No.

Date

Homework 5

2. 1. $H(X,Y) = -\sum_i\sum_j P(X=x_i, Y=y_j)\log P(X=x_i, Y=y_j)$

$= -\sum_j P(Y=y_j)\left(\sum_i P(X=x_i|Y=y_j)\log P(X=x_i|Y=y_j)\right)$

$- \sum_j P(Y=y_j)\log P(Y=y_j)\left(\sum_i P(X=x_i|Y=y_j)\right)$

$= H(X|Y) - \sum_j P(Y=y_j)\log P(Y=y_j)$

$= H(X|Y) + H(Y)$

Similarly, $H(X,Y) = H(Y|X) + H(X)$

2. X, Y independent $\Rightarrow H(Y|X) = H(Y) \Rightarrow H(X,Y) = H(Y|X) + H(X)$

$= H(Y) + H(X)$

$I(X;Y) = H(X) - H(X|Y) = 0$

3. $D_{KL}(p(X,Y)\,\|\,p(X)p(Y)) = -\sum_i\sum_j p(x_i,y_j)\log\frac{p(x_i)p(y_j)}{p(x_i,y_j)}$

$= -\sum_i\sum_j p(x_i,y_j)\log\frac{p(x_i)}{p(x_i,y_j)} - \sum_i\sum_j p(x_i,y_j)\log p(y_j)$

$= \sum_i\sum_j p(y_j|x_i)p(x_i)\log p(y_j|x_i) - \sum_i\sum_j p(x_i|y_j)p(y_j)\log p(y_j)$

$= \sum_i p(x_i)\left(\sum_j p(y_j|x_i)\log p(y_j|x_i)\right) - \sum_j p(y_j)\log p(y_j)\left(\sum_i p(x_i|y_j)\right)$

$= -H(Y|X) - \sum_j p(y_j)\log p(y_j)$

$= -H(Y|X) + H(Y)$

$= I(X;Y)$

3. 1. Likelihood function: $L(\mu,a,b) = \left(\frac{1}{2}\right)^a \mu^b (2\mu)^c \left(\frac{1}{2}-3\mu\right)^d$

$l(\mu,a,b) = \log L(\mu,a,b) = a\log\frac{1}{2} + b\log\mu + c\log(2\mu) + d\log\left(\frac{1}{2}-3\mu\right)$

2. $\hat{a}^{(m)} = \frac{P(A)}{P(A)+\hat{P}(B)}h = \frac{1}{1+2\hat{\mu}^{(m)}}h$, $\hat{b}^{(m)} = \frac{\hat{P}(B)}{P(A)+\hat{P}(B)}h = \frac{2\hat{\mu}^{(m)}}{1+2\hat{\mu}^{(m)}}h$

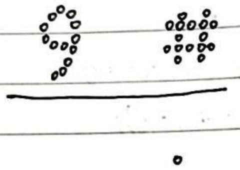
3. $\frac{\partial l}{\partial \mu} = 0$

$\Rightarrow \hat{\mu}^{(m+1)} = \frac{\hat{b}^{(m)}+c}{6(\hat{b}^{(m)}+c+d)} = \frac{c+2\hat{\mu}^{(m)}(h+c)}{6((c+d)+2\hat{\mu}^{(m)}(h+c+d))}$

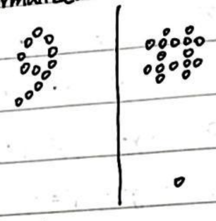
4. Yes. The lower bound of $l(\mu,a,b)$ increase each time.

Date

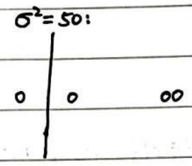
4. 1.(a) Min-cut:



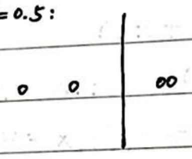
Normalized-cut:



(b) $\sigma^2 = 50$:



$\sigma^2 = 0.5$:



2.(a) Points in the same cluster, their distance ≤ 3.

Between-cluster distance ≥ 4.

σ^2 choose in $[9, 16]$.

(b) "9" has 18 points while "#" has 16.

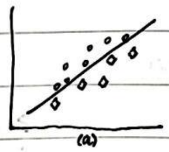
Affinity matrix: $W = \mathrm{diag}(J_{18} - I_{18}, J_{16} - I_{16}, 0)$

Degree matrix: $D = \mathrm{diag}(17 I_{18}, 15 I_{17}, 0)$

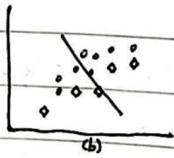
$L = D - W$ has 2 eigenvalues wrt. I_{18}, I_{16}.

So is 17, 15.

5. 1.



