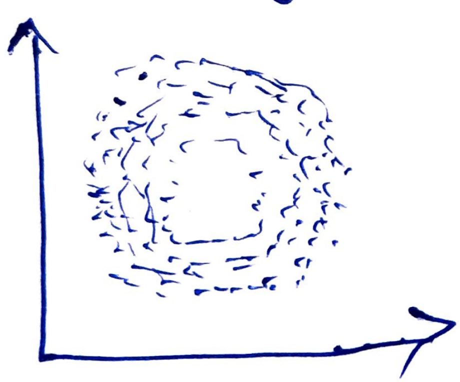
EE5601 : Representation Learning

HW0

2. I am printing the covariance matrix of the output $Y (= PX)$ to check if my non diagonal elements are zero (or very close to zero). This way, I can check if my algorithm is working correctly.

PCA always finds linear principal components to represent data in lower dimension. This may not work always as shown below:



If we apply PCA for this data, it fails to find a good representation of the principal components.

3. a) <u>Binomial</u> <u>distribution:</u>

$$f(x_i;\theta) = {}^{n}C_{x_i}\, p^{x_i} (1-p)^{n-x_i}$$

$$L(X;\theta) = \prod_{i=1}^{N} {}^{n}C_{x_i}\, p^{x_i} (1-p)^{n-x_i}$$

$$\log(L(X;\theta)) = \sum_{i=1}^{N} [\log {}^{n}C_{x_i} + x_i \log p + (n-x_i)\log(1-p)]$$

$$\frac{d}{dp}\log(L(x;\theta)) = \sum_{i=1}^{N} 0 + \frac{x_i}{p} + \frac{(n-x_i)(-1)}{1-p} = 0$$

$$\Rightarrow \frac{\sum x_i}{p} + \frac{\sum(n-x_i)(-1)}{1-p} = 0$$

$$\Rightarrow (1-p)\sum x_i - p\sum(n-x_i) = 0$$

$$\Rightarrow \sum x_i - \cancel{p\sum x_i} - p(nN) + \cancel{p\sum x_i} = 0$$

$$\Rightarrow p = \frac{\sum x_i}{nN}$$

b) Poisson Distribution:

$$f(x_i;\theta) = \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

$$L(x;\theta) = \prod_{i=1}^{N} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!}$$

$$\log(L(x;\theta)) = \sum_i (x_i \log\lambda - \lambda - \log x_i!)$$

$$\frac{d}{d\lambda}\log(L(x;\theta)) = \frac{\sum x_i}{\lambda} - \sum_i 1 = 0$$

$$\Rightarrow \lambda = \frac{\sum x_i}{N}$$

c) Exponential Distribution:

$$f(x_i;\theta) = \lambda e^{-\lambda x_i}$$

$$L(x;\theta) = \prod_{i=1}^{N} \lambda e^{-\lambda x_i}$$

$$\log(L(x;\theta)) = \sum_i \log\lambda - \lambda x_i$$

$$\frac{d}{d\lambda}\log(L(x;\theta)) = \frac{\sum_i 1}{\lambda} - \sum x_i = 0$$

$$\Rightarrow \lambda = \frac{N}{\sum x_i}$$

d) <u>Gaussian Distribution:</u>

$$f(x_i;\theta) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$L(X;\theta) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

$$\log(L(X;\theta)) = \sum_{i=1}^{N} -\log(\sqrt{2\pi}\,\sigma) - \frac{(x_i-\mu)^2}{2\sigma^2}$$

$$\frac{\partial}{\partial\mu} \log(L(X;\theta)) = 0 - \frac{\sum(x_i-\mu)}{\sigma} = 0$$

$$\Rightarrow \quad \mu = \frac{\sum x_i}{N}$$

$$\frac{\partial}{\partial\sigma} \log(L(X;\theta)) = \frac{-N}{\sigma} + \frac{\sum_i (x_i-\mu)^2}{\sigma^3} = 0$$

$$\Rightarrow \quad \sigma^2 = \frac{\sum_i (x_i-\mu)^2}{N}$$

e) <u>Laplacian</u> <u>Distribution:</u>

$$f(x_i;\theta) = \frac{1}{2b} e^{-\frac{|x_i-\mu|}{b}}$$

$$L(x;\theta) = \prod_{i=1}^{N} \frac{1}{2b} e^{-\frac{|x_i-\mu|}{b}}$$

$$\log(L(x;\theta)) = \sum_{i=1}^{N} -\log(2b) - \frac{|x_i-\mu|}{b}$$

The value of μ that maximizes $\log(L(x;\theta))$ is the one that minimizes $\sum_i |x_i-\mu|$

The minimum occurs when μ is the median

$$\Rightarrow \mu = \text{median}(x_i)_{i=1}^{N}$$

$$\frac{d}{db}\log(L(x;\theta)) = -\frac{N}{b} + \frac{1}{b^2}\sum_i |x_i-\mu| = 0$$

$$\Rightarrow b = \frac{\sum_i |x_i-\mu|}{N}$$