Keerti P. Charantimath 19MA20059

Assignment 1 - LA for AI and ML

(Q1) a) $\operatorname{avg}(\alpha x + \beta 1_n) = \frac{1}{n} 1_n^T (\alpha x + \beta 1_n)$

$= \frac{1}{n}(\alpha 1_n^T x) + \frac{1}{\cancel{n}} \beta (\cancel{n})$

$= \alpha \operatorname{avg}(x) + \beta$

b) $\operatorname{std}(\alpha x + \beta 1_n) = \frac{\| \alpha x + \beta 1_n - \operatorname{avg}(\alpha x + \beta 1_n) 1_n \|_2}{\sqrt{n}}$

$= \frac{\| \alpha x + \cancel{\beta 1_n} - (\alpha \operatorname{avg}(x) + \cancel{\beta}) 1_n \|_2}{\sqrt{n}}$

$= \frac{\| \alpha x - \alpha \operatorname{avg}(x) 1_n \|_2}{\sqrt{n}}$

$= |\alpha| \frac{\| x - \operatorname{avg}(x) 1_n \|_2}{\sqrt{n}}$

$= |\alpha| \operatorname{std}(x)$

c) We know that

$$std(x) = \frac{\|x - avg(x) 1_n\|_2}{\sqrt{n}}$$

$$\therefore std^2(x) = \frac{1}{n}(x - avg(x) 1_n)^2 = \sum_{i=1}^{n} \frac{1}{n}(x_i - avg(x))^2$$

It is given that $\exists$ k elements such that $|x_i - avg(x)| \geq a$

$$\therefore std^2(x) \geq \frac{1}{n}\left(\sum_{k} a^2 + \sum_{n-k}(x_i - avg(x))^2\right)$$

as $(x_i - avg(x))^2 \geq 0$ for all x

$$std^2(x) \geq \frac{1}{n} k a^2 \Rightarrow \left[\frac{std(x)}{a}\right]^2 \geq \frac{k}{n}$$

d) $\|x\|_w = \sqrt{\sum_{i=1}^{n} w_i x_i^2}$

we know $\sqrt{x} \geq 0 \quad \forall x \in \mathbb{R}$

$\therefore \|x\|_w \geq 0 \quad \forall \ x \in \mathbb{R}^n$

Suppose

$$\|x\|_w = 0 \Rightarrow \sqrt{\sum_{i=1}^{n} w_i x_i^2} = 0$$

$$\Rightarrow w_i x_i^2 = 0 \quad \forall x_i \quad \text{as } x_i^2 \geq 0$$

$$\Rightarrow x_i = 0 \quad \forall x_i \quad (\text{as } w_i \text{ is constant})$$

$$\therefore \|x\|_w = 0 \Rightarrow x = 0$$

Now

$$\|\alpha x\|_w = \sqrt{\sum_{i=1}^{n} w_i (\alpha x_i)^2} = \sqrt{\sum_{i=1}^{n} \alpha^2 w_i x_i^2} = |\alpha| \sqrt{\sum_{i=1}^{n} w_i x_i^2}$$

$$= |\alpha| \, \|x\|_w \quad \text{where } \alpha \in \mathbb{R}$$

consider

$$\cancel{\|x\|_w^2 = \sum_{i=1}^{n} w_i x_i^2}$$

$$\|x+y\|_w^2 = \sum_{i=1}^{n} (x_i+y_i)^2 w_i$$

$$= \sum_{i=1}^{n} w_i (x_i^2 + y_i^2 + 2 x_i y_i)$$

$$= \|x\|_w^2 + \|y\|_w^2 + 2 \sum_{i=1}^{n} w_i x_i y_i$$

Now

$$\|x\|_w \|y\|_w = \left(\sum_{i=1}^{n} w_i x_i^2\right)^{1/2} \left(\sum_{i=1}^{n} w_i y_i^2\right)^{1/2}$$

we know that

$$\sum_{i=1}^{n} w_i x_i y_i \leq \left(\sum_{i=1}^{n} w_i x_i^2\right)^{1/2} \left(\sum_{i=1}^{n} w_i y_i^2\right)^{1/2} = \|x\|_w \|y\|_w$$

$$\|x+y\|_w^2 = \|x\|_w^2 + \|y\|_w^2 + 2 \sum_{i=1}^{n} w_i x_i y_i$$

$$\leq \|x\|_w^2 + \|y\|_w^2 + 2\|x\|_w \|y\|_w$$

$$= (\|x\|_w + \|y\|_w)^2$$

$$\therefore \|x+y\|_w \leq \|x\|_w + \|y\|_w$$

Hence proved

2019						JULY
S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

23 Friday

3) $z = (A+B)(x+y)$ $A, B \in \mathbb{R}^{m \times n}$, $x, y \in \mathbb{R}^n$

a) add A and B → $(m \times n)$ computations
add x and y → (n) computations
Multiply $(A+B)(x+y)$ → $m(2n-1)$ computations

Total :- $mn + n + 2mn - m$

b) compute Ax → $m(2n-1)$ computations
compute Ay → $m(2n-1)$ computations
compute Bx → $m(2n-1)$ computations
compute By → $m(2n-1)$ computations
Add Ax, ~~and~~ Bx, ~~and~~ Ay and By → $3m$ computations

24 Saturday

Total :- $4m(2n-1) + 3m$

We ~~are~~ need $mn + n + 2mn - m > 4m(2n-1) + 3m$

$\Rightarrow \cancel{3mn} + n - m > \cancel{8}^{5}mn - 4m + 3m$

$\Rightarrow n - \cancel{m} > 5mn - \cancel{m}$

$\Rightarrow \cancel{n} > 5m\cancel{n}$

$\Rightarrow 1 > 5m$

$\Rightarrow \frac{1}{5} > m$

We know $m, n \geq 0$ &

∴ The above condition is never achievable
Thus approach 1 is always better than approach 2

