Q1:

$$L(\phi|x) = \sum_t \log P(x^t|\phi)$$

$$= \sum_t \log \sum_{i=1}^{k} P(x^t|C_i)\cdot P(C_i)$$

$$= \sum_t \log \sum_{i=1}^{k} \pi_i \cdot \frac{m!}{x_1^t!\cdots x_n^t!} p_{i1}^{x_1^t}\cdots p_{in}^{x_n^t}$$

①

$$\frac{\partial\left(L - \alpha\left(1 - \sum_{j=1}^{n} P_{ij}\right)\right)}{\partial P_{ij}} = \sum_t \underbrace{\boxed{\frac{\pi_i \frac{m!}{x_1^t!\cdots x_n^t!} p_{i1}^{x_1^t}\cdots p_{ij}^{x_j^t}\cdots p_{in}^{x_n^t}}{\sum_{i=1}^{k} \pi_i \frac{m!}{x_1^t!\cdots x_n^t!} p_{i1}^{x_1^t}\cdots p_{in}^{x_n^t}}}}_{=\gamma(z_i^t)} \cdot \frac{x_j^t}{P_{ij}} + \alpha = 0$$

$$0 = \sum_t \gamma(z_i^t)\cdot \frac{x_j^t}{P_{ij}} + \alpha$$

$$0 = \sum_t \gamma(z_i^t)\, x_j^t + \alpha\cdot P_{ij}$$

$$0 = \sum_j \sum_t \gamma(z_i^t)\cdot x_j^t + \alpha\cdot \underbrace{\sum_j P_{ij}}_{=1}$$

$$= \sum_t \gamma(z_i^t) \underbrace{\sum_j x_j^t}_{=m} + \alpha$$

$$-\alpha = m\sum_t \gamma(z_i^t)$$

$$\Rightarrow P_{ij} = \frac{\sum_t \gamma(z_i^t)\cdot x_j^t}{-\alpha}$$

$$= \frac{\sum_t \gamma(z_i^t)\cdot x_j^t}{m\cdot \sum_t \gamma(z_i^t)}$$

$$③\ \frac{\partial(L-\beta(1-\sum_i \pi_i))}{\partial \pi_i} = \sum_t \frac{\frac{m!}{x_1^t!\cdots x_n^t!} P_{i1}^{x_1^t}\cdots P_{in}^{x_n^t}}{\sum_i \pi_i \frac{m!}{x_1^t!\cdots x_n^t!} P_{i1}^{x_1^t}\cdots P_{in}^{x_n^t}} + \beta = 0.$$

$$0 = \sum_t \frac{\gamma(z_i^t)}{\pi_i} + \beta.$$

$$0 = \sum_t \gamma(z_i^t) + \pi_i \beta$$

$$0 = \sum_t \underbrace{\sum_i \gamma(z_i^t)}_{=1} + \underbrace{\sum_i \pi_i}_{=1} \beta$$

$$-\beta = N$$

$$\Rightarrow \pi_i = \frac{\sum_t \gamma(z_i^t)}{-\beta} = \frac{\sum_t \gamma(z_i^t)}{N}$$

$$= \frac{N_i}{N}, \quad \text{when } N_i = \sum_t \gamma(z_i^t)$$

Then calculate complete log likelihood

$$① \quad P(x \mid z, P) = \prod_{i=1}^{k} P(x \mid P_i)^{z_i}$$

$$② \quad P(z \mid \pi) = \prod_{i=1}^{k} \pi_i^{z_i}$$

$$①② \Rightarrow \log P(x, z \mid P, \pi) = \sum_{t=1}^{N} \log P(x^t, z^t \mid P, \pi)$$

$$= \sum_{t=1}^{N} \log P(x^t \mid z^t, P) \cdot P(z^t \mid \pi)$$

$$= \sum_{t=1}^{N} \log \prod_{i=1}^{k} \left(P(x^t \mid P_i)^{z_i^t} \cdot \pi_i^{z_i^t} \right)$$

$$= \sum_{t=1}^{N} \sum_{i=1}^{k} \left(z_i^t \log(\pi_i) + z_i^t \log(P(x^t \mid P_i)) \right)$$

Note that $P(x^t|P_i) = \frac{m!}{x_1^t!\cdots x_n^t!} \cdot P_{i1}^{x_1^t}\cdots P_{in}^{x_n^t}$

$$\log(P(x^t|P_i)) = \log\left(\frac{m!}{x_1^t!\cdots x_n^t!}\right) + \sum_{j=1}^{n} x_j^t \cdot \log(P_{ij})$$

$$\log P(x, z|P, \pi) = \sum_{t=1}^{N}\sum_{i=1}^{K} z_i^t \left(\log(\pi_i) + \log(P(x^t|P_i))\right)$$

$$= \sum_{t=1}^{N}\sum_{i=1}^{K} z_i^t \left(\log(\pi_i) + \log\left(\frac{m!}{x_1^t!\cdots x_n^t!}\right) + \sum_{j=1}^{n} x_j^t \cdot \log(P_{ij})\right)$$

E-step:

$$\mathbb{E}_z \log(P(x, z|P^l, \pi))$$

$$= \sum_{t=1}^{N}\cdot\sum_{i=1}^{K} \mathbb{E}(z_i^t|x, P^l)\cdot\left(\log(\pi_i) + \log(P(x^t|P_i^l))\right)$$

$$\mathbb{E}(z_i^t|x, P^l) = \mathbb{E}(z_i^t|x^t, P^l) =$$

$$= 1\cdot P(z_i^t=1|x^t, P^l) + 0\cdot P(z_i^t=0|x^t, P^l)$$

$$= P(z_i^t=1|x^t, P^l)$$

$$= \frac{P(x^t|z_i^t=1, P^l)\cdot P(z_i^t=1|P^l)}{P(x^t|P^l)}$$

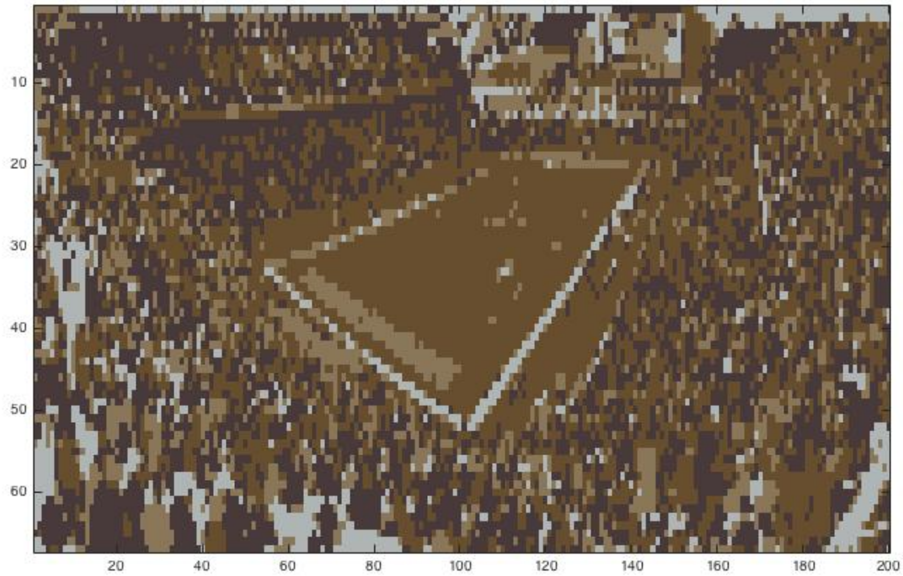
$$= \frac{P(x^t|P_i^l)\cdot\pi_i}{\sum_i P(x^t|P_i^l)\,\pi_i}$$

$$= \frac{\pi_i \frac{m!}{x_1^t!\cdots x_n^t!} P_{i1}^{x_1^t}\cdots P_{in}^{x_n^t}}{\sum_{i=1}^{K}\pi_i \frac{m!}{x_1^t!\cdots x_n^t!} P_{i1}^{x_1^t}\cdots P_{in}^{x_n^t}}$$

$$= \gamma(z_i^t)$$

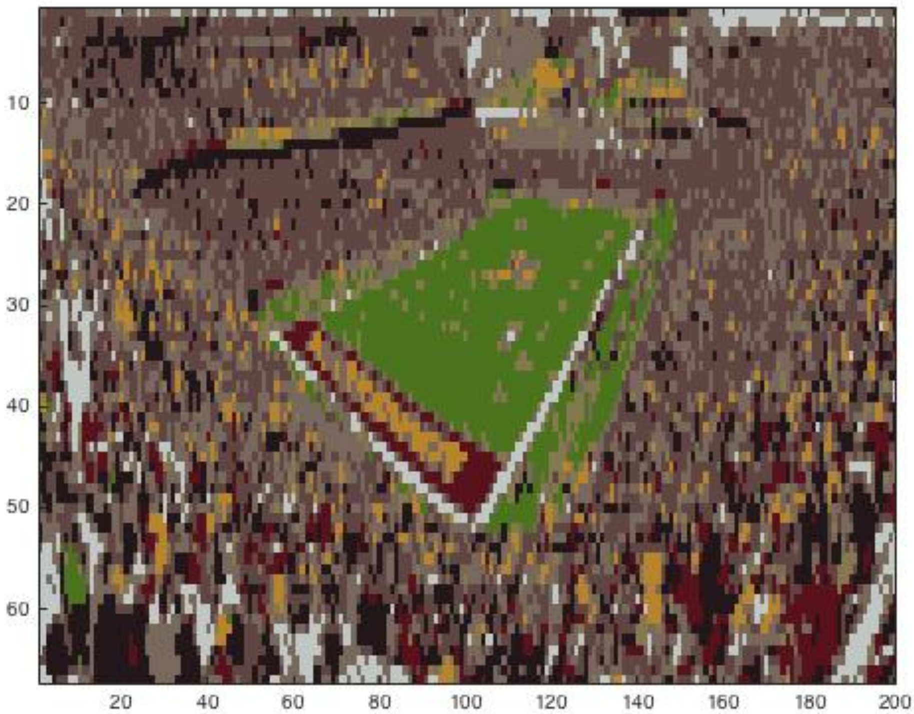
M-step: $P^{l+1} = \arg\max_P \sum_t\sum_i \gamma(z_i^t)\cdot[\log\pi_i + \log(P(x^t|P_i^l))]$

Q2
(a)
k=4,



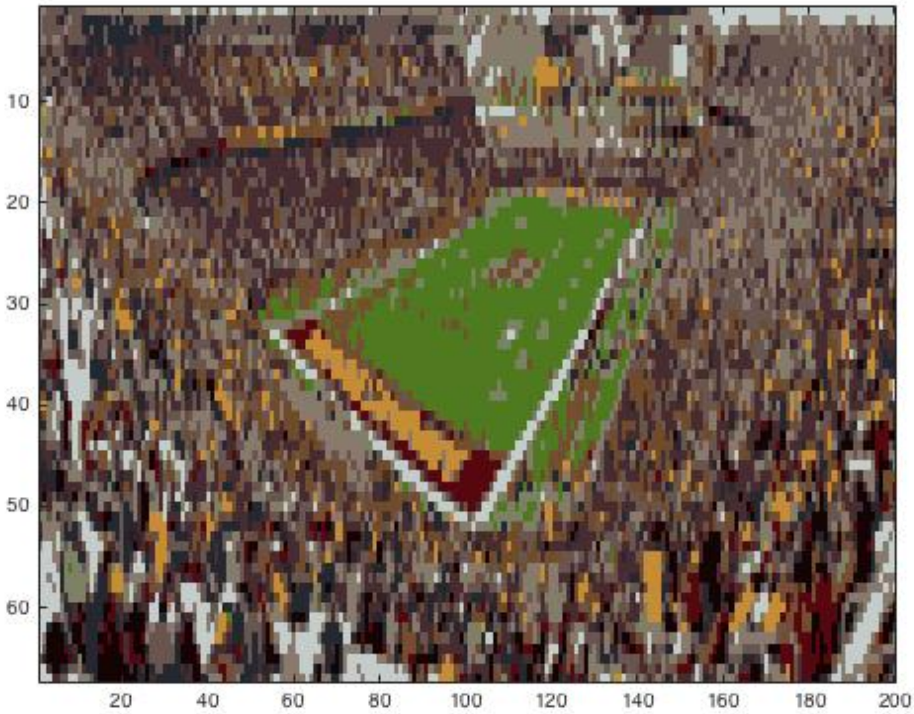
10
20
30
40
50
60
20
40
60
80
100
120
140
160
180
200

k=8,

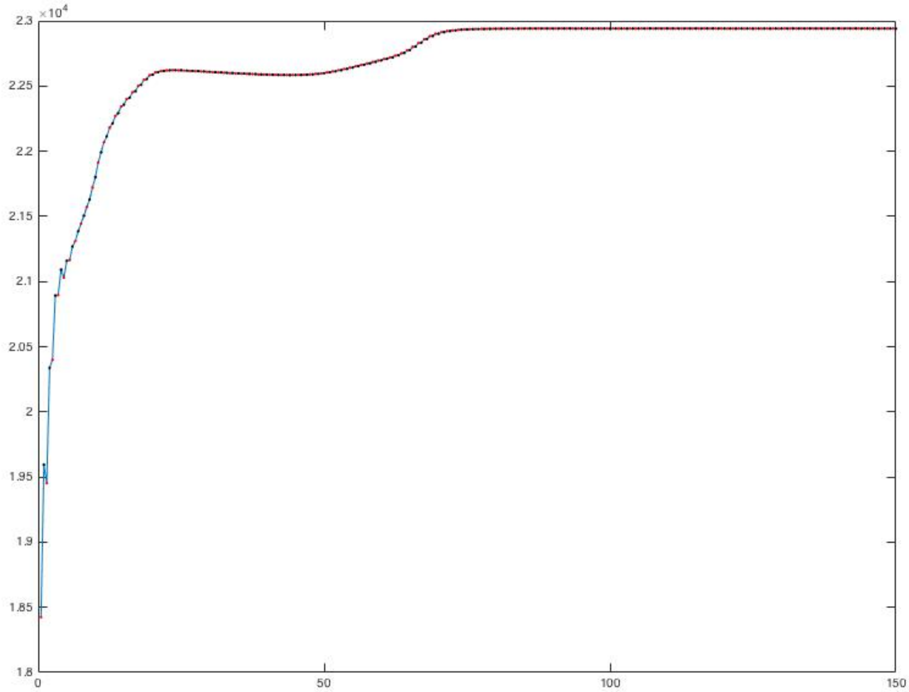


10
20
30
40
50
60
20
40
60
80
100
120
140
160
180
200

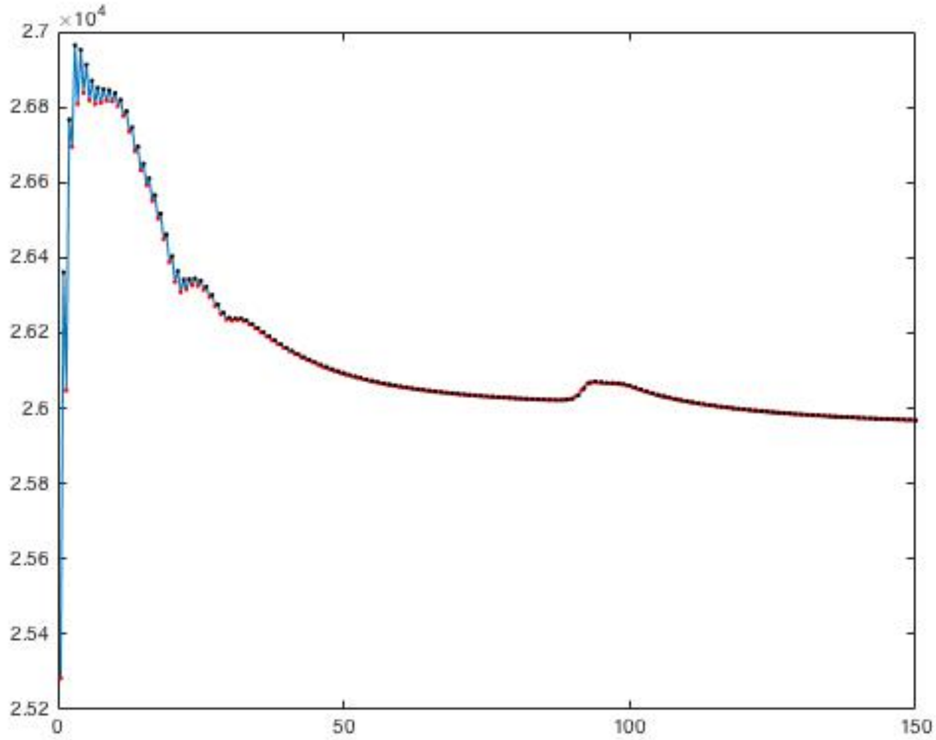
k=12,


10
20
30
40
50
60
20 40 60 80 100 120 140 160 180 200

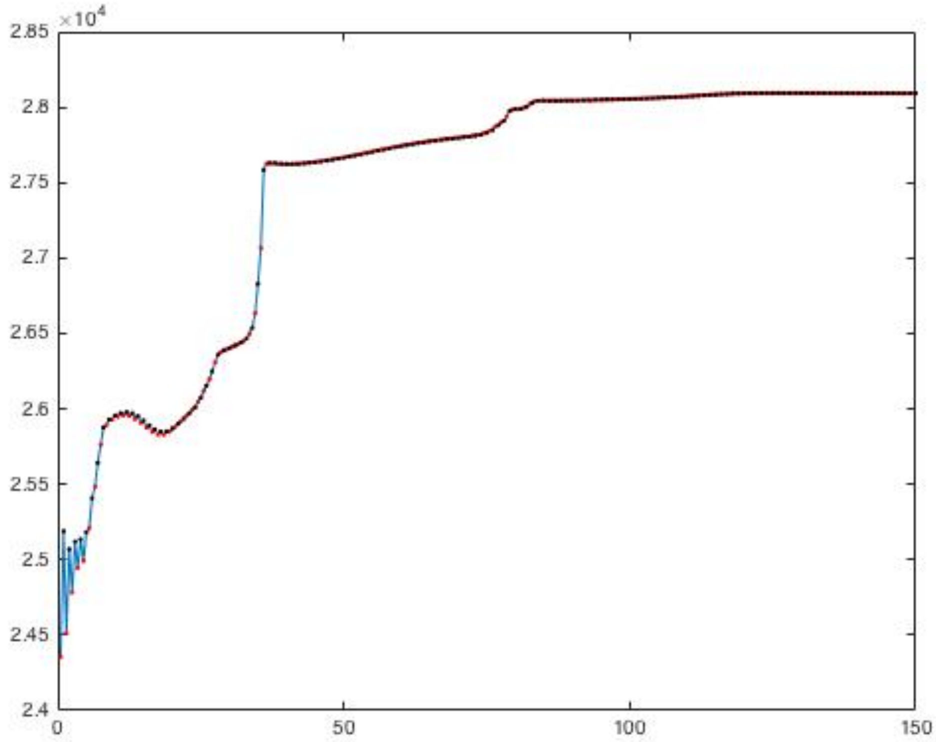
(b)
k=4,


×10^4
2.3
2.25
2.2
2.15
2.1
2.05
2
1.95
1.9
1.85
1.8
0 50 100 150

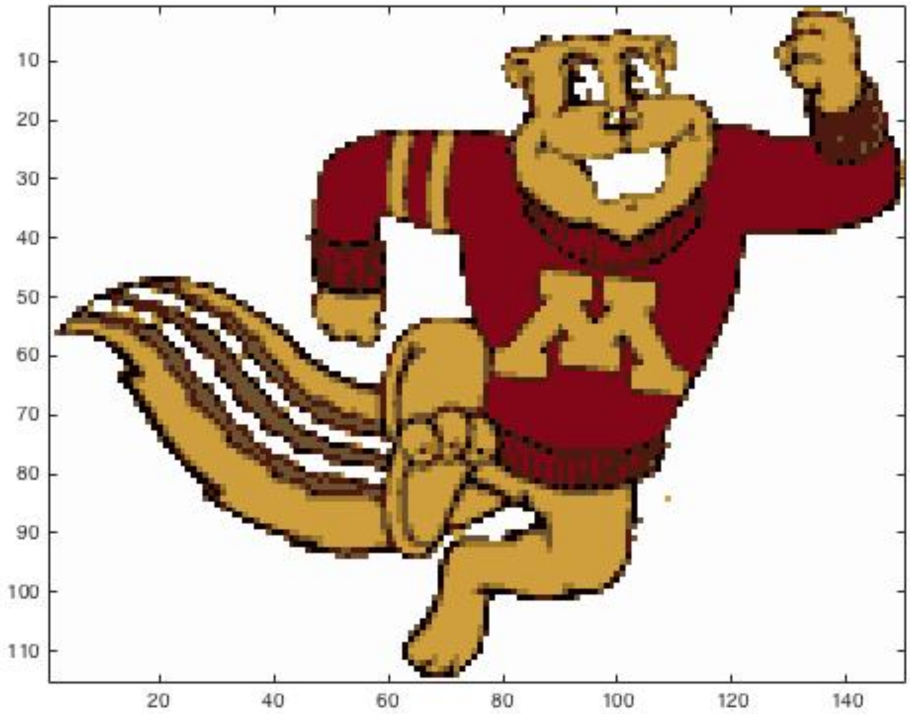
k=8,


×10^4
2.7
2.68
2.66
2.64
2.62
2.6
2.58
2.56
2.54
2.52
0
50
100
150

k=12,


×10^4
2.85
2.8
2.75
2.7
2.65
2.6
2.55
2.5
2.45
2.4
0
50
100
150

(c) Impressed image by k Means



My EM implementation falls, because the Sigma is not revertible and we can no longer use probability density function to calculate the p(i), Q and new Sigma. However, k Means does not need to use the probability density function, but only uses Euclidean distance.