

• **Problem 4 (12 pts)**

For this question you will build classifiers to label images of handwritten digits. Each image is 8 by 8 pixels and is represented as a vector of dimension 64 by listing all the pixel values in raster scan order. The images are grayscale and the pixel values are between 0 and 1. The labels y are $\{0, 1, 2, \dots, 9\}$ corresponding to which character was written in the image. There are 700 training points and 400 test points for each digit; they can be found in `digits_train.txt` and `digits_test.txt`. These data sets can be loaded by using the helper function in `utils.R`.

You will implement both linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA) to classify these images. Recall that conditioning on each class $k \in \{0, 1, \dots, 9\}$, the feature $X | Y = k$ follows a multivariate Gaussian distribution, that is,

$$\mathbb{P}(X = \mathbf{x} | Y = k) = (2\pi)^{-p/2} |\Sigma_k|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_k)^\top \Sigma_k^{-1}(\mathbf{x} - \boldsymbol{\mu}_k)\right\} \quad (0.2)$$

where $\boldsymbol{\mu}_k \in \mathbb{R}^p$ is the conditional mean and $\Sigma_k \in \mathbb{R}^{p \times p}$ is the conditional covariance matrix. For LDA, Σ_k is assumed to be the same across classes. The priors are

$$\pi_k = \mathbb{P}(Y = k), \quad \text{for all } k \in \{0, 1, \dots, 9\}.$$

You will compute the maximum likelihood estimators of the priors π_k , the conditional means $\boldsymbol{\mu}_k$ and the conditional covariance matrices Σ_k for $k \in \{0, 1, \dots, 9\}$, and use the estimators to construct classifiers.

Read carefully the structure of `discriminant_analysis.R`. Include your code for all sub-questions.

1. **(4 pts)** Complete the functions `Comp_priors`, `Comp_cond_means` and `Comp_cond_covs` in the file `discriminant_analysis.R`.
2. **(2 pts)** Complete the functions `Predict_posterior` and `Predict_labels` in the file `discriminant_analysis.R`.
3. **(2 pts)** Use LDA to classify the test data by completing part a in `hw4_starter.R`. Report the misclassification error of LDA.
4. **(2 pts)** Use QDA to classify the test data by completing part b in `hw4_starter.R`. Report the misclassification error of QDA.
5. **(2 pts)** Complete part c in `hw4_starter.R`, i.e. perform LDA and QDA by using the built-in `lda` and `qda` functions and compare with your implementation in terms of both misclassification rates and computational speed.