5) Symmetric $x \to x_k = x_{n-k+1}$
Antisymmetric $x \to x_k = -x_{n-k+1}$ $\Big\}$ $x \in \mathbb{R}^n$

Let y be such that
$y_k = x_{n-k+1}$, $y \in \mathbb{R}^n$

Consider
$a = \cancel{x_k + y_k}\ x + y$

$\Rightarrow a_k = x_k + y_k = x_k + x_{n-k+1}$
and

$a_{n-k+1} = x_{n-k+1} + y_{n-k+1} = x_{n-k+1} + x_{n-(n-k+1)+1}$
$= x_{n-k+1} + x_k$

$\therefore a_k = a_{n-k+1}$

$\therefore a$ is symmetric

Consider $b = x - y$

$\Rightarrow b_k = x_k - y_k = x_k - x_{n-k+1}$
and
$b_{n-k+1} = x_{n-k+1} - y_{n-k+1} = x_{n-k+1} - x_k$

$\therefore b_k = -b_{n-k+1}$

$\therefore b$ is antisymmetric

Now, $\frac{1}{2}(a+b) = \frac{1}{2}(x+y+x-y) = x$

$\therefore x = \frac{1}{2}a + \frac{1}{2}b$ where $\frac{1}{2}a$ is symmetric and $\frac{1}{2}b$ is antisymmetric

S	M	T	W	T	F	S
	1	2	3	4	5	6
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
28	29	30	31			

6) $A \in \mathbb{R}^{n \times m}$

Let columns of A be linearly independent ie

$Ax = 0 \Rightarrow x = 0 \Rightarrow Ix = 0 \Rightarrow$ there exists a C such that

$CA = I$.

$\therefore$ if A has linearly independent columns, it also has a left inverse

Conversely, assume that A has a left inverse C but it has linearly dependent columns

ie $Ax = 0$ for some $x \neq 0$

$\Rightarrow CAx = 0 \Rightarrow Ix = 0 \Rightarrow x = 0$

This contradicts our assumptions.

Thus A has left inverse iff A has linearly independent columns

7) $A = \begin{bmatrix} I_n & x \\ x^T & 0 \end{bmatrix}$

If A is invertible, then columns of ~~et~~ A are linearly independent, i.e

$Ay = 0 \Rightarrow y = 0$ where $y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \in \mathbb{R}^{2n}$

$\Rightarrow I_n y_1 + x y_2 = 0 \Rightarrow y_1 = -x y_2 \rightarrow$ (a)

$x^T y_1 + 0 y_2 = 0 \Rightarrow x^T y_1 = 0 \rightarrow$ (b)

y_1 and y^2 ~~cannot~~ have to be zero simultaneously

~~If $x = 0 \Rightarrow y_1 = 0$ and y_2 can~~ be anything

$\therefore$ condition on x is $\boxed{x \neq 0}$ as if $x \neq 0$, $y_1 = 0$ & $y_2 = 0$

$\therefore$ for A to be invertible $\boxed{x \neq 0}$ is the condition

If inverse exists, then for some
$Ay = b \Rightarrow y = A^{-1} b$

ie $\begin{bmatrix} I_n & x \\ x^T & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$

$I_n y_1 + x y_2 = b_1$, $x^T y_1 = b_2$

$$\Rightarrow y_1 = b_1 - x y_2$$

Substituting in $x^T y_1 = b_2$

$$x^T(b_1 - y_2 x) = b_2$$
$$x^T b_1 - y_2 \|x\|^2 = b_2$$
$$y_2 = \frac{x^T b_1 - b_2}{\|x\|^2}$$

$$\therefore y_1 = b_1 - x\left(\frac{x^T b_1 - b_2}{\|x\|^2}\right) x$$

$$\Rightarrow y_1 = \frac{\left(\|x\|^2 I_n - x x^T\right) b_1 + x b_2}{\|x\|^2}$$

writing in matrix format

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \frac{1}{\|x\|^2} \begin{bmatrix} \|x\|^2 I_n - x x^T & x \\ x^T & -1 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$\therefore A^{-1} = \frac{1}{\|x\|^2} \begin{bmatrix} \|x\|^2 I_n - x x^T & x \\ x^T & -1 \end{bmatrix}$$

8) $A \in \mathbb{R}^{m \times n}$ with linearly independent columns
$\Rightarrow A$ has a ~~ft~~ left inverse

if $\hat{x} \in \mathbb{R}^n$ is a solution, then the left inverse is unique

$$\Rightarrow \hat{x} = (A^T A)^{-1} A^T b \Rightarrow (A^T A)\hat{x} = A^T b \longrightarrow (1)$$

where $(A^T A)^{-1} A^T$ is the left inverse of A known as pseudo inverse

let $y \in \mathbb{R}^n$
Then

$$(Ay)^T b = y^T A^T b = y^T A^T A \hat{x} = (Ay)^T (A\hat{x})$$

let $y = \hat{x} \Rightarrow (A\hat{x})^T b = (A\hat{x})^T (A\hat{x}) = \|A\hat{x}\|_2^2$

$$\therefore \frac{(A\hat{x})^T b}{\|A\hat{x}\|_2 \|b\|_2} = \frac{\|A\hat{x}\|_2^2}{\|A\hat{x}\|_2 \|b\|_2} = \frac{\|A\hat{x}\|_2}{\|b\|_2}$$

