Assignment#3

This assignment calls for using principal component analysis and histogram/Bayesian concepts for classifier design and evaluation. Use the Excel spreadsheet [Assignment\_3\_ Submission\_Template.xlsxPreview the document](https://classroom.ucsc-extension.edu/courses/3126/files/650137/download?verifier=0ViYznhrhnDMIophx13sSCUqMPN75vs14bODJenU&wrap=1) as a template to report your results. **Except for cutting and pasting your results, do not alter the spreadsheet in any other way.** This is the only recognized means of submitting this assignment. An extra sheet has been provided just in case you wish to make [optional] interesting observations.

Read through the instructions and examine the spreadsheet carefully and understand what must be computed or described. Enjoy the transition from "I have been exposed to machine learning" to "I am becoming quite good at machine learning!"

1. Import and uncompress handwritten digit data from yann.lecun.com/exdb/mnist (four files). If you use Python, refer to the Python code suggested in [Read\_MNIST-Improved.html](https://classroom.ucsc-extension.edu/courses/3126/files/608022/download?verifier=TsSUY38AuYQsIaeBatpD5A5nOckuIdYIw8w0j3ns&wrap=1) to read and parse the data.
2. Prepare the data you will use for this assignment by limiting the imported data to two specific digits. The exact two digits that you will use for the rest of the assignment is given in a table following these instructions. The positive and negative class labels have also been identified for you.
3. Prepare your X matrix and target vector T
4. Go through the XZCVP procedure to produce P which is the 2D approximation to X. Produce a 2D scatterplot showing the two classes. The mean vector (mu) and two eigenvectors (v1, v2) calculated in completing this step must also be entered into the Excel template attached.
5. Optional: Write a program, call a builtin function or use the “manual” procedure explained in class to convert any given 784 element feature vector in X to an image. No submission is required.
6. Optional: For a few vectors in X, visually verify that reconstructing the vector from a few principal components leaves the image looking more or less the same. That is, visually verify that reduction in dimension does not significantly compromise the identity of the samples. No submission is required.
7. Using two-dimensional approximations of the data, build a histogram classifier and a Bayesian classifier assuming that the feature vectors are distributed normally in the reduced dimensional space. Note: All the specifications needed to apply the classifiers must be entered into the Excel template attached. For the histogram classifier, the specifications include the ranges (min, max) covered by the two features, number of bins (fixed for you at 32x32, do not change this specification), and the histogram counts. For the Bayesian classifier, the specifications include the number of samples in the two classes, the mean vectors and covariance matrices. In order to use either classifier on a 784D query, the mean vectors and eigenvectors (Step 4 above) are needed.
8. **From the testing set**, exhibit any one representative positive feature vector xp and the corresponding mean-subtracted vector (zp), 2D representation (pp), reconstructed mean-subtracted vector (rp), and reconstructed mean-restored vector (xrecp). Do the same for any one representative negative feature vector xn from the testing set.
9. Run xp and xn through both the classifiers. Check and record the results (e.g. what is xp recognized as? With what probability? See the Excel sheet to understand how to report the results.)
10. Evaluate the testing accuracy of your two classifiers (2D histogram and Bayesian) and record the results.

Note: The helpful document [PCA\_Simple\_Example.html](https://classroom.ucsc-extension.edu/courses/3126/files/608021/download?verifier=5zyW3HfhH5i8cKN1Oc91cewGOZwmOvZByGlJe9s2&wrap=1) has been provided for guidance in the XZCVPR process. Also, note the helper file [Assignment\_3\_Input\_Output.html](https://classroom.ucsc-extension.edu/courses/3126/files/608023/download?verifier=1RJZgHycIXOUzlU0C6H0BakCSzAMelRujpdx0RUa&wrap=1)

|  |  |  |
| --- | --- | --- |
| **Student Name** | **Negative Class** | **Positive Class** |
| George Abraham | 6 | 9 |
| Azhar Abubacker | 4 | 7 |
| Anoop Antony | 4 | 8 |
| Subhasish Bandyopadhyay | 2 | 8 |
| Hyunjung Byun | 1 | 8 |
| Rupa Gangatirkar | 0 | 4 |
| Sree Iyer | 4 | 5 |
| Ganesh Jothikumar | 4 | 9 |
| Kavya Kambi Ravi | 0 | 5 |
| Nebu Koshy | 1 | 7 |
| Amit Kshirsagar | 3 | 4 |
| Sudhir Kulkarni | 3 | 8 |
| Wesley Larsen | 1 | 3 |
| Paul Vincent Legaspi | 2 | 7 |
| Tsui-Yan Leung | 1 | 5 |
| Jong Lin | 2 | 3 |
| Sean McMahon | 1 | 4 |
| Rekha Menon | 8 | 9 |
| Tripti Nashier | 5 | 9 |
| Omer Orhan | 2 | 4 |
| Chandramouly Rathinasababathy | 4 | 6 |
| Dhruv Sampat | 0 | 2 |
| Ajay Sapre | 0 | 3 |
| Uma Sarasamma | 2 | 9 |
| Uchit Shah | 6 | 7 |
| Himanshu Singh | 6 | 8 |
| Prateek Singh | 0 | 1 |
| Dennis Thomas | 3 | 7 |
| Senthilrajan Vasanthi Srinivass | 1 | 2 |
| Ravisankar Videla | 7 | 9 |
| Hsing-Yi Wang | 7 | 8 |
| Vincent Wang | 2 | 6 |
| Jimmy Wong | 0 | 8 |
| Yusuke Yakuwa | 5 | 6 |
| Carissa Yao | 5 | 7 |
| Sneha Ravikumar Yewankar | 2 | 5 |