ECE8560 Spring 2017
(Takehome #2)

Canvas submission only

Assigned 2/23/2017; Due 3/16/2017 11:59 PM

1 Overview

This is a comprehensive, four part take-home assignment which follows and builds upon Takehome #1 and uses the same dataset (H and S_T). **This is to be done individually**. I recommend you start as soon as possible. It is highly unlikely you can achieve a credible result by starting a few days before the due date.

The objectives are:

1. To assess your previously developed Bayesian, minimum error classifier S_T.
2. To determine a separating hyperplane between each pair of classes.
3. To consider k-NNR classification on S_T with k=1,3,5.
4. To employ PCA to reduce the number of features of the previous problem from the original 4 to the 2 'best' (decoupled and maximum variance).

Key remarks are:

1. Do all 4 parts of the assignment. Use this document as a checklist for your submission.
2. Open book and notes, but **no collaboration**. (This is an individual effort.)

3. **Submit to Canvas by the deadline.**

4. **Clarity of the presentation (in addition to technical correctness) counts significantly**.

2 Data

This is a $c = 3$ class problem. You were given the $d = 4$ training and test data (H and S_T). Refer to Takehome #1 for more details.

3 Part 1: Assessing Classification Results from Takehome #1

3.1 Previous Effort

In takehome #1, you classified all 15000 test vectors by processing `test_sp2017_v19` sequentially. Since `test_sp2017_v19` contains one (transposed) feature vector per line, this was done by writing the class (from your Bayesian classifier) to a corresponding file named

`<yourname>-classified-takehome1.txt`

with one integer per corresponding line. Your answer was therefore an ASCII (text) file with one ASCII integer entry, a_i, per line, where

$$a_i = \begin{cases} 1 & \text{if sample i is classified as class } w_1 \\ 2 & \text{if sample i is classified as class } w_2 \\ 3 & \text{if sample i is classified as class } w_3 \end{cases}$$

Use your file from Takehome #1 for the following part. Recall I have a copy.

3.2 Assessment Knowing True Class

In S_T, the feature vectors are arranged by class in the following sequence: Starting from the first sample, the sequence for the true class is 3-1-2-3-2-1 and this sequence repeats throughout S_T.

Using the text file `<yourname>-classified-takehome1.txt` in the zipped archive you submitted, show your classifier performance on S_T **using a confusion matrix**. Also, estimate $P(error)$.

Confusion Matrix. A 'confusion matrix', also referred to as a contingency table or an error matrix, is a tabular representation that allows visualization of the performance of a classification algorithm. Rows correspond to ground truth (true class); columns indicate accumulated resulting classifier decisions. A diagonal confusion matrix indicates perfect classification, i.e., estimated P(Error)=0.

4 Part 2: Separating Hyperplanes

In this part, you are to implement a (simple) Ho-Kayshap hyperplanar classifier *for the given training set*. **Consider the classes pairwise**. You will thus generate the 3 hyperplanes from the respective subsets of H. In each case, provide the exact parameters of the respective hyperplane(s) and their classification performance on the pair of classes used in training.

Following this, classify S_T using the above hyperplanes and show your classifier performance on S_T using a confusion matrix. Also, estimate $P(error)$.

5 Part 3: $k - NNR$ Strategies

In this approach, develop, apply and assess a k-NNR classifier for S_T using H. Generate results (classification error and confusion matrix) as a function of k. For this problem, use $k = 1$, 3, and 5. You may use a brute-force matching strategy, but I'll be more impressed if you implement something more computationally efficient.

6 Part 4: PCA

Repeat the effort of takehome #1, but only using 2 'best' features (which are linear combinations of the original features). Use PCA on H to determine projected features which maximize variance. Show, in detail, how these new features are derived from the given 4-D $\underline{x}$. Apply this feature reduction result to S_T, using the 2 new derived features. Show your classifier performance

using a confusion matrix, estimate $P(error)$, and compare this with your results from Takehome #1 (assessed in Part 1, above).

7 Format of the Report Results

This aspect is critical. If you make it difficult for me to asses your effort, I won't. The final report must be in your solution archive in a PDF-format file named `results_takehome2.pdf`. The report results **must be in the following order**:

1. p. 0: Title page (<name>, <CU username>, ECE 8560, Takehome #2)
2. p. 1: Confusion matrix and P(error) estimate from Part 1.
3. p. 2: The 3 hyperplanes from the respective subsets of H in Part 2.
4. p. 3: Hyperplane classifier performance on S_T, as a confusion matrix and estimate of $P(error)$ from Part 2.
5. p. 4: k-NNR classifier results on S_T using H. Classification error and confusion matrix as a function of k from Part 3.
6. p. 5: PCA results (feature derivation, use and assessment on S_T) from Part 4.
7. p. 6: Comparison and analysis of Parts 1-4.
8. Pages beyond 6: Anything else you feel is relevant.

8 Additional Notes and Constraints

8.1 Format of the Electronic Submission

The final **zipped archive** is to be named **<yourname>-takehome2.zip**, **where <yourname> is your (CU) assigned user name**. You must upload this to the ECE8560 Canvas page prior to the deadline for your solution to be considered.

The minimal contents of this archive are described below.

1. Include a `readme.txt` file listing the contents of the archive and a brief description of each file. Include 'the pledge' here. Here's the pledge:

   Pledge:
   On my honor I have neither given nor received aid on this exam.

2. Put the results for Parts 1-4 each in a single directory.

3. Include **all** (your) source code used in your simulations.

4. Indicate your engineering decisions and solution derivations for each part.

5. Estimate your classifier performance for each part by showing the confusion matrix and P(Error), using the test set data (now with known class).

6. Compare the results of Parts 1 thru 4.

7. All documentation should be in pdf in a file named `<yourname>-takehome2.pdf` **No MS Word (doc) files.** The quality and structure of the documentation is very important.