

MAT160 - Final Project

Ahmed H. Mahmoud

June, 11th 2017

Problem No.1

Part A: Our code consists of first computing the left K singular vectors or rank K approximation $U_k^{(j)}$ (j represents the class) using the MATLAB function `svds` for each class $X^{(j)}$ using the training data set. For each test data point y_i , we compare the data point with all computed singular vectors (10 classes) and the inference value of the data point is the class of minimal error. The error of data point y_i w.r.t class j is computed as $E_j y_i = ||y_i - U_k^{(j)}(U_k^{(j)T} y_i)||_2$. Figure 1 shows the first five left singular vectors plotted as images for the first three classes (digits 0, 1 and 2). We notice that the first left singular vector (associated with largest singular value) closely resemble the corresponding digit followed by the second singular vector.

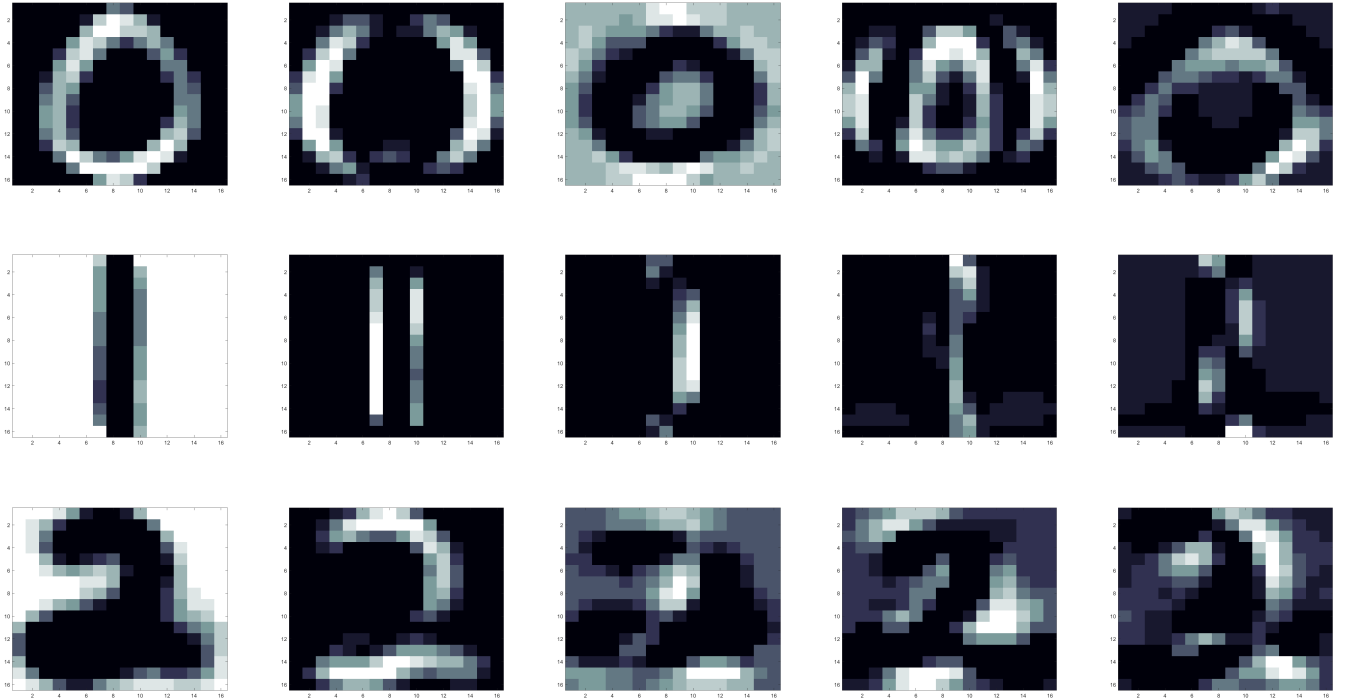
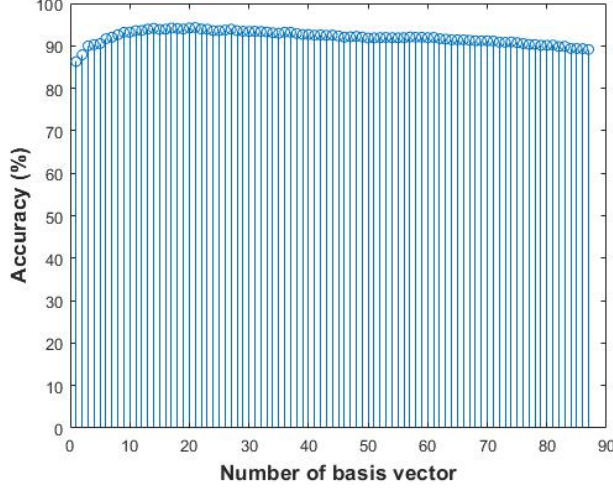


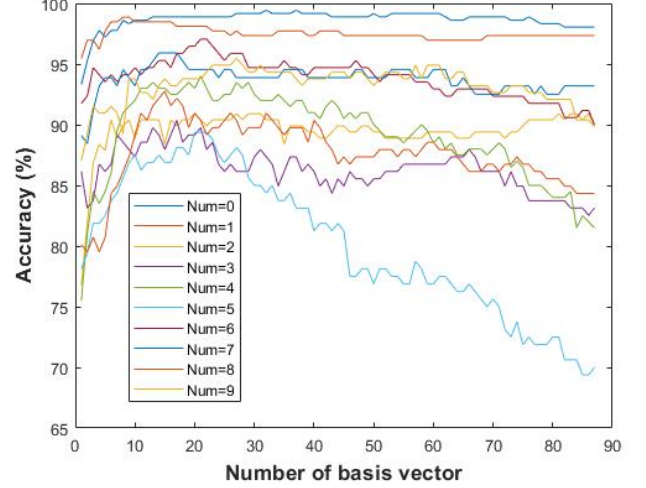
Figure 1: First left (associated with largest singular values) singular vectors for the first three classes.

