MA234 · Introduction to Big Data Science

Assignment 6

刘乐奇 / 12011327

May 19, 2022

1 Exercise 1

It was done on the website.

2 Exercise 2

1. We have

$$
\begin{aligned}
H(X,Y) &= -\sum_{i=1}^{n}\sum_{j=1}^{n} P(X=x_i, Y=y_j)\log P(X=x_i, Y=y_j) \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{n} P(Y=y_j|X=x_i)P(X=x_i)[\log P(X=x_i) + \log P(Y=y_j|X=x_i)] \\
&= -\sum_{i=1}^{n} P(X=x_i)\log P(X=x_i)\sum_{j=1}^{n} P(Y=y_j|X=x_i) \\
&- \sum_{i=1}^{n} P(X=x_i)(\sum_{j=1}^{n} P(Y=y_j|X=x_i)\log P(Y=y_j|X=x_i)) \\
&= H(X) + H(Y|X)
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
H(X,Y) &= -\sum_{i=1}^{n}\sum_{j=1}^{n} P(X=x_i, Y=y_j)\log P(X=x_i, Y=y_j) \\
&= -\sum_{i=1}^{n}\sum_{j=1}^{n} P(X=x_i|Y=y_j)P(Y=y_j)[\log P(Y=y_j) + \log P(X=x_i|Y=y_j)] \\
&= -\sum_{j=1}^{n} P(Y=y_i)\log P(Y=y_j)\sum_{i=1}^{n} P(X=x_i|Y=y_j) \\
&- \sum_{j=1}^{n} P(Y=y_i)(\sum_{i=1}^{n} P(X=x_i|Y=y_j)\log P(X=x_i|Y=y_j)) \\
&= H(Y) + H(X|Y)
\end{aligned}
$$

2. If X and Y are independent, then $P(X,Y) = P(X)P(Y)$, $P(X|Y) = P(x)$ and $P(Y|X) = P(Y)$. Thus,

$$
\begin{aligned}
I(X;Y) &= H(X) - H(X|Y) \\
&= H(X) + \sum_{j=1}^{n} P(Y=y_i)(\sum_{i=1}^{n} P(X=x_i|Y=y_j)\log P(X=x_i|Y=y_j)) \\
&= H(X) + \sum_{j=1}^{n} P(Y=y_i)(\sum_{i=1}^{n} P(X=x_i)\log P(X=x_i)) \\
&= H(X) - H(X) \\
&= 0
\end{aligned}
$$

3. We have

$$\begin{aligned}
D_{KL}(p(X,Y)||p(X)p(Y)) &= -\sum_i\sum_j p(x_i,y_j)\log\frac{p(x_i)p(y_j)}{p(x_i,y_j)} \\
&= -\sum_i\sum_j p(x_i,y_j)\log(p(x_i)p(y_j)) + \sum_i\sum_j p(x_i,y_j)\log p(x_i,y_j) \\
&= -\sum_i\sum_j p(x_i,y_j)[\log p(x_i)+\log p(y_j)] - H(X,Y) \\
&= -\sum_i\sum_j p(x_i,y_j)\log p(x_i) - \sum_i\sum_j p(x_i,y_j)\log p(y_j) - H(X,Y) \\
&= H(X)+H(Y)-H(X,Y) \\
&= H(X)+H(Y)-(H(Y)+H(X|Y)) \\
&= H(X)-H(X|Y) \\
&= I(X;Y)
\end{aligned}$$

4. By Jensen's inequality: $f(E(x)) \le E(f(x))$, we have

$$\begin{aligned}
D_{KL}(P||Q) &= -\sum_i p_i \log\frac{q_i}{p_i} \\
&= E_{p_i}(-\log\frac{q_i}{p_i}) \\
&\ge -\log E(\frac{q_i}{p_i}) \\
&= -\log\sum_i p_i\frac{q_i}{p_i} \\
&= -\log\sum_i q_i \\
&= -\log(1) = 0
\end{aligned}$$

3 Exercise 3

Y	A	B	C	D
X	a	b	c	d
P	$\frac{1}{2}$	μ	2μ	$\frac{1}{2}-3\mu$

1. By Multinoulli distribution, we have

$$p(z) = \prod_{i=1}^{4} P_i^{z_i}$$

and since

$$\begin{aligned}
&p(x|z_i=1) = p(x|\mu) \\
&i.e.,\ p(x|z) = \prod_{i=1}^{4} p(x|\mu)^{z_i}
\end{aligned}$$

we have

$$p(x) = \sum_z p(z)p(x|z) = \sum_{i=1}^{4} p_i p(x|\mu)$$

So

$$
\begin{aligned}
l(\mu, a, b) &= \log p(\{x_n\}_{n=1}^{4}|\mu) \\
&= \sum_{i=1}^{4} \log \sum_{j=1}^{4} p_j p(x_i|\mu)
\end{aligned}
$$

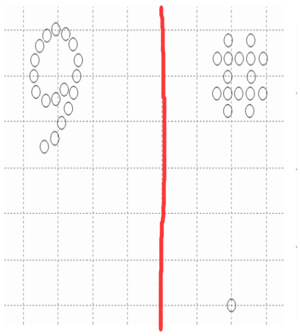
2.

$$Q_i(z_k) = p(z_k = 1|x_i, \mu^{(m)}) = \frac{p_k p(x_i|\mu)}{\sum_{j=1}^{4} p_j p(x_i|\mu)}$$

3.

