HW2 Tao Ni

2019年3月9日 星期六 上午12:08

Q1: a. $p(x) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x-u)^T\Sigma^{-1}(x-u)\right]$

$$l(u,\Sigma|x) = \prod_{t=1}^{N} p(x^t)$$

$$= \prod_{t=1}^{N} \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(x^t-u)^T\Sigma^{-1}(x^t-u)\right]$$

For model 2:

$$\mathcal{L}(u,\Sigma|x) = \sum_{t=1}^{N} -\frac{d}{2}\log 2\pi - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x^t-u)^T\Sigma^{-1}(x^t-u)$$

$$\frac{\partial \mathcal{L}}{\partial \Sigma^{-1}} = \frac{N}{2}\Sigma - \frac{1}{2}\frac{\partial}{\partial \Sigma^{-1}}\sum_{t=1}^{N} tr\left[(x^t-u)^T(x^t-u)\Sigma^{-1}\right]$$

$$= \frac{N}{2}\Sigma - \frac{1}{2}\sum_{t=1}^{N}(x^t-u)(x^t-u)^T = 0$$

$$\therefore \Sigma = \frac{\sum_{t=1}^{N}(x^t-u)(x^t-u)^T}{N}$$

For model 3, $\Sigma = \alpha I$

$$\mathcal{L}(u,\alpha I|x) = \sum_{t=1}^{N} -\frac{d}{2}\log 2\pi - \frac{d}{2}\log\alpha - \frac{1}{2\alpha}(x^t-u)^T(x^t-u)$$

$$\frac{\partial \mathcal{L}}{\partial \alpha} = 0 \Longrightarrow -\frac{1}{2\alpha}d\cdot N + \frac{1}{2\alpha^2}\sum_{t=1}^{N}(x^t-u)^T(x^t-u) = 0$$

$$\therefore \alpha = \frac{\sum_{t=1}^{N}(x^t-u)^T(x^t-u)}{d\cdot N}$$

$$\therefore \alpha_i = \frac{\sum_{t=1}^{N} r_i^t(x^t-u)^T(x^t-u)}{d\cdot \sum_{t=1}^{N} r_i^t}$$

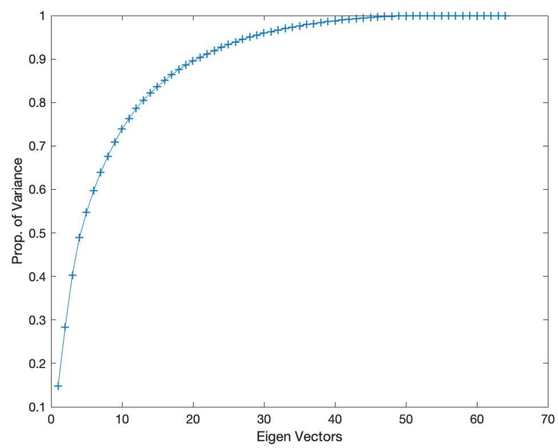
c.

Error Rate	Data Pair 1	Data Pair 2	Data Pair 3
Model 1	0.2	0.23	0.12
Model 2	0.17	0.56	0.45
Model 3	0.24	0.55	0.05

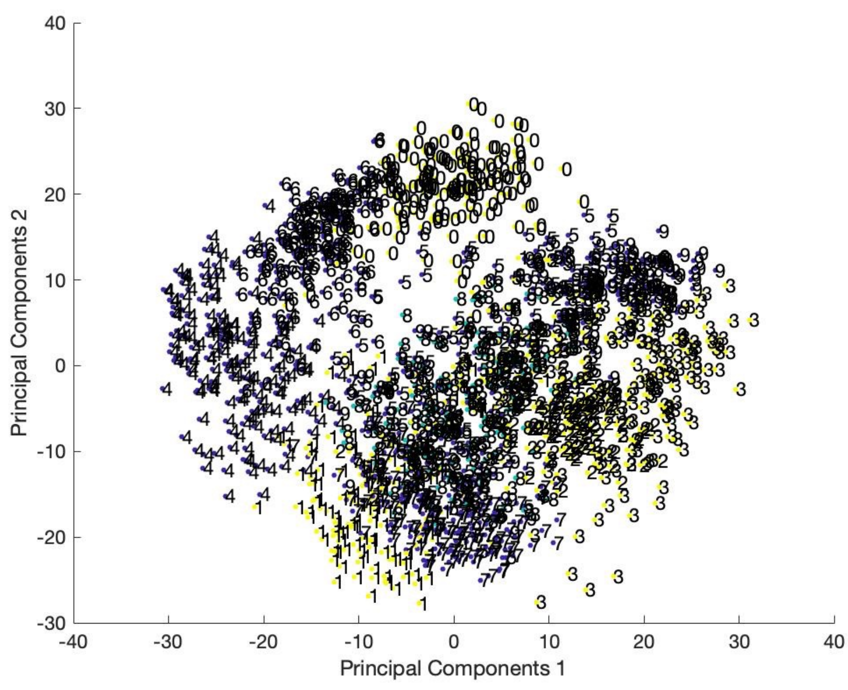
Matching relationships are marked out by the red numbers in the table, based on the rule that each data pair follows a Gaussian distribution of the model that has the lowest prediction error rate.
As we go from model 1 to model 3, we are reducing the number of parameters to be estimated to reduce model complexity. Each data pair achieves the lowest prediction error rate when they are trained by the model that assumes distribution that is identical to the actual distribution of that data pair. Since the three data pairs are sampled from different distributions, we get different matching models for each pair.

