Introduction to Pattern Recognition Homework 2 Report

Andrés Ponce(彭思安) 0616110

April 28, 2020

1 Coding

1.1 Compute the mean vectors $m_i, (i = 1, 2)$ of each 2 classes on training data

To calculate the **mean vectors** for each class, we need to know which of the input vectors belong to C_1 and C_2, respectively. This is done by just checking which rows in `y_test` are 0 and 1, since those tell us which class the test input vectors belong to. After separating them and getting the mean for each column separately, we end with the result

$$m_1 = [2.47107265\ 1.97913899] \quad m_2 = [1.82380675\ 3.03051876]$$

1.2 Compute the within-class scatter matrix S_W on training data.

The within class matrix S_W measures how much the points in ever class differ from the mean. The within class variance for the dataset is the sum of the within-class variance for all the classes in our dataset. The result is a $2x2$ matrix with the values:

$$S_W = \begin{bmatrix} 140.40035447 & -5.30881553 \\ -5.30881553 & 138.14297637 \end{bmatrix}$$

1.3 Compute the between-class scatter matrix S_B on training data.

The between-class covariance matrix S_B measures how much the two class means vary from each other. The result is another $2x2$ matrix:

$$S_B = \begin{bmatrix} 0.41895314 & -0.68052227 \\ -0.68052227 & 1.10539942 \end{bmatrix}$$

1.4 Compute the Fisher's linear discriminant W on training data.

1.5 Project the testing data by linear discriminant to get the class prediction by nearest-neighbor rule and calculate your accuracy score on testing data.

1.6 Plot the 1) best projection line on the training data and slow the slope and intercept on the title 2)colorize the data with each class 3) project all data points on your projection line.

2 Questions

2.1 Show that maximization of the class separation criterion given by $L(\lambda, w) = w^T(m_2 - m_1) + \lambda(w^T w - 1)$ with respect to w, using a Lagrange multiplier to enforce the constraint $w^T w = 1$, leads to the result that $w \propto (m_2 - m_1)$.

Lagrange Multipliers allow us to maximize a function L given a constraint g. To do this, we have the initial equation $\nabla L(w, \lambda) = \alpha \nabla g(w, \lambda)$. Where α is the Lagrange Multiplier in this equation. Since we apply the multiplier to each variable individually, then we have to find the multiplier based on the component functions for L and g. Thus we have to solve the system of equations

$$L_\lambda = \alpha g_\lambda \qquad L_w = \alpha g_w$$

So, we have to differentiate with respect to λ and w, since we are dealing with the gradient vector. When we set up the equations, we get the system of equations

$$w^T w - 1 = 0 \tag{1a}$$

$$m_2 - m_1 = \alpha \tag{1b}$$

$$w^T w = 1 \tag{1c}$$

Then, multiplying (1c) by (1b), we get

$$(m_2 - m_1)(w^T w) = \alpha$$

which, when divided by $m_2 - m_1$ gives

$$w^T w = \frac{\alpha}{m_2 - m_1}$$

Since $w_T w$ is proportional to $m_2 - m_1$, then w will also be. Thus, using Lagrange Multipliers we have shown that $w \propto m_2 - m_1$

2.2 Using eq 2 and eq 3, derive the result eq 4 for the posterior class probability in the two-class generative model with Gaussian densities, and verify the results eq 5 and eq 6 for the parameters w and w_0.

$$\frac{1}{1 + exp(-\alpha) = \phi(\alpha)} \tag{2}$$

$$a = ln\frac{p(x|C_1)p(C_1)}{p(x|C_2)p(C_2)} \tag{3}$$

$$p(C_1|x) = \phi(w^T x + w_0) \tag{4}$$

$$w = \sum^{-1}(\mu_1 - \mu_2) \tag{5}$$

$$w_0 = -\frac{1}{2}\mu_1^T \sum{}^{-1}\mu_1 + \frac{1}{2}\mu_2^T \sum{}^{-1}\mu_2 + ln\frac{p(C_1)}{p(C_2)} \tag{6}$$