2019 AUGUST

S	M	T	W	T	F	S
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	31

9)

$$y_t \equiv \hat{y}_t = \sum_{j=1}^{n} u_{t-j+1} h_j \quad , \quad t = \{1, 2 \ldots T\}$$

$$\hat{y}_1 = \sum_{j=1}^{n} h_j u_{2j} = h_1 u_1$$

$$\hat{y}_2 = h_2 u_2 + h_2 u_1$$

$$\vdots$$

$$\hat{y}_T = h_1 u_T + h_2 u_{T-1} + \cdots + h_n u_{T+1-n}$$

$$\therefore \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ | \\ \hat{y}_T \end{bmatrix} = \begin{bmatrix} u_1 & 0 & 0 & \cdots & 0 \\ u_2 & u_1 & 0 & \cdots & 0 \\ u_3 & u_2 & u_1 & & \vdots \\ \vdots & & & & 0 \\ u_T & u_{T-1} & \cdots & & u_{T+1-n} \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ | \\ h_n \end{bmatrix}$$

$$b = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \vdots \\ y_T \end{bmatrix}$$

$$\therefore \|Ah - b\|_2^2 = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \cdots + (y_T - \hat{y}_T)^2$$

$\therefore$ A is a Toepliz matrix

10) a)

for every x_i

→ calculation of distance from 1 representative → $(n+n+n-1+1) = 3n+1$ operations
(we subtract each dim, square it, ~~and~~ add each of them)
∴ $3kn$ operations for ~~all~~ distances from all representatives to x_i
(we also take root at the end)

Calculating minimum → k operations

Total operations for updation of $x_i = 3kn+k$
" " " " " of all x_i = $(3kn+k)N$

b) To update the cluster representative, we add the every x_i once and each x_i has n dimensions, hence, we have total of $n(N)$ additions and as we have k representatives, we average the value of each dimension k times. Hence, there are $(n \times k)$ divisions

Total operations ~~$= (N \times n) + (n \times k)$~~ $= n(N+k)$
$= Nn + nk$

For 10 iterations, the total number of operations performed are
$10(n(N+k) + (3kn+k)N)$

Question 4

Part (a)

Array A

```
A

array([[1, 0, 0, 0, 1, 1],
       [0, 0, 1, 1, 1, 0],
       [1, 1, 0, 0, 1, 0],
       [1, 0, 1, 1, 1, 0],
       [0, 1, 1, 0, 0, 0],
       [0, 0, 0, 0, 1, 0],
       [1, 1, 0, 1, 0, 1],
       [1, 0, 0, 1, 0, 1]])
```

Array B

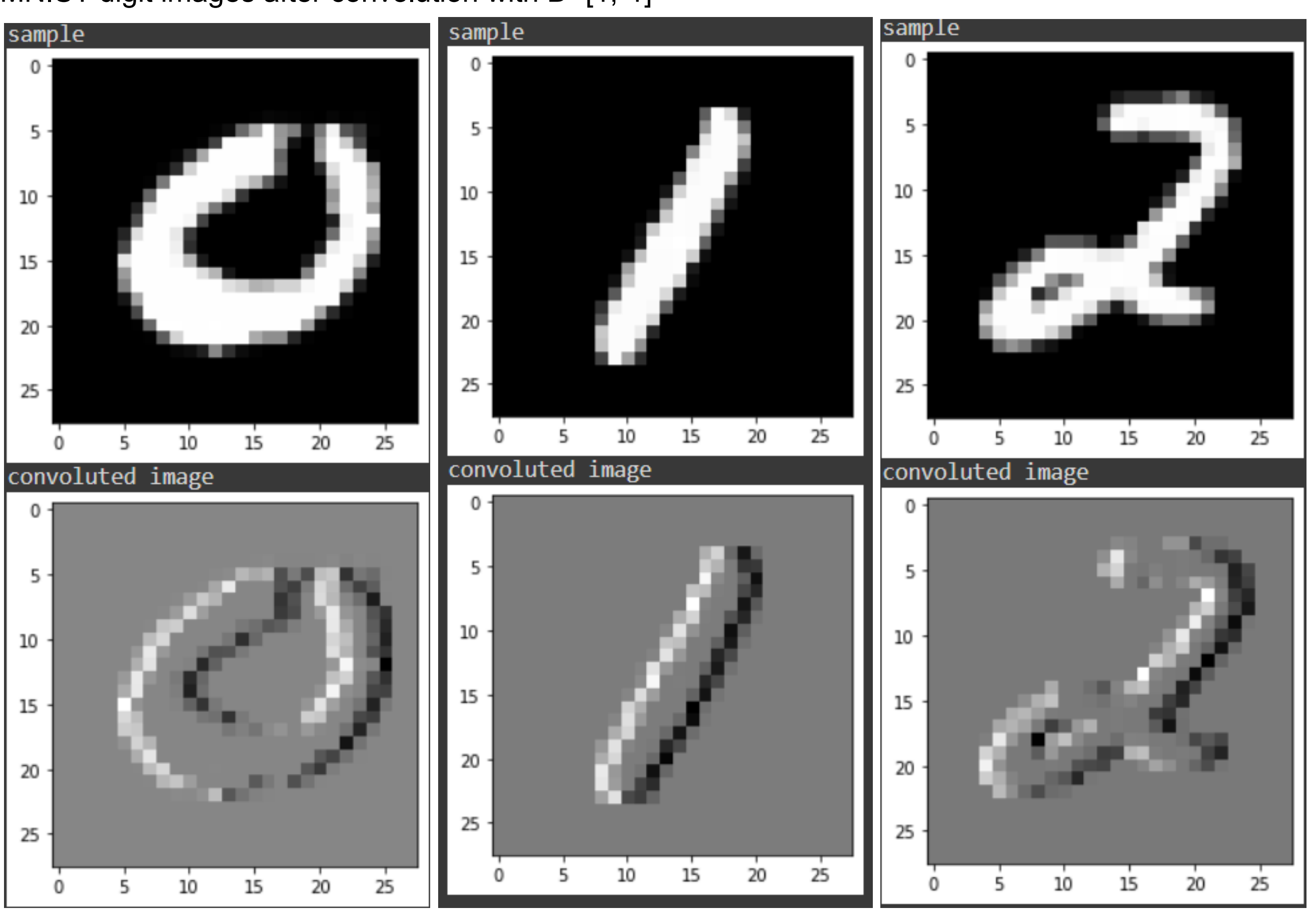
```
B

array([[-1,  1],
       [ 1, -1]])
```