Q2. b. plot:

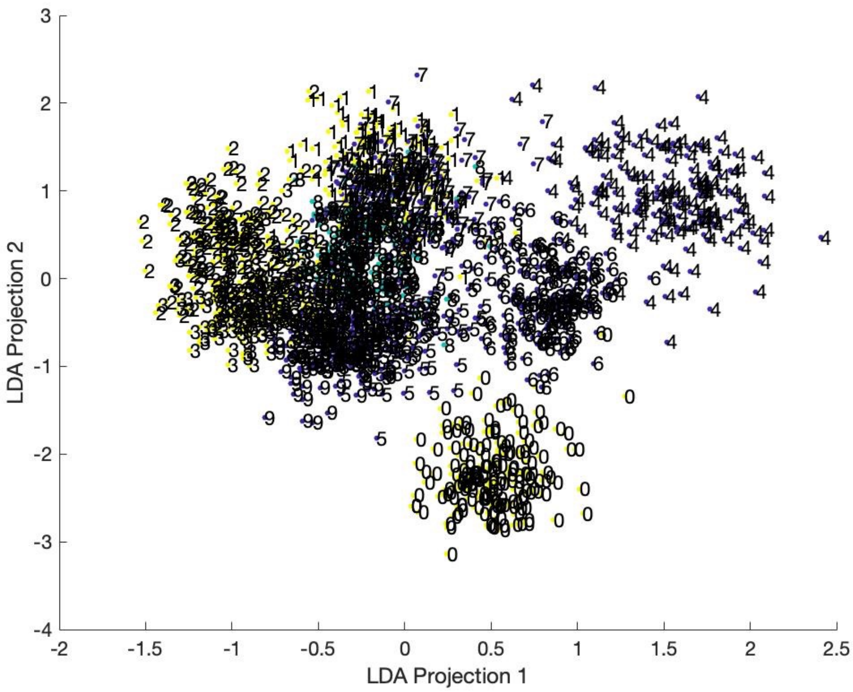
K=21



c.

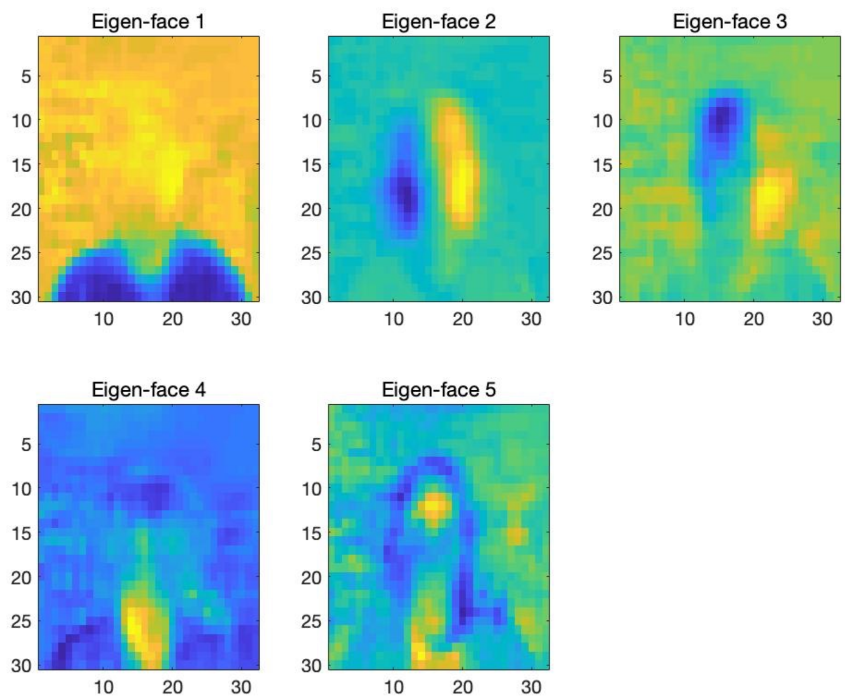


e.



LDA Projection 1
LDA Projection 2

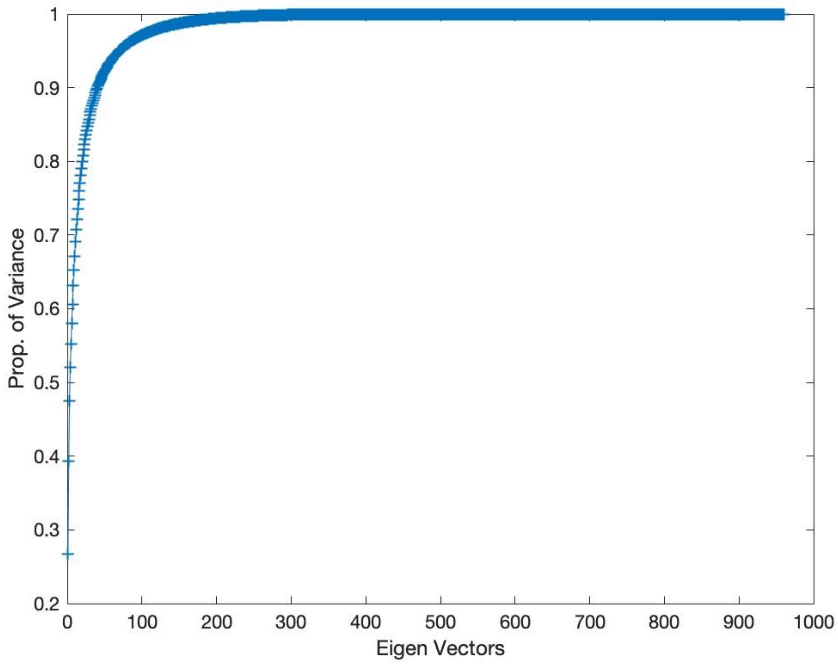
Q3.a.



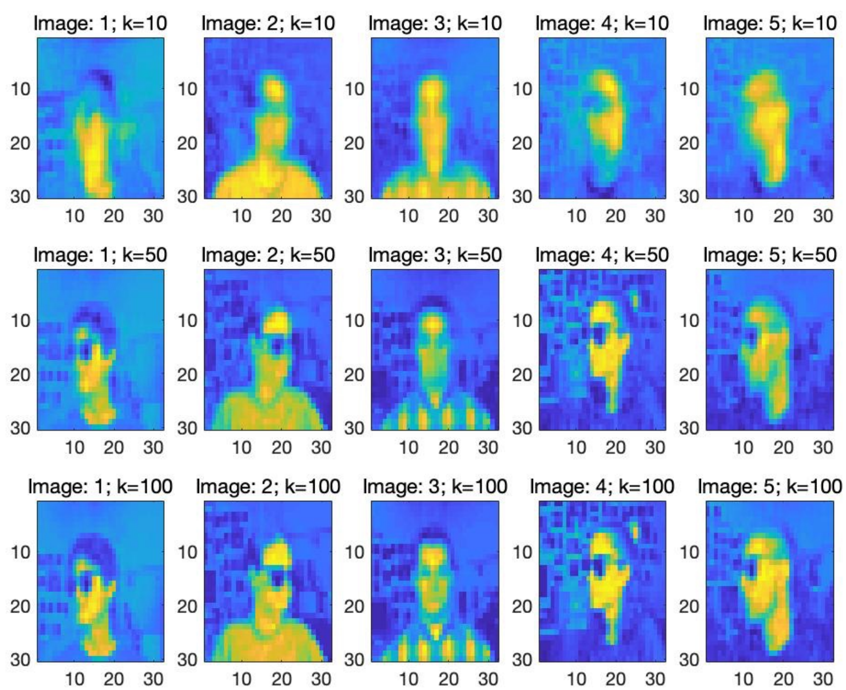
Eigen-face 1
Eigen-face 2
Eigen-face 3
Eigen-face 4
Eigen-face 5

b.

k=41



c.



We can see from the plot, as we increase the number of principle components, these images are becoming more and more close to the original images. This is mainly because that when we keep more and more principal components we are preserving more and more variance among the images.