5.

84 values. CODE :- 25.py , 25.ipynb

a.

Finding the (84, 1) values i.e column vector of 84 values.

Method :

- First flatten the array.
- Find the covariance matrix of the array
- Find the eigen values using eig and take top 84 vector values from them.

Now repeat the same process for each digit.

Here we have selected the 84 vectors based on the eigenvalues of corresponding vectors. The eigenvectors with top 84 eigenvalues are taken.

b.

There are 2 processes that we need to do in the code.

gen(i, arr) → generates (84,1) data from (784,1) datatype.

retransform(i, arr) → generates back (28,28) image back from (84,1)

Retransformation:

⇒ Generation of (28,28) image back from (84,1)

To understand how retransformation is done

$(84,784)^T \cdot (784,1) \rightarrow (84,1)$ ⇒ $A = X - \mu$

$C^T A = V$

Now coming to retransformation :-

Our best guess is to transform it back to our homespace which is defined by the

84 vectors we have been using.

Best guess is to vector/data if we know the eigen-components and the PCA reduction at that vector.

Definition :-

$\langle AV \rangle = A^T V$ is like dot product of A, V column vectors.

Frobius Norm :

$\|A\|_2$ for one-column is same as frobius but

$$\|A\|_{Fro} = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} |a_{ij}|^2}$$

Let $A + n_1 V_1 + n_2 V_2 + \cdots + n_n V_n$ be the best approximation we can make by minimizing frobius norm where

$n_1, n_2, \ldots, n_n \subseteq \mathbb{R}$

$V_1, V_2, \ldots, V_n$ are eigen vectors / principal directions.

$u \rightarrow$ mean.

→ Let A be the original vector

$\|A - (u + n_1 V_1 + n_2 V_2 + \cdots + n_n V_n)\|_2$ should be minimum.

We see that,

$$\|A - (u + n_1 V_1 + n_2 V_2 + \cdots + n_n V_n)\|_{Foo} \cdot \|A - (u + n_1 V_1 + \cdots + n_n V_n)\|_{Foo}$$

$$= \text{trace}\left(\langle (A-u) - (n_1 V_1 + \cdots + n_n V_n), (A-u) - (n_1 V_1 + \cdots n_n V_n) \rangle \right)$$

Observe that $\langle V_i V_j \rangle = 0$ [symmetric matrix ⇒ eigen vectors are orthogonal]

$$= \text{trace}\left(\langle A-u \; A-u \rangle + \sum_{i=1}^{n} n_i^2 (\langle V_i V_i \rangle) - n_i \langle V_i A-u \rangle - n_i \langle A-u V_i \rangle \right)$$

$$= \text{trace}\left(\langle A-u\, A-u \rangle + \sum_{i=1}^{n} n_i^2 \langle v_i v_i \rangle - 2n_i \langle v_i\, A-u \rangle \right)$$

$$= \text{trace}\left(\langle A-u\, A-u \rangle + \sum_{i=1}^{n} n_i^2 - 2n_i \langle A-u\, v \rangle \right)$$

should be minimum.

We can clearly see that for (1-D) (1-column matrix) trace $(\langle AB \rangle) = \langle AB \rangle$ and for

$$\sum_{i=1}^{n} n_i^2 - 2n_i \langle v_i A-u \rangle \text{ to be minimum}$$

$$n_i = \langle v_i\, A-u \rangle = \langle A-u\, v_i \rangle$$

Hence the best approximation of A is

$$u_1 + \sum_{i=1}^{n} \langle A-u\, v_i \rangle v_i$$

where $\langle A-u\, v_i \rangle$ = ith value of PCA analysis of A

$$PCA(A) = (A-u)^T \cdot v = J$$

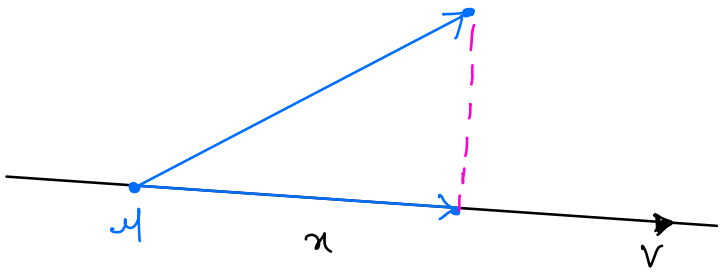
$\langle A - \bar{u} \, V_i \rangle = J[0, i] \Rightarrow$ 1 x n vectors

Best approximation gives PCA J is

$$\bar{u} + \sum_{i=1}^{n} J[0, i] \, V_i$$

Hence we can see that PCA component which are determent by $\langle A \, V_i \rangle$ are the best approximations to that coordinates

In <u>one-dimension</u> :



$\Rightarrow$ $u + x \cdot v$ is the best guess

$\Rightarrow$ where v is the vector

In the same way.

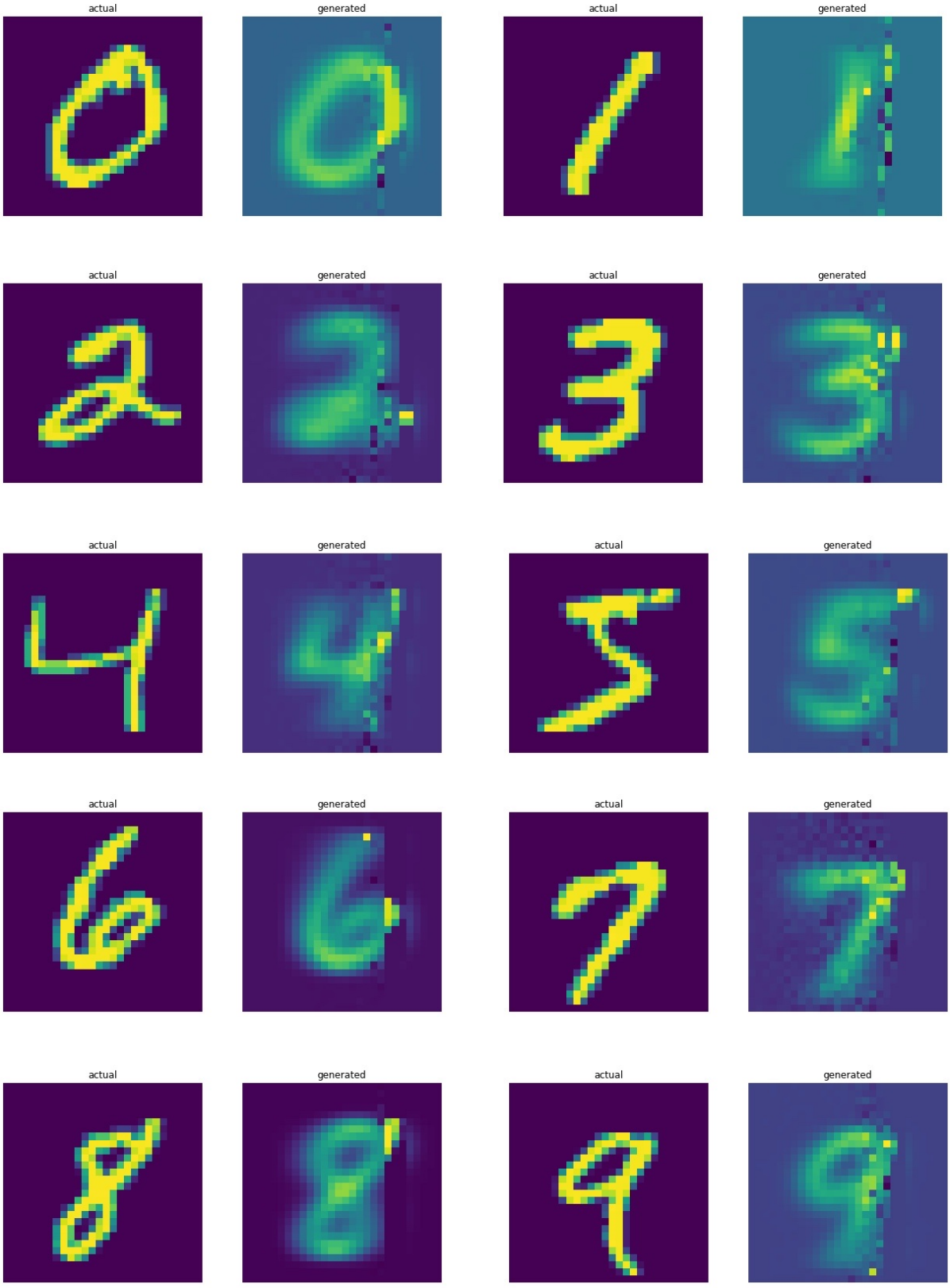
* New expected data is

$a + x_1 v_1 + x_2 v_2 + \cdots + x_n v_n$ where

dimensional space is of sizen.

$\Rightarrow$ This is what that has been implemented in the answer

Another way to look at it is first we project our data into the specific orthogonal vectors and the data in the remaining directions is nullified. While retransforming we trace back the values only along the direction in the specific orthonormal vectors since the data in the other direction is lost. But since the other directions have very less eigen values their components wont make much of difference. We can clearly see this from the following images.


actual
generated
actual
generated
actual
generated
actual
generated
actual
generated
actual
generated
actual
generated
actual
generated
actual
generated