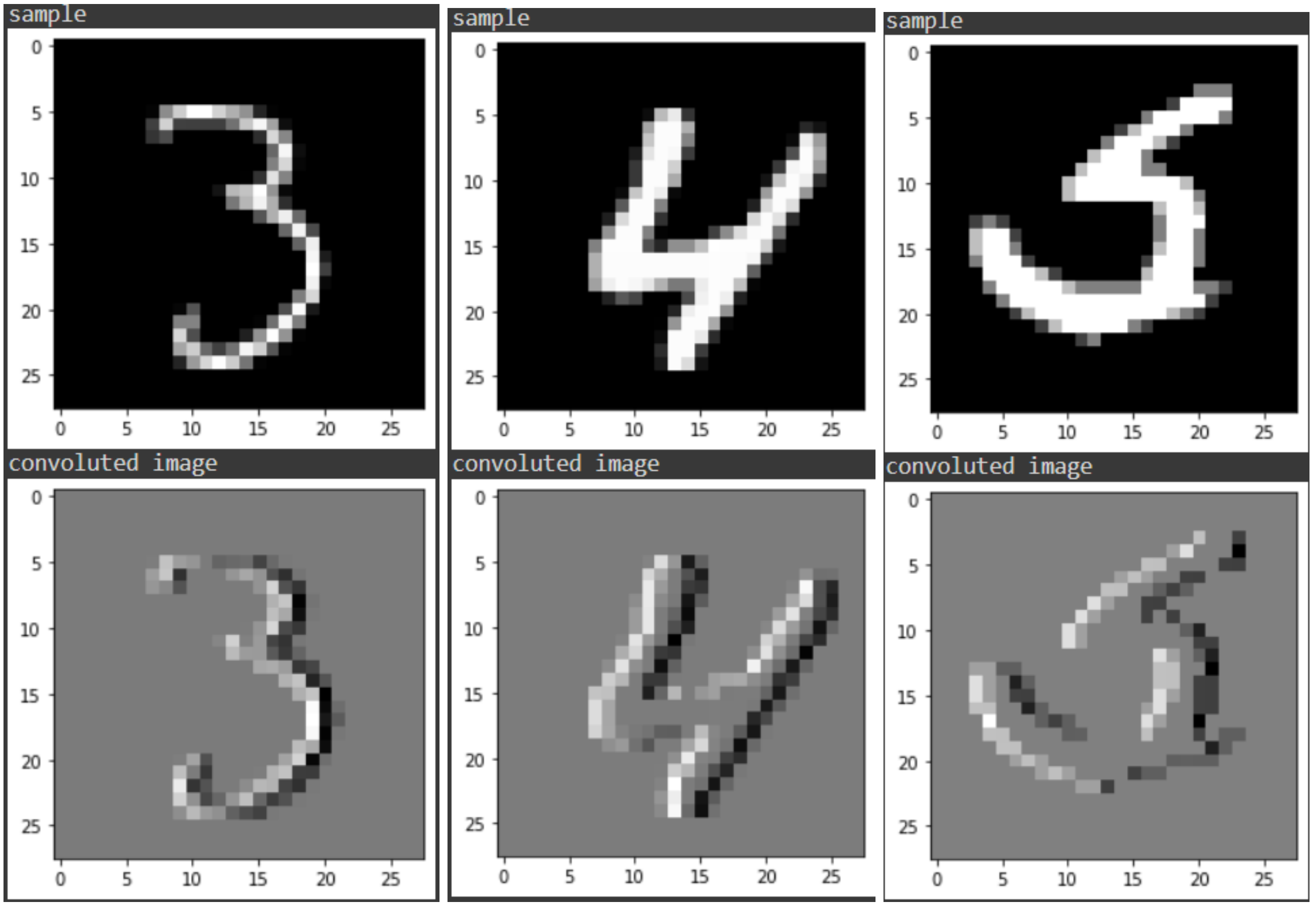
Array C (C=A*B)

