ISyE 6740 HW 2 Di (Emma) Wu

1. PCA for face recognition.

1. The plot are displayed in ~~Q1~~ attachments.

2. Yes. The test pictures are different when projected on different dominant component. (~~out1 & out2~~ output1 & out2 are diff)

2. PCA: yet another interpretation

① First find x_0 which makes sum of Euclidean distance of x_0 and all data points.

$$L_0(x_0) = \sum_{i=1}^{n} \|x_0 - x_i\|^2$$

$$m = \frac{1}{n}\sum_{i=1}^{n} x_i$$

$$L_0(x_0) = \sum_{i=1}^{n} \|(x_0 - m) - (x_i - m)\|^2$$

$$= \sum_{i=1}^{n} \|x_0 - m\|^2 + \sum_{i=1}^{n} \|x_i - m\|^2 - 2\sum_{i=1}^{n}(x_0 - m)(x_i - m)$$

$$= \sum_{i=1}^{n} \|x_0 - m\|^2 + \sum_{i=1}^{n} \|x_i - m\|^2$$

$\Rightarrow$ $x_0 = m$, sum of distance is ~~the~~ least.

$x_0 = m$

$$f_1(x_i) = m + x_i v$$

$$L_1(a_1, \ldots a_n, v) = \sum_{i=1}^{n} \|x_i - f_1(x_i)\|^2$$

$$= \sum_{i=1}^{n} \|x_i - (m + x_i v)\|^2 = \sum_{i=1}^{n} a_i^2 \|v\|^2 - 2\sum_{i=1}^{n} a_i v^t (x_i - m)$$
$$+ \sum_{i=1}^{n} \|x_i - m\|^2$$

$\Rightarrow \frac{\partial L}{\partial a_i} = 0$

$\Rightarrow a_i = v^t (x_i - m)$

$\frac{\partial L}{\partial v} = 0$

$\Rightarrow L(v) = -v^t (x_i - m)(x_i - m)^t v + \sum_{i=1}^{n} \|x_i - m\|^2$

set $S = \sum_{i=1}^{n} v^t (x_i - m)(x_i - m)^t v$

$u = v^t S v - \lambda (v^t v - 1)$

set $\frac{\partial u}{\partial v} = 0$

$S v = \lambda v$

$\Rightarrow \frac{S}{n-1} v = \frac{\lambda}{n-1} v$

That is the first principal component.

3. Use Euclidean distance ~~as~~.
Choose 10 nearest neibors. Since dataset is large.
See plot in ~~the~~ attachments
~~code is from~~ code use standard isomap algorithms.

4. (a).

$$l(\Delta, h; x) = \sum_i n_i \log n_i$$

(b)

$$l(h) = l(\Delta, h; x) + \lambda(\sum_i \Delta_i h_i - 1)$$

$$l'(h) = \frac{n_i}{h_i} - \lambda\Delta_i$$

$$l'(h) = 0$$

$$\sum_{i} \frac{1}{h_i} I = \lambda \sum_i \Delta_i$$

$$\lambda(\sum_i \Delta_i h_i - 1) = 0$$

$$\sum_i n_i = N$$

$$\Rightarrow \lambda = N.$$

$$h_i = \frac{n}{N\Delta_i}$$

(c) non-parametric dencity estimation does not use F.

non-parametric doesn't mean no parameters.

It's can not be explained by fixed amount of parameters.

- Histogram is ~~not the~~ a good way to estinate density for high-dimensional data.

we can also use kernel density estimation.

- parametric method needs to make assumptions on prob density. F.

5. $$p(x) = \sum_{z \in Z} P(z) p(x|z) = \sum_{z \in Z} \prod_{k=1}^{K} \pi_k^{z_k} N(x|\mu_k, \Sigma_k)^{z_k}$$

b/c $\sum_k \pi_k = 1$

$$p(x) = \sum_{k=1}^{K} \pi_k N(x|\mu_k, \Sigma_k)$$

(b) $$p(z_k^n = 1|x_n) = \frac{p(x_n|z_k^n - 1) p(z_k^n - 1)}{\sum_{i=1}^{K} p(x_n|z_i^n) p(z_i^n)}$$

$$= \frac{\pi_k N(x_n|\mu_k, \Sigma_k)}{\sum_{i=1}^{K} \pi_i N(x_n|\mu_i, \Sigma_i)}$$

(c)

$$l = \log p(x|\pi, \mu, \Sigma) = \sum_{n=1}^{N} \log \sum_{k=1}^{K} \pi_k N(x_n|\mu_k, \Sigma_k)$$

$$-\sum_{n=1}^{N} \frac{\pi_k N(x_n|\mu_k, \Sigma_k)}{\sum_{i=1}^{K} \pi_i N(x_i|\mu_i, \Sigma_i)} (x_n - \mu_k) = 0$$

$$\mu_k = \frac{\sum_{n=1}^{N} p(z_k^n|x_n) x_n}{\sum_{n=1}^{N} p(z_k^n|x_n)} \qquad \frac{\partial l}{\partial \Sigma_k} = \frac{\sum_{n=1}^{N} p(z_k^n|x_n)(x_n - \mu_k)^2}{\sum_{n=1}^{N} p(z_k^n|x_n)}$$

$$\sum_{k=1}^{K} \pi_k = 1$$

$$\log p(X \mid \pi, \mu, \Sigma) + \lambda\left(\sum_{k=1}^{K} \pi_k - 1\right) = 0$$

$$\sum_{n=1}^{N} \frac{\mathcal{N}(x_n \mid \mu_k, \Sigma_k)}{\sum_{i=1}^{K} \pi_i \mathcal{N}(x_n \mid \mu_i, \Sigma_i)} + \lambda = 0$$

$$\pi_k = \frac{\sum_{n=1}^{N} p(z_k^n \mid x_n)}{N}$$

(d) $\Sigma = \varepsilon I$

$$p(x \mid \mu_k, \Sigma) = \exp\left(-\frac{\|x - \mu_k\|^2}{2\varepsilon}\right) / (2\pi\varepsilon)^{\frac{1}{2}}$$

$$l = \sum_N \sum_K z_k^n \{\ln \pi_k + \log \mathcal{N}(x_n \mid \mu_i, \Sigma_i)\}$$

$$\mathbb{E}(l) = \sum_N \sum_K p(z_k^n \mid x_n) \{\ln \pi_k + \log \mathcal{N}(x_n \mid \mu_i, \Sigma_i)\}$$

$$\Rightarrow p(z_k^n \mid x_n) = \frac{\pi_k \exp\left(-\frac{\|x - \mu_k\|^2}{2\varepsilon}\right)}{\sum_{i=1}^{K} \pi_i \exp\left(-\frac{\|x - \mu_k\|^2}{2\varepsilon}\right)}$$

(e) $$\mathbb{E}(x) = \int x \sum_{k=1}^{K} \pi_k p(x \mid k)\, dx = \sum_{k=1}^{K} \int x\, p(x \mid k)\, dx = \sum_{k=1}^{K} \pi_k \mu_k$$

$$\mathrm{cov}(x) = \mathbb{E}(x - \mathbb{E}(x))^2 = \sum_{k=1}^{K} \pi_k(\Sigma_k + \mu_k^2) - \left(\sum_{k=1}^{K} \pi_k \mu_k\right)^2$$