Homework 2, Question 1

Yanlin Li, Student #: 1003770305

February 18, 2021

(a)

The likelihood is:

$$P(x,t|\mu_1,...,\mu_9,\Sigma) = \prod_{n=1}^{N}(\prod_{k=0}^{9}\pi_k N(x_n|\mu_k,\Sigma))^{t_{nk}})$$

The log-likelihood is:

$$\ln P(x,t|\mu_1,...,\mu_9,\Sigma) = \sum_{n=1}^{N}\sum_{k=0}^{9} t_{nk}(\ln \pi_k + \ln N(x_n|\mu_k,\Sigma))$$

(1). For π_k
$max_{\pi k} \ln P$, subject to $\sum_{k=0}^{9}\pi_k = 1$
Using lagrangian:

$$L = \sum_{n=1}^{N}\sum_{k=0}^{9} t_{nk}\ln\pi_k + \lambda(\sum_{k=0}^{9}\pi_k - 1)$$

$$\frac{dL}{d\pi_k} = \sum_{n=1}^{N}\frac{t_{nk}}{\pi_k} + \lambda = 0$$

$$\hat{\pi_k} = -\frac{1}{\lambda}\sum_{n=1}^{N} t_{nk}$$

To solve λ:

$$\sum_{k=0}^{9}\hat{\pi_k} = 1 \Rightarrow -\frac{1}{\lambda}\sum_{k=0}^{9}\sum_{n=1}^{N} t_{nk} = 1$$

Also:

$$\sum_{k=0}^{9}\sum_{n=1}^{N} t_{nk} = 1 \Rightarrow -\frac{N}{\lambda} = 1 \Rightarrow \lambda = -N$$

Thus we get the MLE:

$$\hat{\pi_k} = \frac{1}{N}\sum_{n=1}^{N} t_{nk}$$

(1). For μ_k

$$\frac{d}{d\mu_k}\ln P = \frac{d}{d\mu_k}(\sum_{n=1}^{N}\sum_{k=0}^{9} t_{nk}\ln N(x_n|\mu_k,\Sigma))$$

$$= \sum_{n=1}^{N} \sum_{k=0}^{9} t_{nk} \frac{d}{d\mu_k} (-\frac{1}{2}(x_n - \mu_k)^T \Sigma^{-1}(x_n - \mu_k))$$

$$= \sum_{n=1}^{N} t_{nk} \Sigma^{-1}(x_n - \mu_k) = 0$$

Solve the function to get the MLE:

$$\hat{\mu_k} = \frac{\sum_{n=1}^{N} t_{nk} x_n}{\sum_{n=1}^{N} t_{nk}}$$

(b)

Here is the image for means of each class:

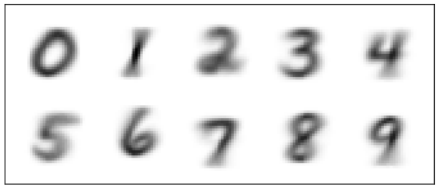

Figure 1: weights

(c)

The number of training images is 10000. The results are:
1. Training accuracy is 0.8939
2. Test accuracy is 0.825

(d)

The accuracy on training data is higher with Gaussian Discriminant Analysis. However, the Logistic Regression model perform better on test data. The differences are small in both cases (within 0.03).

(e)

Here are the two plots required:

[(a)]

[(b)]

Figure 2: (a). 10 images from digit 0 ;(b). 10 images from digit 3

The codes begin on the next page.