)^2 - 4 \times 1 \times 620}}{2 \times 1}$$

$$\lambda = \frac{384 \pm \sqrt{147456 - 2480}}{2}$$

$$\lambda = \frac{384 \pm \sqrt{144976}}{2}$$

$$\lambda = \frac{384 \pm 380.75}{2}$$

So, we have two possible solutions:

$$\lambda_1 = \frac{384 + 380.75}{2}$$
$$= 382.375$$
$$\lambda_2 = \frac{384 - 384.75}{2}$$
$$= 1.625$$

Given the eigenvalue equation $A^T A\mathbf{v} = \lambda\mathbf{v}$ with

$$A^T A = \begin{bmatrix} 30 & 100 \\ 100 & 354 \end{bmatrix}$$

the calculated eigenvalues and corresponding eigenvectors are as follows:

1. For $\lambda = 1.625$, the corresponding eigenvector is any non-zero multiple of $\begin{bmatrix} 100 \\ -28.375 \end{bmatrix}$.

2. For $\lambda = 382.375$, the corresponding eigenvector is any non-zero multiple of $\begin{bmatrix} 100 \\ 352.375 \end{bmatrix}$.

Therefore, the eigenpairs are (1.625, $\begin{bmatrix} 100 \\ -28.375 \end{bmatrix}$) and (382.375, $\begin{bmatrix} 100 \\ 352.375 \end{bmatrix}$).

b) To find the eigenvalues of AA^T, where A is the given matrix

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 9 \\ 4 & 16 \end{bmatrix}$$

we can use the property that the eigenvalues of AA^T are the same as the eigenvalues of $A^T A$. From our previous discussion, we found that the eigenvalues of $A^T A$ are 1.625 and 382.375. Therefore, the expected eigenvalues of AA^T are also 1.625 and 382.375.

c) Given the matrix AA^T:

$$AA^T = \begin{bmatrix} 2 & 6 & 12 & 20 \\ 6 & 20 & 42 & 72 \\ 12 & 42 & 90 & 156 \\ 20 & 72 & 156 & 272 \end{bmatrix}$$

- For the eigenvalue $\lambda = 1.625$, the corresponding eigenvector $\mathbf{v}$ satisfies the system of equations:

$$A \times a^T \times v = \lambda \times v$$

$$\begin{bmatrix} 2 & 6 & 12 & 20 \\ 6 & 20 & 42 & 72 \\ 12 & 42 & 90 & 156 \\ 20 & 72 & 156 & 272 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = 1.625 \times \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix}$$

so value of v is non-zero multiple of $\begin{bmatrix} -1.32 \\ -1.60 \\ -0.82 \\ 1 \end{bmatrix}$

- For the eigenvalue $\lambda = 382.375$, the corresponding eigenvector $\mathbf{v}$ satisfies the system of equations:

$$A \times a^T \times v = \lambda \times v$$

$$\begin{bmatrix} 2 & 6 & 12 & 20 \\ 6 & 20 & 42 & 72 \\ 12 & 42 & 90 & 156 \\ 20 & 72 & 156 & 272 \end{bmatrix} \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix} = 382.375 \times \begin{bmatrix} w \\ x \\ y \\ z \end{bmatrix}$$

so value of v is non-zero multiple of $\begin{bmatrix} -0.37 \\ 0.27 \\ 0.57 \\ 1 \end{bmatrix}$

d)
- 1-dimensional encoding (z) of columns of A using PCA = A. for first value of v

$$A^T \times A = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 9 \\ 40 & 16 \end{bmatrix} \times \begin{bmatrix} 100 \\ -28.375 \end{bmatrix} = \begin{bmatrix} 71.625 \\ 86.5 \\ 44.625 \\ -54 \end{bmatrix}$$

- 2-dimensional encoding (z) of columns of A using PCA = A. eigenvectors of

$$A^T \times A = \begin{bmatrix} 1 & 1 \\ 2 & 4 \\ 3 & 9 \\ 4 & 16 \end{bmatrix} \times \begin{bmatrix} 100 & 100 \\ -28.375 & 352.375 \end{bmatrix}$$

$$= \begin{bmatrix} 71.625 & 452.375 \\ 86.5 & 1609.5 \\ 44.625 & 3471.375 \\ -54 & 6038 \end{bmatrix}$$

Problem 5

Design a CNN that can detect the face in the given input image

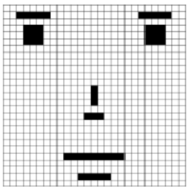


Figure 2 Face Image

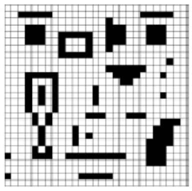


Figure 1 Input Image

a) What strategy will you use? How many layers? How many filters?

b) Describe the filters for each layer in detail.

c) Show how the convolutions will work and demonstrate that your CNN will be able to detect this pattern.

Solution.

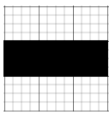
a) Strategy, Layers and Filters

- We can create small features to detect each part of face and then we will combine this features to get or detect complete face
- we will mainly use 3 layers which will describe in below part solution
- we will mainly taking 5 filter in first layer and then to get combine part like nose, eyes, mouth we again use 3 filters. And then we can get eighter full face or partial face(here while creating filters, we are treating black part as 1 and white part as 0. And to to detect part required we are putting high value correspondig to it)

b) Filters and Layers details

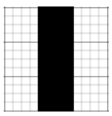
layer 1 :-

Filter 1 : Horizontal filter(denoted by letter H)



This filter will detect horizontal blocks of size 3×1 which belongs may be mouth, nose or eyes

Filter 2 : Vertical filter(denoted by letter V)

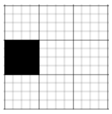


This filter will detect Vertical blocks of size 1×3 which belongs may be nose

Filter 3 : Block filter(denote by letter B)

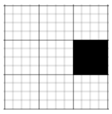
This filter will detect blocks of size 3×3 which belongs may be eyes

Filter 4 : Left filter(denote by letter L)



This filter will detect blocks of size 1×1 which belongs may be eye part or mouth

Filter 5 : Right filter(denote by letter R)



This filter will detect blocks of size 1×1 which belongs may be eye part or mouth

layer 2 :-

Filter 1 : Nose filter(denoted by letter N)

-	-	-
-	V	-
-	H	-

This filter will detect Nose part of face (combination of vertical and horizontal block which are detected in first layer)

Filter 2 Mouth filter: (denoted by letter M)

-	-	-
H	H	H
L	H	R

This filter will detect Mouth part of face (Combination of two horizontal strips. First one is of size 1×9 which is combination of again three horizonatal filter. And another below one strip is of size 1×5 which is combination of left + horizonatal + right filter)

Filter 3 : Eyes filter(denote by letter E)

L	H	L
-	B	-
-	-	-

This filter will detect eyes part of face(This is combination of eyebrow and block feature. Here eyebrow part is combination of left + horizontal + right And Block is feature detected in layer 1)

layer 3 :-

Filter 1 : Full face(denoted by letter F)

E	-	E
-	N	-
-	M	-

This filter will detect full face (combination of two eyes, one nose and mouth)

Filter 2 Mouth filter: (denoted by letter M)

-	E	-
-	-	N
-	-	M

This filter will detect partial face (Combination of of one nose, mouth and one eye)

c) Demostration