C. Sai Sreenivas
CS17BTECH11012

EE5601: Representation Learning

2. If the ~~reduced~~ covariance matrix of the reduced matrix is diagonal and if the squared projection error (distance b/w data points and the low-dimensional linear subspace) is being minimized (or) not.

Where PCA fails :

(1) If the values in a single dimension are so large then PCA ends up in taking that dimension as the principal component as the variance is high.

(2) PCA tries to map the data to a lower dimensional linear subspace but if the data is spread non-linearly in 1D then applying PCA leads to a straight line.

(3) It doesn't take the information of the classes when it is used in the classification.

3.

(a) **Binomial** : $\{x_1, x_2 \dots x_m\} \Rightarrow$ i.i.d samples.

$$f_X(x, n; \theta) = \binom{n}{x} \cdot (p)^x \cdot (1-p)^{n-x} \qquad (\theta = p)$$

$$L(x; n; \theta) = \prod_{i=1}^{m} f_X(x_i, n; \theta)$$

$$\Rightarrow \log(L(X, n; \theta)) = \log\left(\prod_{i=1}^{m} f_X(x_i, n; \theta)\right)$$

$$= \log(f_X(x_1, n; \theta)) + \log(f_X(x_2, n; \theta)) + \dots$$

$$= \sum_{i=1}^{m} \log(f_X(x_i, n; \theta))$$

$$\Rightarrow \log(L(X, n; \theta)) = \sum_{i=1}^{m} \log\left(\binom{n}{x_i} \cdot p^{x_i} \cdot (1-p)^{n-x_i}\right)$$

$$= \sum_{i=1}^{m} \log\binom{n}{x_i} + \log p^{x_i} + \log(1-p)^{n-x_i}$$

$$= \sum_{i=1}^{m} \log\binom{n}{x_i} + x_i \cdot \log p + (n - x_i)\log(1-p)$$

$$\Rightarrow \log(L(X, n; \theta)) = \sum_{i=1}^{m} \log\binom{n}{x_i} + \left(\sum_{i=1}^{m} x_i\right) \cdot \log p + \cancel{mn\sum}$$

$$+ \left(mn - \sum_{i=1}^{m} x_i\right) \cdot \log(1-p)$$

$$P_{MLE} = \arg\max_{p} \log[L(X,n;\theta)]$$

$$\Rightarrow \frac{\partial}{\partial p} \log[L(X,n;\theta)] = 0$$

$$\Rightarrow 0 + \left(\sum_{i=1}^{m} x_i\right) \cdot \frac{1}{p} + (-1) \cdot \frac{1}{1-p} \cdot \left(mn - \sum_{i=1}^{m} x_i\right) = 0$$

$$\Rightarrow \left(\sum_{i=1}^{m} x_i\right) \cdot \frac{1}{p} = \frac{1}{1-p} \cdot \left(mn - \sum_{i=1}^{m} x_i\right)$$

$$\Rightarrow \boxed{P_{MLE} = \frac{\sum_{i=1}^{m} x_i}{mn}}$$

(b) Poisson: $\{x_1, x_2, \cdots x_n\} \Rightarrow$ i.i.d samples

$$f_X(x;\theta) = \frac{e^{-\lambda} \cdot \lambda^{x}}{x!} \qquad (\theta = \lambda)$$

$$L(X;\theta) = \prod_{i=1}^{n} f_X(x_i;\theta)$$

$$\Rightarrow \log(L(X;\theta)) = \sum_{i=1}^{n} \log(f_X(x_i;\theta))$$

$$= \sum_{i=1}^{n} \left[-\lambda + x_i \log\lambda - \log x_i!\right]$$

$$\lambda_{MLE} = \arg\max_{\lambda} \log[L(X;\theta)]$$

$$\Rightarrow \frac{\partial}{\partial \lambda} \log[L(X;\theta)] = 0$$

$$\Rightarrow (-1)\cdot\left(\sum_{i=1}^{n} 1\right) + \frac{1}{\lambda}\cdot\sum_{i=1}^{n} x_i + 0 = 0$$

$$\Rightarrow \quad n = \frac{1}{\lambda}\cdot\sum_{i=1}^{n} x_i \Rightarrow \boxed{\lambda_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

(c) Exponential: $\{x_1, x_2 \dots x_n\} \Rightarrow$ i.i.d samples

$$f_X(x;\theta) = \lambda \cdot e^{-\lambda x}$$

$$L(X;\theta) = \prod_{i=1}^{n} f_X(x_i;\theta)$$

$$\Rightarrow \log(L(X;\theta)) = \sum_{i=1}^{n} \log(f_X(x_i;\theta))$$

$$= \sum_{i=1}^{n} \log(\lambda \cdot e^{-\lambda x_i})$$

$$= \sum_{i=1}^{n} [\log\lambda - \lambda x_i]$$

$$\log(L(X;\theta)) = \left(\sum_{i=1}^{n} 1\right).\log\lambda - \lambda \sum_{i=1}^{n} x_i$$

