Assignment 3

This assignment is due **Monday November 2** at **11:59PM**.
Solutions should be turned in through the assignment FTP site in PDF form. The name of the PDF file should be ps3_YourStudentID, e.g., "ps3_9762578.pdf". Your MATLAB code (with comments) should be included in the PDF file too. The FTP site is 140.114.71.2 port 1235. You can use the same ID and password of the course website to login.

---

1. **(30 points)**

   Consider a linear classification model in terms of dimensionality reduction given by $\mathbf{y} = \mathbf{W}^{\mathrm{T}}\mathbf{x}$, where the matrix $\mathbf{W}$ consists of $\{\mathbf{w}_j \,|\, j = 1, \ldots, D'\}$ as linearly independent columns. Show that the following problem is equivalent to maximizing the criterion $J(\mathbf{W}) = \mathrm{Tr}\{\mathbf{s}_{\mathrm{W}}^{-1}\mathbf{s}_{\mathrm{B}}\}$ in PRML (4.50), if we choose suitable weights for d_{nm} and d'_{nm} .

$$\text{Maximize} \sum_{j=1}^{D'} \sum_{k,k',k\neq k}^{K} \sum_{n\in\mathcal{C}_k, m\in\mathcal{C}_{k'}} \|\mathbf{w}_j^{\mathrm{T}}\mathbf{x}_n - \mathbf{w}_j^{\mathrm{T}}\mathbf{x}_m\|^2 \, d'_{nm} \,,$$

$$\text{subject to} \sum_{k}^{K} \sum_{n,m\in\mathcal{C}_k} \|\mathbf{w}_j^{\mathrm{T}}\mathbf{x}_n - \mathbf{w}_j^{\mathrm{T}}\mathbf{x}_m\|^2 \, d_{nm} = 1 \,, j = 1, \ldots, D' \,.$$

   Give a brief explanation of this result.

2. **(20 points)** Bishop PRML Exercise 4.15.

   (1) Show that the Hessian matrix $\mathbf{H}$ for the logistic regression model, given by PRML (4.97), is positive definite.

   (2) Show that the error function $E(\mathbf{w})$ in PRML (4.90) is a convex function of $\mathbf{w}$ and that it has a unique minimum.

3. **(50 points)** Fisher's Linear Discriminant (Matlab Experiments)

   In this problem you will try binary classification with the following four data sets: australian.mat, fourclass.mat, german.mat, heart.mat. Each data set consists of training data and test data. The labels are either $+1$ or -1. Your program needs to learn the Fisher's linear discriminants from the training data, and then use the linear discriminants to predict the labels of the test data.

(1) For each classification task, use the found direction to project the training data to one dimension, and draw a plot of the projected data.

(2) Come up with a scheme for choosing the threshold y_0 so that you can classify a new data point $\mathbf{x}$ by checking if $y(\mathbf{x}) \geq y_0$ or $y(\mathbf{x}) < y_0$.

(3) Compute the error rates of predicting the labels of the test data. Report the four results in a table.

(As before, your code must be included in the PDF.)