Pattern Recognition
Exercise set 3

1. **[6 points] (Bayes minimum error rate classifier)**

   (a) Implement a training function for Bayes classifier assuming that class-specific density functions are multivariate Gaussians. Inputs for the training function are all the training feature vectors and the corresponding class labels. Outputs: mean vectors, covariance matrices, and prior probabilities for each class.

   (b) Implement a minimum error rate classifier function. Inputs are the outputs of the training function and the test sample(s) to be classified. Output: predicted class label(s).

   (c) Test your classifier using the i-vector data. Report the speaker identification accuracy.

2. **[4 points] (Bayes minimum risk classifier)** Suppose that speaker #7 is known to be an unreliable person. In order to prevent this person from fooling the speaker identification system, implement and test minimum risk classifier using decision costs c_{ij} $(i, j = 1, \ldots, 20)$, where

$$c_{ij} = \begin{cases} 0, & \text{when } i = j, \\ 6, & \text{when } i \neq 7 \text{ and } j = 7, \\ 1, & \text{otherwise.} \end{cases}$$

   The training function is the same as in the minimum error rate classifier. The classifier function needs to be partially updated to include the decision cost matrix (formed from values c_{ij}) as an additional input parameter and to perform the classification based on conditional risks. Test sample should be classified to the class that has the lowest conditional risk.

   By classifying the test set samples, determine how many times speaker #7 gets misclassified as some other person when using minimum error rate and minimum risk classifiers (do we obtain the desired result?). How does the overall speaker identification accuracy get affected by using minimum risk classifier with the given detection costs as opposed to the minimum error rate classifier?

3. Given a feature vector $\boldsymbol{x} \in \mathbb{R}^d$ from one of the possible classes $\{\omega_1, \ldots, \omega_M\}$, we recall that Bayes minimum error rate classifier finds the class that maximizes the posterior probability:

$$\omega^* = \underset{j=1,\ldots,M}{\operatorname{argmax}} \; P(\omega_j|\boldsymbol{x}).$$

   (a) **[2 points]** Assuming multivariate Gaussians for the class-conditional densities, show that the above classification rule is equivalent to

$$\omega^* = \underset{j=1,\ldots,M}{\operatorname{argmax}} \; g_j(\boldsymbol{x})$$

   where

$$g_j(\boldsymbol{x}) = -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_j)^\intercal \boldsymbol{\Sigma}_j^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_j) + \log P(\omega_j) - \frac{1}{2}\log|\boldsymbol{\Sigma}_j|.$$

   (b) **[Bonus (2 points)]** Show that if we further assume $\boldsymbol{\Sigma}_j = \sigma^2 \boldsymbol{I}$, we can express the classication rule as

$$\omega^* = \underset{j=1,\ldots,M}{\operatorname{argmax}} \; h_j(\boldsymbol{x})$$

   where

$$h_j(\boldsymbol{x}) = \boldsymbol{x}^\intercal \boldsymbol{w}_j + w_{j0}.$$

   What are $\boldsymbol{w}_j$ and w_{j0}?

Bonus exercises can give you extra points, which makes it possible to get more than one-third of the course points from the exercises.

Submit your answers to Moodle by November 30, Thu, 23:55.
Late submissions:
Before December 1, Fri, 7:00: -10% of points
Before December 2, Sat, 12:00: -30% of points
After December 2, Sat, 12:00: Not accepted without a good reason that has to be given well before the deadline.

The submission should be an archive (zip, tar, etc.) that contains following files:

- `answers.pdf`: Contains answers to questions in pdf-format. If you wish, you may include scanned (readable) handwritten answers in your answer file (to avoid math typesetting). In case you give some answers in program code comments, mention it in the `answers.pdf`. Include your full name.
- `main.m` / `main.py`: A script that outputs the answers for all programming tasks. Write your full name in the first line as a comment. If using Python, please use Python3 + NumPy + matplotlib.
- Possibly other code files that your main script calls.
- Include the required data set files (ivectors.txt) in the archive so that the main script is runnable right after unpacking the archive.

Use the following naming convention for the archive file:
`<first name>_<last name>_ex<exercise set number>`
For example:
`ville_vestman_ex3.zip`