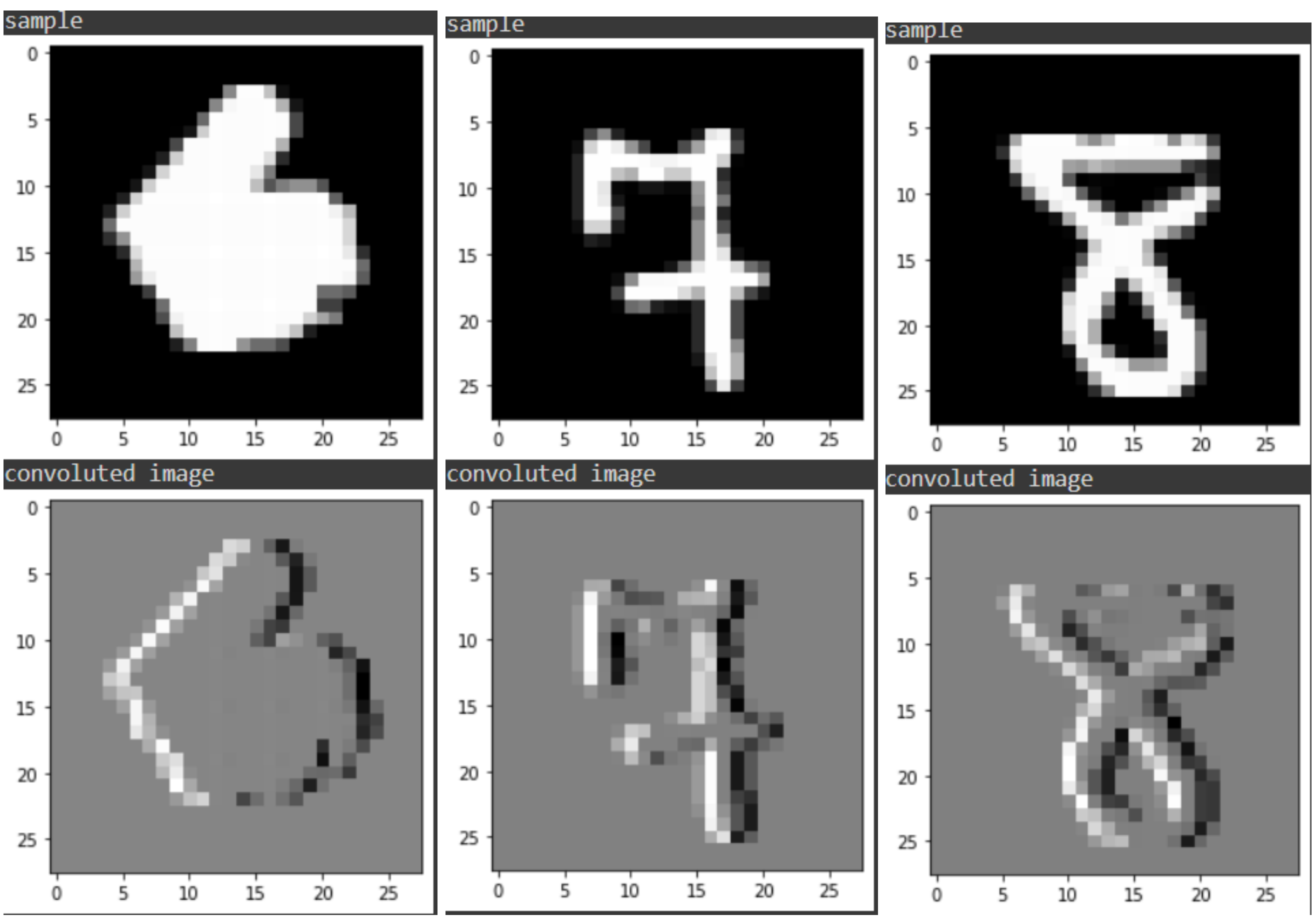
```
C

array([[-1.,  1.,  0.,  0., -1.,  0.,  1.],
       [ 1., -1., -1.,  0.,  1.,  1., -1.],
       [-1.,  0.,  2.,  0., -1.,  0.,  0.],
       [ 0.,  1., -2.,  0.,  1.,  0.,  0.],
       [ 1., -2.,  1.,  1.,  0., -1.,  0.],
       [ 0.,  1.,  0., -1., -1.,  1.,  0.],
       [-1.,  0.,  1., -1.,  2., -2.,  1.],
       [ 0.,  1., -1.,  0.,  0.,  0.,  0.],
       [ 1., -1.,  0.,  1., -1.,  1., -1.]])
```

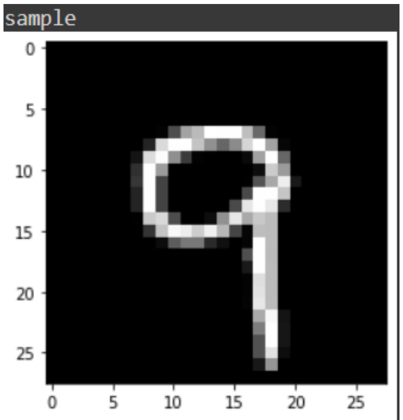
Part (b)

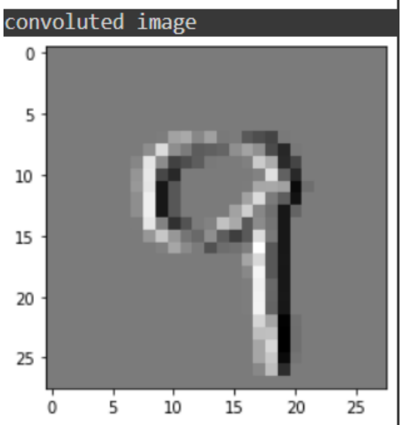
MNIST digit images after convolution with D=[1,-1]




sample
sample
sample
convoluted image
convoluted image
convoluted image


sample
sample
sample
convoluted image
convoluted image
convoluted image





Inference: The convolution **is detecting the vertical edges in the images.** Basically, if a pixel is surrounded on the left side by a value close to its own, then the convolution value of the resultant pixel goes close to zero. If the **pixel on the left has a pixel of value very far from its own then the convolution value is either very high positive or very low negative** depending on if the left side is black and the right is white or vice versa**.** This is how this convolution works

This happens as current_pixel=pixel_left*(1)+pixel_right*(-1) results in a very high or very low value when it hits a vertical edge, else its value remains close to zero. Thus it detects the vertical edges.

