Assignment 4

Bilal Baig (215734320)

`bilalb@yorku.ca`

April 6, 2021

Note: This assignment is mainly for you to review several basic generative models. You have to work individually. **You must use the same mathematical notations in textbook or lecture slides to answer these questions.** You must use this latex template to write up your solutions. Remember to fill in your information (name, student number, email) at above. No handwriting is accepted. Direct your queries to Hui Jiang (`hj@eecs.yorku.ca`)

Exercise 1

Bayesian Decision Theory (20 marks)

(a)   Assume that we are allowed to reject an input as unrecognizable in a pattern-classification task. For an input $\mathbf{x}$ belonging to class ω, we can define a new loss function for any decision rule $g(\mathbf{x})$ as follows:

$$l\big(\omega, g(\mathbf{x})\big) = \begin{cases} 0 & : & g(\mathbf{x}) = \omega \\ 1 & : & g(\mathbf{x}) \neq \omega \\ \lambda_r & : & \text{rejection,} \end{cases}$$

where $\lambda_r \in (0,1)$ is the loss incurred for choosing a rejection action. Derive the optimal decision rule for this three-way loss function.

(b)   What would happen if we set $\lambda_r > 1$?

Your answers:

1. $x \longrightarrow g(x)\epsilon\{0, 1, \lambda_r\}$
    $g^*(x) = \arg\max_k \Pr(\{0, 1, \lambda_r\}_k) \cdot p(x|\{0, 1, \lambda_r\}_k)$

2. If γ_r was set to greater than 1, then the risk function would be negative and would result in unsatisfactory reuslts.

Exercise 2

Gaussian Models (20 marks)
Derive the maximum likelihood estimation (MLE) for multivariate Gaussian models with a diagonal covariance matrix, i.e. $\mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ with $\mathbf{x}, \boldsymbol{\mu} \in \mathbb{R}^d$ and

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1 & & \\ & \ddots & \\ & & \sigma_d \end{bmatrix}$$

Show that the MLE of $\boldsymbol{\mu}$ is the same as Eq.(11.3) on page 238 and that of $\{\sigma_1, \cdots, \sigma_d\}$ equals to the diagonal elements in Eq.(11.4) on page 239.

Your answers:

1. $|\Sigma| = (\sigma_1 \cdot \sigma_2 \ldots \sigma_{d-1} \cdot \sigma_d)$

$$(x - \mu)^T \Sigma^{-1} (x - \mu) = \begin{bmatrix} x_1 - \mu_1 \ldots x_d - \mu_d \end{bmatrix} \begin{bmatrix} \frac{1}{\sigma_1}(x_1 - \mu_1) \\ \ldots \\ \frac{1}{\sigma_d}(x_d - \mu_d) \end{bmatrix}$$

$$= (\frac{1}{\sigma_1}(x_1 - \mu_1)^2 + \ldots + \frac{1}{\sigma_d}(x_d - \mu_d)^2)$$

$$p_{\mu,\Sigma}(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} e^{\frac{-1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

$$= \frac{1}{(2\pi)^{d/2}(\sigma_1 \cdot \sigma_2 \ldots \sigma_{d-1} \cdot \sigma_d)^{1/2}} e^{\left(-\frac{1}{2\sigma_1}(x_1-\mu_1)^2 - \ldots - \frac{1}{2\sigma_d}(x_d-\mu_d)^2\right)}$$

At this point we can see that this is just the product of d univariate gaussian models and so we know that the μ and σ would have to be the same

Exercise 3

Gaussian Mixture Models (40 marks)
You will solve a simple binary classification problem (class A vs. class B) using simple multivariate Gaussian models as well as Gaussian mixture models. Assume two classes have equal prior probabilities. Each observation feature is a three-dimensional (3D) vector. You can download the data set from: `http://www.eecs.yorku.ca/~hj/MLF-gaussian-dataset.zip`.
You will use several different methods to build such a classifier based on the provided training set, and then the estimated models will be evaluated on the provided test set. You will have to implement all training and test methods from scratch.

1. (10 marks) Build a simple classifier using multivariate Gaussian models. Each class is modeled by a single 3D Gaussian distribution. You should consider the following structures for the covariance matrices:

   - Each Gaussian uses a diagonal covariance matrix.
   - Each Gaussian uses a full covariance matrix.

   Use the provided training data to estimate the Gaussian mean vector and covariance matrix for each class based on MLE. Report the classification accuracy of the MLE-trained models as measured by the test set for each choice of the covariance matrix.

2. (30 marks) Improve the Gaussian classifier from the previous step by using a GMM to model each class. You need to use the k-means clustering method to initialize all parameters in the GMMs, and then improve the GMMs based on the EM algorithm. Investigate GMMs that have 2, 4, 8, or 16 Gaussian components, respectively. Determine the best model configuration in terms of the number of Gaussian components and the covariance matrix structure (diagonal vs. full) for this data set.

The csv data format: All training samples are given in the file *train-gaussian.csv*, and all test samples are given in the file *test-gaussian.csv*. Each line represents a feature vector in the format as follows:

$$y, \ x1, \ x2, \ x3,$$

where $y \in \{A, B\}$ is class label, and $[x1 \ \ x2 \ \ x3]$ is a 3D feature vector.
   You can use the method (item 4) in

```
https://colab.research.google.com/drive/1FyahMGAE22716sUCrNXPvTrKTwS615Hd#
                        scrollTo=onKkhFTQ1aAJ
```

to load this data set in Python.

What to submit?

You must submit:

1. one PDF document (using this latex template) for your solutions to all written questions and all results and discussions for your programming assignments.

2. one zip file that includes all of your Python codes (e.g., *.ipynb if you use Jupyter notebooks) and a readme file for TA to run your codes.

from eClass before the deadline. No late submission will be accepted.