(a)



(b)

2. (a) $\mu = (0,0,0,0,0)^T$

(b) $XX^T \Rightarrow \begin{cases} \sigma_1^2 = 42, & u_1^T = \frac{\sqrt{14}}{14}(-1,3,-2,0,0)^T \\ \sigma_2^2 = 12, & u_2^T = \frac{\sqrt{6}}{6}(0,0,0,-1,2,-1)^T \end{cases}$

$X^TX \Rightarrow \begin{cases} \sigma_1^2 = 42, & v_1 = \frac{\sqrt{3}}{3}(-1,-1,-1,0,0) \\ \sigma_2^2 = 12, & v_2 = \frac{\sqrt{2}}{2}(0,0,0,1,1) \end{cases}$

No.
Date . .

$$X = \begin{pmatrix} -\frac{\sqrt{14}}{14} & 0 \\ \frac{3\sqrt{14}}{14} & 0 \\ -\frac{\sqrt{14}}{7} & 0 \\ 0 & -\frac{\sqrt{6}}{6} \\ 0 & \frac{\sqrt{6}}{3} \\ 0 & -\frac{\sqrt{6}}{6} \end{pmatrix} \begin{pmatrix} \sqrt{42} & 0 \\ 0 & 2\sqrt{3} \end{pmatrix} \begin{pmatrix} -\frac{\sqrt{3}}{3} & -\frac{\sqrt{3}}{3} & -\frac{\sqrt{3}}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

(c) $\Sigma = Cov(X) = \frac{1}{5}X^TX$ X^TX's max eigenvalue is $\sqrt{42}$, $v_1 = \frac{\sqrt{3}}{3}(1,1,1,0,0)$

The first principle component is $\frac{\sqrt{3}}{3}(x_1+x_2+x_3)$

(d) $\hat{x}_1, \cdots \hat{x}_6 = \sqrt{3}, -3\sqrt{3}, 2\sqrt{3}, 0, 0, 0$

$Var(\hat{x}) = \frac{1}{6}\sum_{i=1}^{6}(\hat{x}_i - \bar{\hat{x}})^2 = \frac{1}{6}\cdot 42 = 7$

(e) $\frac{1}{6}\sum_{i=1}^{6}\|x_i - \hat{x}_i\|_2^2 = 2$

6. 1. $$\begin{cases} \dfrac{\partial \sum_{i=1}^{N}\|x_i-\mu-V_q\alpha_i\|^2}{\partial \mu} = -2\sum_{i=1}^{N}(x_i-\mu-V_q\alpha_i) = 0 \\ \dfrac{\partial \sum_{i=1}^{N}\|x_i-\mu-V_q\alpha_i\|^2}{\partial \alpha_i} = -2V_q^T(x_i-\mu-V_q\alpha_i) = 0 \end{cases} \Rightarrow \begin{cases} \hat{\mu} = \frac{1}{N}\left(\sum_{i=1}^{N}x_i - V_q\sum_{i=1}^{N}\alpha_i\right) \\ \hat{\alpha}_i = V_q^T(x_i-\mu) \end{cases}$$

2. Suppose $D = diag(d_1, \cdots, d_p)$, $V_q = (x_1, \cdots, x_q)$, $V = (v_1, \cdots, v_p)$

$$\min_{V_q}\sum_{i=1}^{N}\|(x_i-\bar{x}) - V_qV_q^T(x_i-\bar{x})\|^2 = \min_{V_q} Tr(\tilde{X}(I_p - V_qV_q^T)\tilde{X}^T)$$
$$= \min_{V_q} Tr(UDV^T(I_p - V_qV_q^T)VDU^T)$$
$$= \min_{V_q} Tr(DV^T(I_p - V_qV_q^T)VD)$$
$$= \min_{V_q} Tr(D^2(I_p - (V^TV_q)(V^TV_q)^T))$$
$$= \max_{V_q} Tr(D^2(V^TV_q)(V^TV_q)^T)$$

$$Tr(D^2(V^TV_q)(V^TV_q)^T)$$
$$= \sum_{i=1}^{p}\sum_{j=1}^{q}(d_i v_i^T x_j)^2$$
$$= \sum_{i=1}^{p} d_i^2 \sum_{j=1}^{q}(v_i^T x_j)^2 \quad \left(V \text{ and } V_q \text{ are orthogonal} \Rightarrow \begin{cases} \sum_{j=1}^{q}(v_i^Tx_j)^2 \le 1 \\ \sum_{i=1}^{p}(v_i^Tx_j)^2 = 1 \Rightarrow \sum_{i=1}^{p}\sum_{j=1}^{q}(v_i^Tx_j)^2 = q \end{cases} \right)$$
$$= \sum_{i=1}^{p} y_i d_i^2$$
$$\le \sum_{j=1}^{q} d_j^2$$

$$V^TV_q = \begin{pmatrix} I_q \\ 0 \end{pmatrix}_{p\times q} \Rightarrow Tr(D^2(V^TV_q)(V^TV_q)^T) = \sum_{j=1}^{q} d_j^2$$

So, V_q consists of the first q columns of V.