ENGG 5202: Assignment #2

Due on Thursday, March 10, 2015

Kai Chen

Problem 1

1.1

$$\begin{aligned} l(\theta_1) &= \log p(D_1|\omega_1,\theta_1) \\ &= \log p(x_{11}|\omega_1,\theta_1) + \log p(x_{12}|\omega_1,\theta_1) \\ &= 2\log\theta_1 - \theta_1(x_{11}+x_{12}) \end{aligned}$$

Let

$$\nabla l(\theta_1) = 0$$

We get

$$\hat{\theta_1} = \frac{2}{x_{11}+x_{12}} = \frac{1}{3}$$

Similarly,

$$\hat{\theta_2} = \frac{2}{x_{21}+x_{22}} = \frac{1}{6}$$

1.2

Given $\theta_1 = \frac{1}{3}$ and $\theta_2 = \frac{1}{6}$, we have

$$p(x|\omega_1) = \begin{cases} 0 & x < 0 \\ \frac{1}{3}\mathrm{e}^{-\frac{x}{3}} & x \geq 0 \end{cases}$$

$$p(x|\omega_2) = \begin{cases} 0 & x < 0 \\ \frac{1}{6}\mathrm{e}^{-\frac{x}{6}} & x \geq 0 \end{cases}$$

$$\begin{aligned} g(x) &= p(\omega_1|x) - p(\omega_2|x) \\ &= \ln\frac{p(x|\omega_1)}{p(x|\omega_2)} - \ln\frac{p(\omega_1)}{p(\omega_2)} \end{aligned}$$

Let

$$g(x) = 0$$

We have

$$\begin{aligned} x^* &= \frac{\ln\theta_1 - \ln\theta_2}{\theta_1 - \theta_2} \\ &= 6\ln 2 \approx 4.16 \end{aligned}$$

1.3

The classification rule is when $x > x^*$, class label is 2, when $0 < x < x^*$, class label is 1.
So the expected classification error

$$\begin{aligned} Error &= \int_0^{x^*} p(x|\omega_2)p(\omega_2)\mathrm{d}x + \int_{x^*}^{\infty} p(x|\omega_1)p(\omega_1)\mathrm{d}x \\ &= \frac{1}{2}\int_0^{6\ln 2} \mathrm{e}^{-\frac{x}{6}}\mathrm{d}x + \frac{1}{2}\int_{6\ln 2}^{\infty}\frac{1}{3}\mathrm{e}^{-\frac{x}{3}}\mathrm{d}x \\ &= \frac{3}{8} \end{aligned}$$

Problem 2

a_{12} is updated in the M step and

$$\begin{aligned}\hat{a}_{12} &= \frac{\sum_{t=2}^{T}\xi_{t-1}(\omega_1,\omega_2)}{\sum_{t=2}^{T}\sum_{j'=1}^{c}\xi_{t-1}(\omega_1,\omega_{j'})} \\ &= \frac{\sum_{t=2}^{T}P(Z_{t-1}=\omega_1,Z_t=\omega_2|X,\theta^{old})}{\sum_{t=2}^{T}\sum_{j'=1}^{c}\xi_{t-1}(\omega_1,\omega_{j'})}\end{aligned}$$

Because a_{12} is initialized as 0, therefore for any $t>1$,we have

$$P(Z_{t-1}=\omega_1,Z_t=\omega_2|X,\theta^{old})=0$$

So in all subsequence updates of the EM algorithm $\hat{a}_{12}=0$.

Problem 3

3.1

The sample size for the "same density" in $\mathcal{R}^d$ is n_1^d.
If $n_1=100$, the sample size needed is 100^{20}.

3.2

$$\begin{aligned}p &= \frac{k/n}{V} \\ &= k/n \\ &= nl_d^d(p)\end{aligned}$$

So

$$l_d(p)=p^{\frac{1}{d}}$$

We can calculate

$$\begin{aligned}l_5(0.01)&=0.398 \\ l_5(0.1)&=0.63 \\ l_{20}(0.1)&=0.89\end{aligned}$$

3.3

If the dimension is high, the edge length of Parzen window will be greater and approximate to 1, leading to larger errors and a "flat" and "uniform" estimated density.

Problem 4

4.1

If it is Gaussian distribution and the class covariances are identical, then the dicision boundary found by two methods will be the same.

4.2

The results are shown in table 1 and figure 1.
I observe that LR is more robust than FLD when dealing with outliners and noises. The two algorithms have similar testing error rates when trained using **train1**, but FLD has significantly larger error rate when trained using **train1_2**, because the label of FLD can be infinite while the sigmoid function in LR turns it to (-1, 1), which can handle such outliners.

Table 1: Dataset 1 results

Train set	Test set	Algorithm	Error rate
train1	test1	FLD	0.07
train1	test1	LR	0.067
train1_2	test1	FLD	0.151
train1_2	test1	LR	0.067

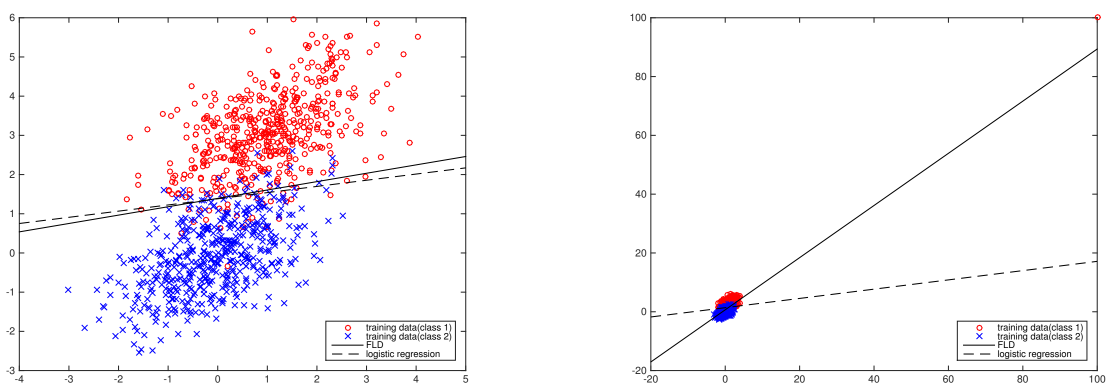

Figure 1: Train data and decision boundaries (left: train1, right: train1_2)

4.3

The error rates are shown in Table 2. The Gaussian asumption is not reasonable for the representation of digit images. The values of x are discrete and binary, so it is not Gaussian. FLD and LR are sensitive to the Gaussian assumption, because the derivation of these methods are based on Gaussian.

Table 2: Dataset 2 results

Train set	Test set	Algorithm	Error rate
train2	test2	FLD	0.23
train2	test2	LR	0.1425