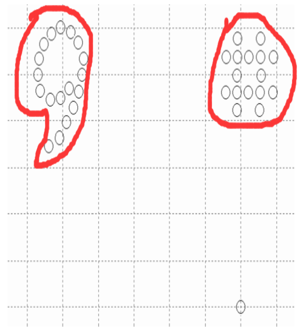
4. It will not always converge to a local optimum of μ. Because it is influenced by the initial value of $\mu^{(m)}$.

4 Exercise 4



(a)

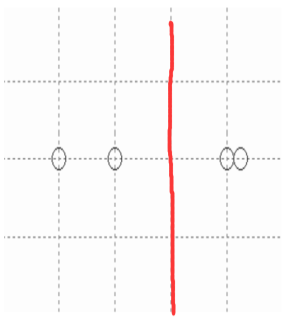
Figure 1: result of min-cut



(a)

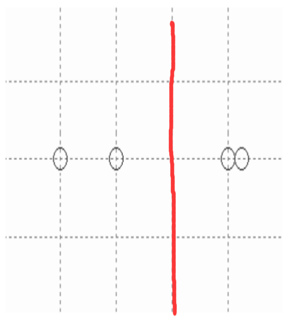
Figure 2: result of min-cut

1. (a)



(b)

Figure 3: result for $\sigma = 50$



(b)

Figure 4: result for $\sigma = 0.5$

   (b)

2. (a) Choose $\sigma^2 = 9$. Since the shortest Euclidean distance between the main two parts is greater than 9 but less than 16 (i.e. $3 < ||x_i - x_j||_2 < 4$), then choosing $\sigma^2 = 9$ is enough.

   (b)

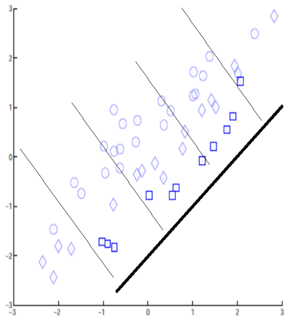

(c) First PCA component

Figure 5: PCA

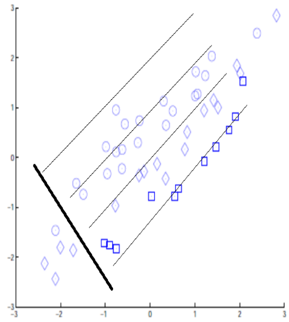

(d) First LDA component

Figure 6: LDA

5 Exercise 5

1.

2. (a) The sample mean of the data set is

$$\overline{x} = [0, 0, 0, 0, 0]$$

(b) Since

$$XX^T = \begin{bmatrix} -3 & -9 & 6 & 0 & 0 & 0 \\ -9 & 27 & -18 & 0 & 0 & 0 \\ 6 & -18 & 12 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & -4 & 2 \\ 0 & 0 & 0 & -4 & 8 & -4 \\ 0 & 0 & 0 & 2 & -4 & 2 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 14 & 14 & 14 & 0 & 0 \\ 14 & 14 & 14 & 0 & 0 \\ 14 & 14 & 14 & 0 & 0 \\ 0 & 0 & 0 & 6 & 6 \\ 0 & 0 & 0 & 6 & 6 \\ 0 & 0 & 0 & 6 & 6 \end{bmatrix}$$

we have

$$det(XX^T - \lambda I) = \begin{vmatrix} -3-\lambda & -9 & 6 & 0 & 0 & 0 \\ -9 & 27-\lambda & -18 & 0 & 0 & 0 \\ 6 & -18 & 12-\lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & 2-\lambda & -4 & 2 \\ 0 & 0 & 0 & -4 & 8-\lambda & -4 \\ 0 & 0 & 0 & 2 & -4 & 2-\lambda \end{vmatrix} = 0$$

by solving it, we have: $\lambda_1 = \lambda_2 = \lambda_3 = 0, \lambda_4 = 12, \lambda_5 = 42$. So the eigen vectors are $u_1 = [0, 0, 0, 0.89, 0.45, 0]^T, u_2 = [0.17, -0.51, -0.85, 0, 0, 0]^T, u_3 = [0, 0, 0, 0.18, -0.37, -0.91]^T, u_4 = [0.95, 0.32, 0, 0, 0, 0]^T, u_5 = [0, 0, 0, -0.41, 0.82, -0.41]^T, u_6 = [-0.27, 0.80, -0.53, 0, 0, 0]^T$. Therefore,

$$D = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3.46 & 0 \\ 0 & 0 & 0 & 0 & 6.48 \end{bmatrix}$$

$$U = \begin{bmatrix} 0 & 0.17 & 0 & 0.95 & 0 & -0.27 \\ 0 & -0.51 & 0 & 0.32 & 0 & 0.80 \\ 0 & -0.85 & 0 & 0 & 0 & -0.53 \\ 0.89 & 0 & 0.18 & 0 & -0.41 & 0 \\ 0.45 & 0 & -0.37 & 0 & 0.82 & 0 \\ 0 & 0 & -0.91 & 0 & -0.41 & 0 \end{bmatrix}$$