<u>**Question 11**</u>

Value of N: 100 images for each digit i.e 100*10 data-points
N=1000

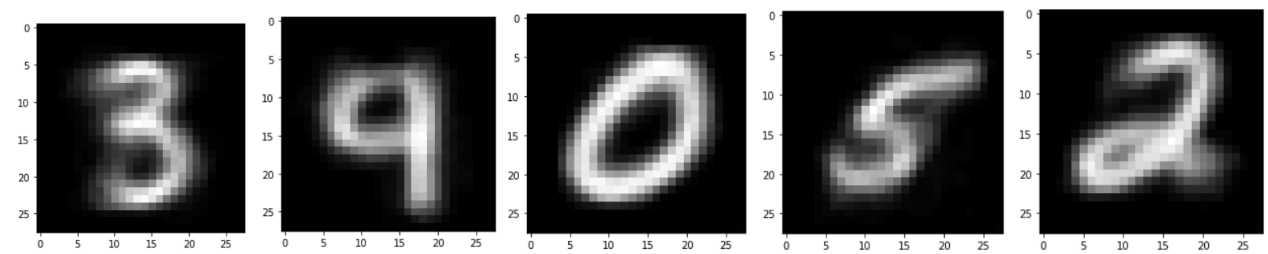
Value of n: the images are 28*28 pixels i.e 784 pixels
n=784

(i) Random initialisation of cluster representatives

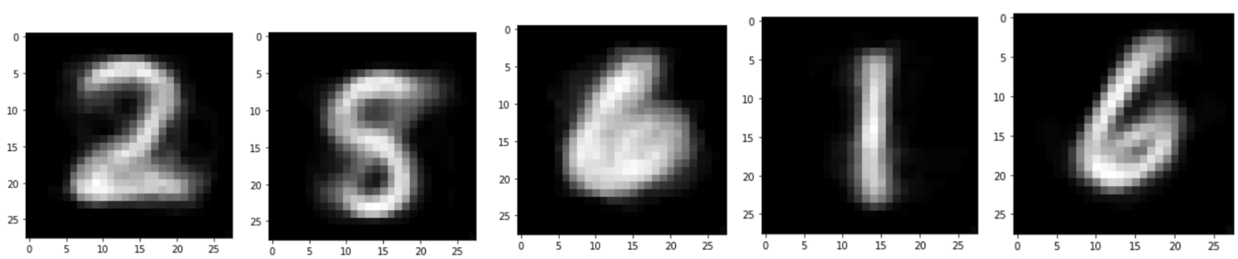
Part (a)

Number of iterations to convergence = **18**

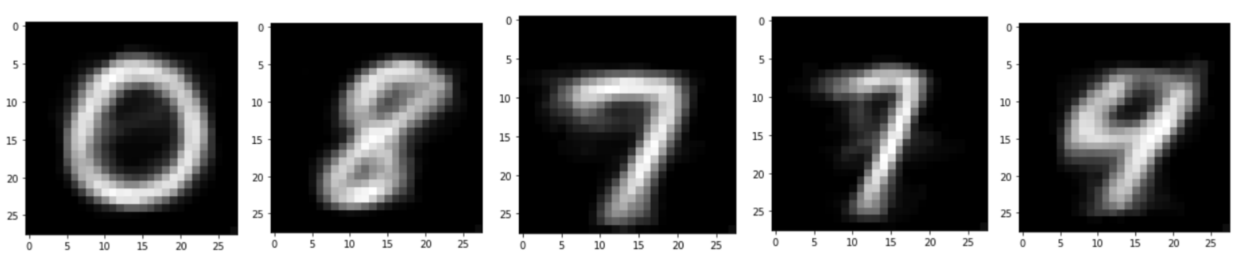
Cluster Representatives



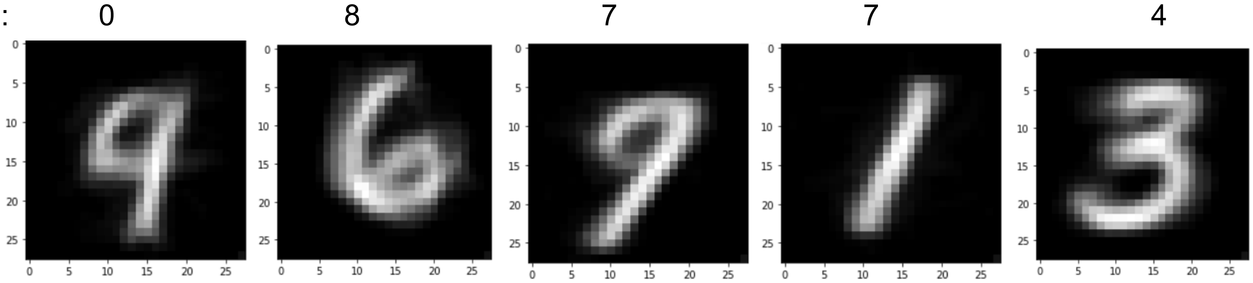
Inferred values: 3 9 0 5 2



Inferred values: 2 8 6 1 6



Inferred values: 0 8 7 7 4

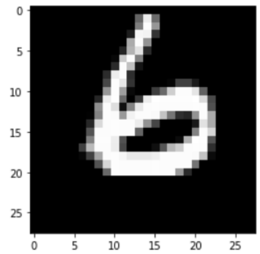


Inferred values: 4 6 9 1 3

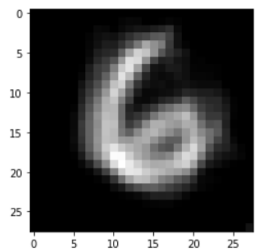
Part (b)

Testing Accuracy:
36 testing images out of 50 were assigned to the right representatives.
Hence the testing accuracy is 36/50=0.72

