ECE8560 Spring 2017
Takehome #1

(Canvas submission only)
Assigned 1/26/2017
Due 2/16/2017 11:59 PM.
Late policy:
Solution is graded if received by the deadline.

Contents

1 Overview

This is a comprehensive take-home quiz which is to be done **individually**. The objective is to design and test a Bayesian, minimum error classifier using a given H and S_T. We may re-use this data for subsequent take-home quizzes.

Key remarks are:

1. Do all parts of the quiz. 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. The $d = 4$ training and test data is available on the course Canvas page in files `train_sp2017_v19` and `test_sp2017_v19`. Each file consists of 15,000 $d = 4$ dimensional vectors in physical coordinates, **with one feature vector (transposed) per line**. The first 5000 samples in `train_sp2017_v19` correspond to class w_1 (H_1), the second 5000 samples correspond to class w_2 (H_2) and the third 5000 samples correspond to class w_3 (H_3). Using `test_sp2017_v19`, the objective is to determine the class for each sample using your resulting Bayesian classifier design (based upon `train_sp2017_v19.`) Note that, at this time, you do not know the correct class for each vector in `test_sp2017_v19`. **Note: do not do any training with the test data.**

3 Bayesian Classifier Design and Implementation

The objective is to develop a Bayesian classifier for this data which minimizes classification error. I am interested in both your engineering

judgment as well as the appropriateness and performance of your resulting classifier. **You will need to determine and justify the appropriate statistical model, including apriori probabilities and pdfs for each class**.

3.1 Models and Parameters

Figures 3 and 4 of Chapter 2 indicate the Bayesian model and classifier structure.

3.2 Reality Check

Training data may be used to check the reasonableness of your classifier. When you are done designing the classifier, classify (test) each sample in `train_sp2017_v19`, and estimate P(Error) using this *training data*. This would also be a good reality check on the quality of your classifier (or the difficulty of the problem, or both).

4 Testing and Reporting Classification Results

In order for me to determine the classification error in your work, you need to classify each feature vector in `test_sp2017_v19`, and communicate the result to me in machine manipulable form. To check the quality of your classifier, you will put your classification result for each test set vector in a file as detailed below.

The simplest way to state this is that is you will classify all 15000 test vectors by processing `test_sp2017_v19` sequentially. Since `test_sp2017_v19` contains one (transposed) feature vector per line, this is done by writing the class (from your Bayesian classifier) to a corresponding file named

`<yourname>-classified-takehome1.txt`

with one integer per corresponding line. The desired format of this part of your answer is therefore an ASCII (text) file[1] with one ASCII integer entry, a_i, per line, where

[1]The reason for this format is to enable me to judge your classifier.

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

Please name this text file as indicated and include it in the zipped archive you submit.

5 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_takehome1.pdf`. The report results **must be in the following order**:

1. p. 0: Title page (<name>, <CU username>, ECE 8560, Takehome #1)
2. p.1-2: Indicate your **engineering decisions and associated rationale**, e.g., density function form, parameter estimation, $P(w_i)$, design of the classifier, etc.
3. p. 3: Show the exact form of the discriminant function used for each class.
4. p. 4: Estimate your P(Error), using the **training data with known class**.
5. Pages beyond 4: Anything else you feel is relevant.

6 Additional Notes and Constraints

6.1 Format of the Electronic Submission

The final **zipped archive** is to be named **<yourname>-ece8560-takehome1.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 all results in the single, top-level directory.

3. All documentation should be in a single pdf file named `<yourname>-takehome1.pdf with the structure indicated in Section 5.` **No MS Word (doc) files.**

4. Include all (your) source code used in your simulations.

5. Include the classifier result text file `<yourname>-classified-takehome1.txt`