Similarly, we can get

$$V = \begin{bmatrix} -0.02 & 0.81 & 0 & 0 & 0.58 \\ 0.71 & -0.39 & 0 & 0 & 0.58 \\ -0.70 & -0.42 & 0 & 0 & 0.58 \\ 0 & 0 & -0.71 & 0.71 & 0 \\ 0 & 0 & 0.71 & 0.71 & 0 \end{bmatrix}$$

The SVD of the data set is

$$X = \begin{bmatrix} 0 & -0.27 \\ 0 & 0.80 \\ 0 & -0.53 \\ -0.41 & 0 \\ 0.82 & 0 \\ -0.41 & 0 \end{bmatrix} \begin{bmatrix} 3.46 & 0 \\ 0 & 6.48 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 0.71 & 0.71 \\ 0.58 & 0.58 & 0.58 & 0 & 0 \end{bmatrix}$$

or equivalently,

$$X = \begin{bmatrix} -0.27 & 0 \\ 0.80 & 0 \\ -0.53 & 0 \\ 0 & -0.41 \\ 0 & 0.82 \\ 0 & -0.41 \end{bmatrix} \begin{bmatrix} 6.48 & 0 \\ 0 & 3.46 \end{bmatrix} \begin{bmatrix} 0.58 & 0.58 & 0.58 & 0 & 0 \\ 0 & 0 & 0 & 0.71 & 0.71 \end{bmatrix}$$

(c) The covariance matrix is

$$\begin{aligned} C &= \frac{1}{n-1}(X - 1_n\overline{x})^T(X - 1_n\overline{x}) \\ &= \frac{1}{4}X^T X \end{aligned}$$

So to find the eigenvalues and eigenvectors of C, we need to do eigen-decomposition on C. From (b), we have computed the eigenvalues of X as $\lambda_1 = \lambda_2 = \lambda_3 = 0, \lambda_4 = 12, \lambda_5 = 42$ and corresponding eigenvectors as $v_1 = [-0.02, 0.71, -0.70, 0, 0]^T, v_2 = [0.81, -0.39, -0.42, 0, 0]^T, v_3 = [0, 0, 0, -0.71, 0.71]^T, v_4 = [0, 0, 0, 0.71, 0.71]^T, v_5 = [0.58, 0.58, 0.58, 0, 0]^T$. So we have the eigenvalues of $C = \frac{1}{4}X^T X = \frac{1}{4}VDU^TUDV^T = \frac{1}{4}VD^2V^T$ as $\lambda_1 = \lambda_2 = \lambda_3 = 0, \lambda_4 = 3, \lambda_5 = 10.5$ and eigenvectors as $w_1 = [-0.02, 0.71, -0.70, 0, 0]^T, w_2 = [0.81, -0.39, -0.42, 0, 0]^T, w_3 = [0, 0, 0, -0.71, 0.71]^T, w_4 = [0, 0, 0, 0.71, 0.71]^T, w_5 = [0.58, 0.58, 0.58, 0, 0]^T$.
Therefore, the first principle component for the original data points is

$$\widetilde{x}_1 = w_1^T x = -0.02x_1 + 0.71x_2 - 0.70x_3$$

(d) Since we have

$$W = \begin{bmatrix} -0.02 & 0.81 & 0 & 0 & 0.58 \\ 0.71 & -0.39 & 0 & 0 & 0.58 \\ -0.70 & -0.42 & 0 & 0 & 0.58 \\ 0 & 0 & -0.71 & 0.71 & 0 \\ 0 & 0 & 0.71 & 0.71 & 0 \end{bmatrix}$$

then

$$W^T X^T X W = \begin{bmatrix} 0.0014 & 0 & 0 & 0 & -0.2436 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 12.0984 & 0 \\ -0.2436 & 0 & 0 & 0 & 42.3864 \end{bmatrix}$$

So for $Y = XW$

$$\begin{aligned} Var(Y) &= \frac{1}{n-1} tr(Y^T Y) \\ &= \frac{1}{6} tr(W^T X^T X W) \\ &= \frac{1}{6} \times 54.4862 \\ &= 9.0810 \end{aligned}$$

(e) The reconstruction error is

$$\begin{aligned} RE(W) &= tr(X^T X) - tr(W^T X^T X W) \\ &= -0.4862 \end{aligned}$$

6 Exercise 6