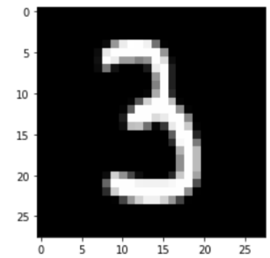
Sample assignments



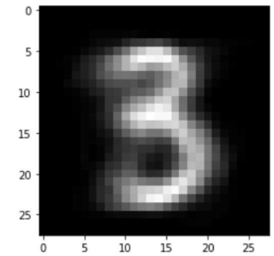
Testing Image



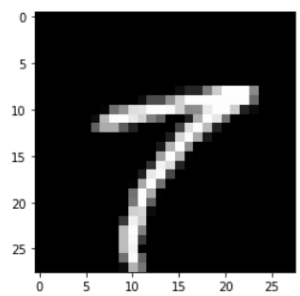
Cluster Representative



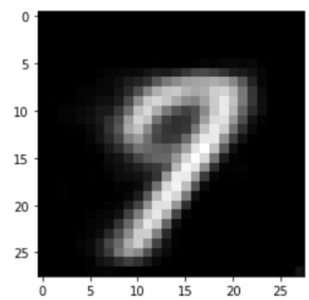
Testing Image



Cluster Representative



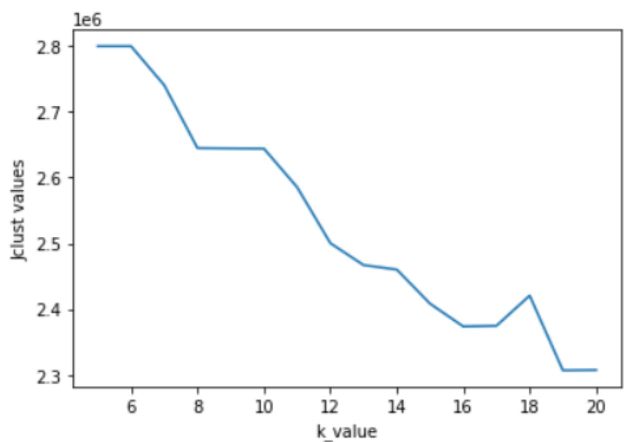
Testing Image



Cluster Representative

Part (c)

Jclust vs k-values plot for k=5 to k=20



Jclust values for k=5 to k=20

```
array([2.79914950e+06, 2.79914950e+06, 7.00000000e+00, 8.00000000e+00,
       9.00000000e+00, 2.79914950e+06, 2.58520197e+06, 2.50024676e+06,
       2.46701823e+06, 2.46021684e+06, 2.40850419e+06, 2.37392778e+06,
       2.37490806e+06, 2.39861555e+06, 2.30728904e+06, 2.30781640e+06])
```

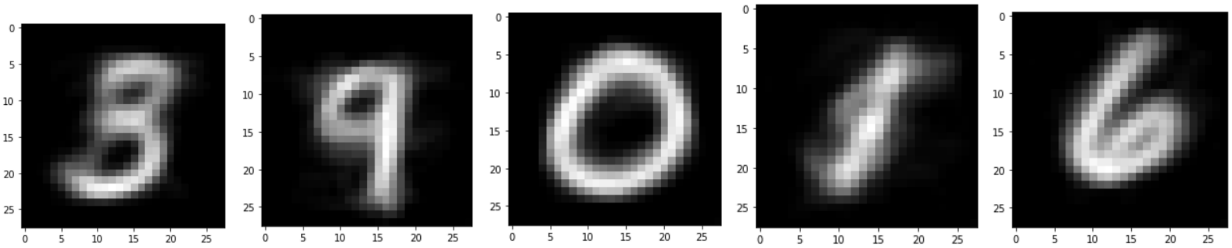
We see that the above graph elbows at k=16, hence we can infer that k is the right cluster size for this problem.

(i) Choose cluster representatives from the given data set

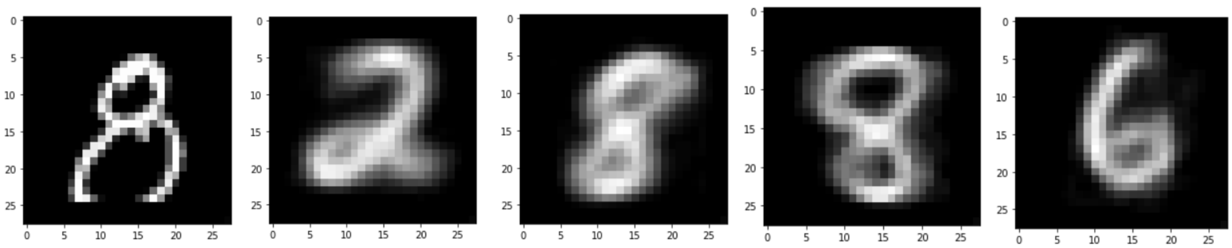
Part (a)

Number of iterations to convergence = **10**

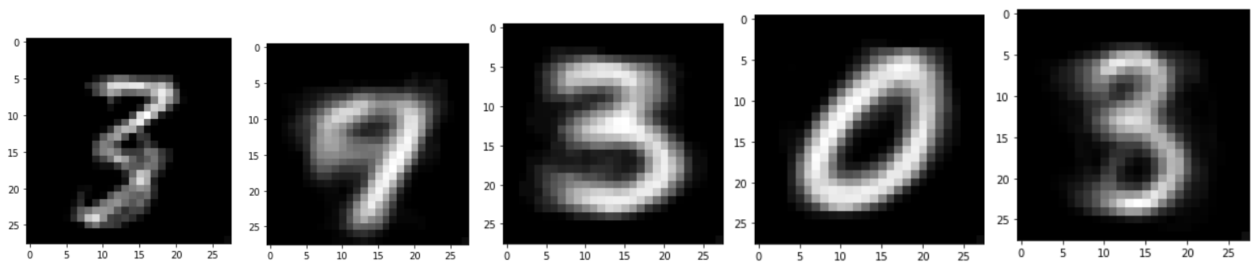
Cluster Representatives



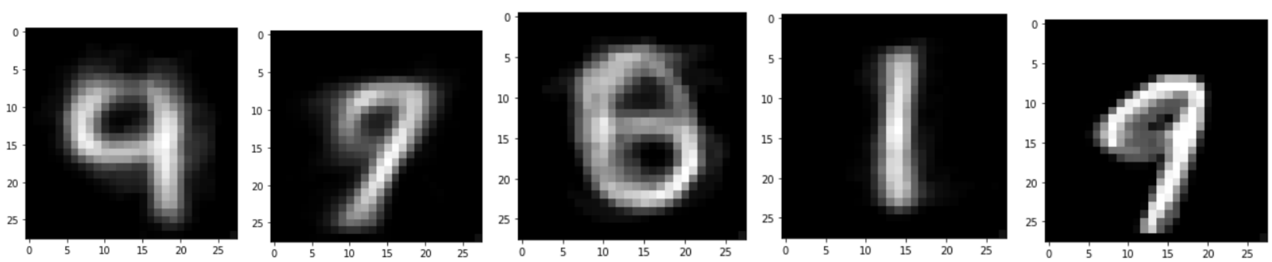
Inferred values: 3 9 0 5 6



Inferred values: 8 2 8 8 6



Inferred values: 3 9 3 0 3



Inferred values: 4 7 6 1 9

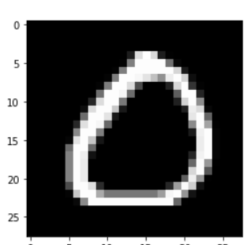
Part (b)

Testing Accuracy:

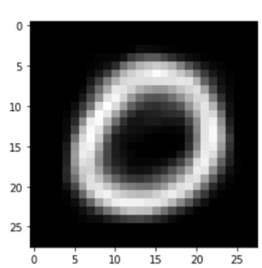
38 testing images out of 50 were assigned to the right representatives.

Hence the testing accuracy is 38/50=0.76

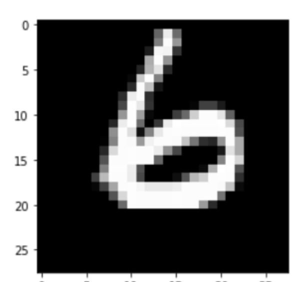
Sample assignments



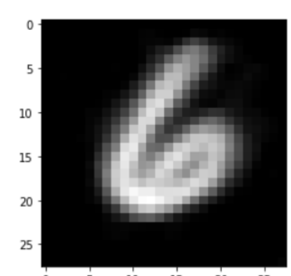
Testing Image



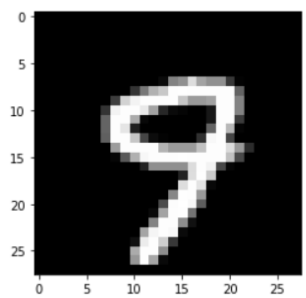
Cluster Representative



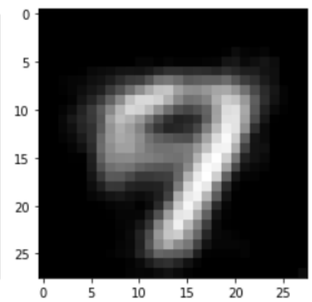
Testing Image



Cluster Representative



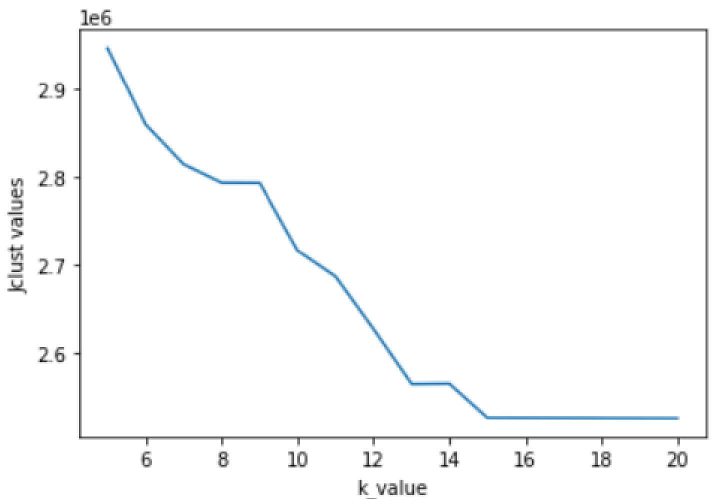
Testing
Image



Cluster
Representative

Part (c)

Jclust vs k-values plot for k=5 to k=20



Jclust values for k=5 to k=20

```
array([2946604.21794634, 2860014.40923455, 2814880.04271228,
       2793957.88250673, 2793772.66298047, 2717031.29912491,
       2687502.62951631, 2627797.2492993 , 2565751.73612171,
       2566166.93799964, 2527406.75803674, 2527261.32842551,
       2527110.70843512, 2526963.47843786, 2526817.95743129,
       2526675.38831129])
```

We see that the above graph elbows at k=15, hence we can infer that k is the right cluster size for this problem.

The initial choice of representative has surely affected the performance. We see that **choosing the initial representatives randomly takes more iterations to converge as compared to choosing representatives from the training set.**

We also see that **although the accuracy of both methods is similar, the second method is more reliable.** This can be inferred from the smoother Jclust vs k_value graph in the case of choosing representatives from the training set. Although the Jclust vs k_value graph from the method of choosing the representatives randomly has the same trend, it has random spikes and dips which shows that the accuracy values are close to the maximum but the maximum value is not always guaranteed as randomly generating representatives can result in some noise bad representatives.

Codes for Question 4 and 11 are present in the following link:
https://colab.research.google.com/drive/15WTrjqI32NaFickNV0ZCxPp8vATNzV92?usp=sharing