$$= n.\log\lambda - \lambda \sum_{i=1}^{n} x_i$$

$$\lambda_{MLE} = \arg\max_{\lambda} \log[L(X;\theta)]$$

$$\Rightarrow \frac{\partial}{\partial \lambda} \log[L(X;\theta)] = 0$$

$$\Rightarrow n \cdot \frac{1}{\lambda} - 1 \cdot \sum_{i=1}^{n} x_i = 0$$

$$\Rightarrow \therefore \boxed{\lambda_{MLE} = \frac{n}{\sum_{i=1}^{n} x_i}}$$

(d) Gaussian : $\{x_1, x_2, \ldots x_n\} \Rightarrow$ i.i.d samples

$$f_X(x;\underline{\theta}) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \quad (\underline{\theta} = [\mu, \sigma]^T)$$

$$L(X;\theta) = \prod_{i=1}^{n} f_X(x_i;\theta)$$

$$\log[L(X;\underline{\theta})] = \sum_{i=1}^{n} \log(f_X(x_i;\underline{\theta}))$$

$$= \sum_{i=1}^{n} \log\left[\frac{1}{\sqrt{2\pi}\,\sigma} \cdot \exp\left[\frac{-(x_i-\mu)^2}{2\sigma^2}\right]\right]$$

$$\log[L(X;\underline{\theta})] = \sum_{i=1}^{n} \left[\log\left[\frac{1}{\sqrt{2\pi}\,\sigma}\right] - \frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

$$= n \cdot \log\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) - \frac{1}{2\sigma^2} \cdot \sum_{i=1}^{n} (x_i-\mu)^2$$

$$\mu_{MLE} = \arg\max_{\mu} \log[L(x;\underline{\theta})]$$

$$\Rightarrow \frac{\partial}{\partial\mu} \log[L(x;\underline{\theta})] = 0$$

$$\Rightarrow 0 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} 2(x_i - \mu)(-1) = 0$$

$$\Rightarrow \sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\Rightarrow \sum_{i=1}^{n} x_i - \mu n = 0 \rightarrow \boxed{\mu_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n}}$$

$$\sigma^2_{MLE} = \arg\max_{\sigma} \log[L(x;\underline{\theta})]$$

$$\Rightarrow \frac{\partial}{\partial\sigma} \log[L(x;\underline{\theta})] = 0$$

$$\Rightarrow n \cdot \sqrt{2\pi}\sigma \cdot \frac{1}{\sqrt{2\pi}} \cdot \frac{-1}{\sigma^2} - \frac{-2}{2\sigma^3} \cdot \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

$$\Rightarrow \frac{-n}{\sigma} + \frac{1}{\sigma^3} \cdot \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

$$\rightarrow n\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2$$

$$\Rightarrow \boxed{\sigma^2_{MLE} = \frac{\sum_{i=1}^{n} (x_i - \mu)^2}{n}}$$

(e) Laplacian: $\{x_1, x_2, \ldots x_n\} \Rightarrow$ i.i.d samples

$$f_X(x; \underline{\theta}) = \frac{1}{2b} \cdot e^{\frac{-|x-\mu|}{b}} \qquad \left(\underline{\theta} = [\mu, b]^T\right)$$

$$L(X; \underline{\theta}) = \prod_{i=1}^{n} f_X(x_i; \underline{\theta})$$

$$\log[L(X; \underline{\theta})] = \sum_{i=1}^{n} \log[f_X(x_i; \underline{\theta})]$$

$$= \sum_{i=1}^{n} \log\left[\frac{1}{2b} \cdot e^{\frac{-|x_i-\mu|}{b}}\right]$$

$$= \sum_{i=1}^{n} \left[-\log 2b - \frac{|x_i-\mu|}{b}\right]$$

$$\log[L(X; \underline{\theta})] = -n\log 2b - \frac{\sum_{i=1}^{n} |x_i-\mu|}{b}$$

$$b_{MLE} = \arg\max_{b} \log[L(X; \underline{\theta})]$$

$$\frac{\partial}{\partial b} \log[L(X; \underline{\theta})] = 0$$

$$\Rightarrow \quad \frac{-n}{b} + \frac{1}{b^2} \cdot \sum_{i=1}^{n} |x_i-\mu| = 0$$

$$\Rightarrow \therefore \quad \boxed{b_{MLE} = \frac{\sum_{i=1}^{n} |x_i-\mu|}{n}}$$

Scanned by CamScanner

$$\mu_{MLE} = \underset{\mu}{\arg\max} \log[L(X;\underline{\theta})]$$

$$\frac{\partial}{\partial \mu} \log[L(X;\underline{\theta})] = 0$$

$$\Rightarrow \quad 0 - \frac{1}{b} \cdot \sum_{i=1}^{n} \frac{\partial |x_i - \mu|}{\partial \mu} = 0 \quad - (1)$$

$$|x_i - \mu| = \begin{cases} x_i - \mu & \text{if } x_i > \mu \\ \mu - x_i & \text{if } x_i < \mu \\ 0 & x_i = \mu \end{cases} \Rightarrow \frac{\partial}{\partial \mu} |x_i - \mu| = \begin{cases} -1 & x_i > \mu \\ 1 & x_i < \mu \\ 0, & x_i = \mu \end{cases}$$

Now for eq (1) to be zero

No. of -1's must be equal to No. of 1's

$\therefore$ μ must be the median

$$\Rightarrow \boxed{\mu_{MLE} = \text{Median}(X)}$$