Part B: We computed the accuracy as a function of the number of basis vector K . The accuracy is computed as the percentage of the number of images that has been classified correctly to all images that has been tested. Figure 2(a) shows the relation between accuracy and the number of basis vector. The maximum accuracy **94.32%** with **22** basis vector. The table to the right show the accuracy values for 5, 10, and 20 basis vector.

# Basis Vector	Accuracy
5	90.3%
10	93.2%
20	93.9 %



(a)



(b)

Figure 2: The accuracy of classified images as a function of the number of basis vectors K (a) for all images (b) per class/number.

Part C & D: In order to realize if it is beneficial to use different number of basis vector for different classes, we compute per-class accuracy for number of basis vector (from 2 up to 88). Figure 2(b) shows the per-class accuracy for different basis vector. We can see for class of *Number 5*, the accuracy decreases with increasing the number of basis vectors. For other classes like *Number 0* and *Number 1* the accuracy stays same (or marginally decreases) as the number of basis increases. For class of *Number 5*, it is better to use 21 vector basis for which the accuracy is 89.3%. Table 3 shows the maximum accuracy obtained for all class along with the number of basis vector associated with the maximum accuracy. We notice that class of *Number 1* only need 9 basis vector due to symmetry and simplicity of the shape of number one, while *Number 9* need up to 28 basis vector as it is not symmetric and the most complex.

In order to investigate the class of *Number 9* more thoroughly, we draw the few instances for which the number got mis-predicated using 28 basis vector as shown in Figure 4. Most of the mis-predicated images are indeed badly written and looks like other number like the first row in Figure 4 where the number looks more like *Number 4* more than *Number 9*.

Class	Max Accuracy (%)	# Basis Vector
<i>Number 0</i>	99.4429	33
<i>Number 1</i>	98.8636	9
<i>Number 2</i>	90.9091	21
<i>Number 3</i>	90.3614	18
<i>Number 4</i>	94.0000	22
<i>Number 5</i>	89.3750	21
<i>Number 6</i>	97.0588	22
<i>Number 7</i>	95.9184	15
<i>Number 8</i>	92.7711	16
<i>Number 9</i>	95.4802	28

Figure 3: The maximum accuracy obtained per class by varying the number of basis vectors. For each class, we report here the maximum accuracy along with the number of basis vector associated with it.

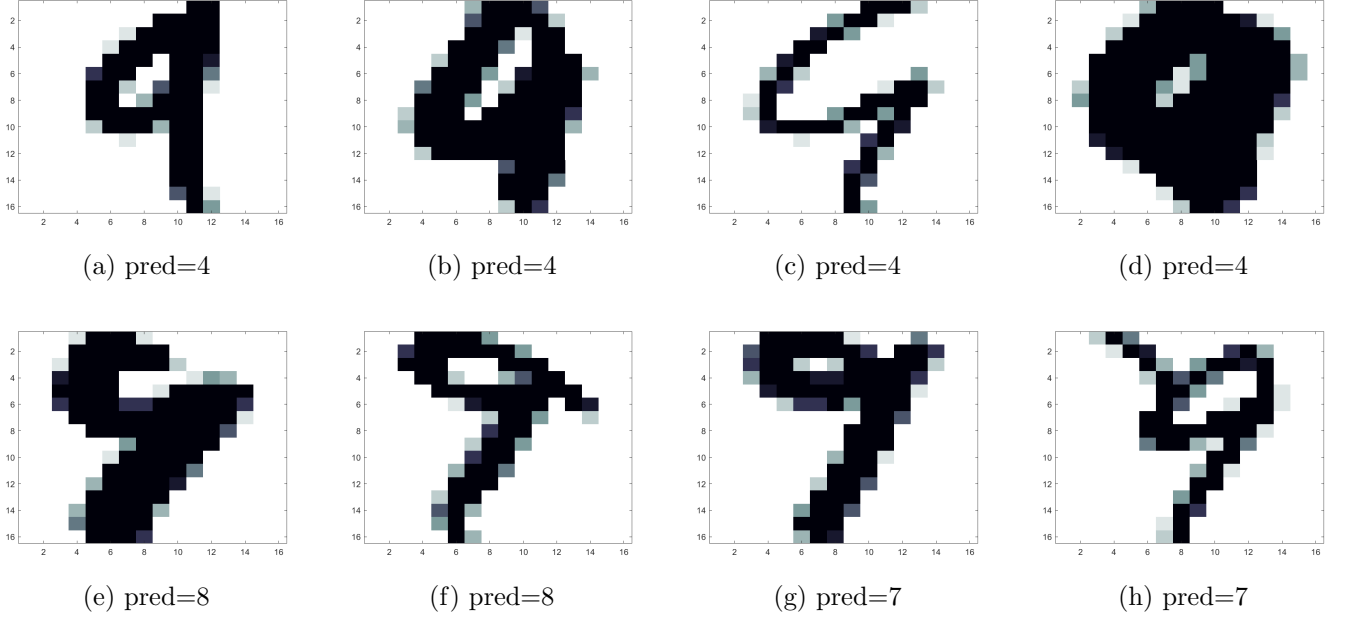


Figure 4: Using 28 basis vector where the maximum accuracy is obtained for class of *Number 9*, we draw the few instances where the images is classified incorrectly along with the wrong predication

Problem No.2

Part A: