Cameron Wonchoba
wonch002
5239321

Problem1: Below are my solutions to problem one.

Cameron Wonchoba
Wonch002
5239321

1. We note from the problem description...

$$P(x|C_i) = P(x|P_{i1}, P_{i2}, \dots P_{in})$$

Prior/weigh → sum of x ← mixture

$$P(x) = \sum_{i=1}^{k} P(x|C_i)P(C_i) = \frac{\pi_i \, m!}{x_1! \dots x_n!} P_{i1}^{x_1} \cdots P_{in}^{x_n}$$

↑ data ↑ Parameters for distribution

Note:

$$\sum_{j=1}^{n} x_j = m \qquad \sum_{i=1}^{k} \pi_i = 1$$

E-STEP

$$Q(\phi,\phi^l) = E[\underbrace{\mathcal{L}(\theta|X,Z)}_{\text{likelihood}} \,|\, \underset{\text{data}}{X}, \underset{\text{Parameters}}{\theta^l}]$$ Note: $\theta = \{\pi_i, P_{i1} \dots P_{in}\}$

$$= \sum_t \sum_i E[z_i^t|X,\theta^l]\left[\log \pi_i + \log\left[\frac{m!}{x_1! \dots x_n!} P_{i1}^{x_1} \cdots P_{in}^{x_n}\right]\right]$$

$$E[z_i^t|X,\theta^l] = p(z_i^t|x^t,\theta^l)$$

$$= \boxed{\frac{P(x^t,\theta^l|z_i^t)\,\overset{\pi_i}{P(z_i^t)}}{\sum_{j=1} P(x^t,\theta^l|z_j^t)\,P(z_j^t)}}$$

← z_i^t = our latent variable or label..

π_i = our distribution "weight"

→ Probability of Parameters and data given label...

Cameron Wonchoba
wonch002
5239321

1 M-Step...

Use Lagrange... ~~multiplier~~

We want to solve for θ^{l+1}...

$$\theta^{l+1} = \underset{\theta}{\text{argmax}}\, Q(\theta|\theta^l)$$

$$\hookrightarrow \sum_t \sum_i h_i^t [\log \pi_i + \log p_i(x^t|\theta^t)]$$

$$= \sum_t \sum_i h_i^t \log \pi_i + \sum_t \sum_i h_i^t \log p_i(x^t|\theta^t)$$

$$= \sum_t \sum_i h_i^t \log \pi_i + \sum_t \sum_i h_i^t \log \left[\frac{m!}{\prod_i x_i!} \prod_j^n p_{ij}^{x_j} \right]$$

Now, we have constraints...

$$\sum_i \pi_i = 1, \quad \sum_{j=1}^n p_{ij} = 1, \quad \sum_{j=1}^n x_j = m$$

Now lagrange...

$$\nabla_{\pi_i} \sum_t \sum_i h_i^t \log \pi_i - \lambda\left(\sum_i \pi_i - 1\right) = 0$$

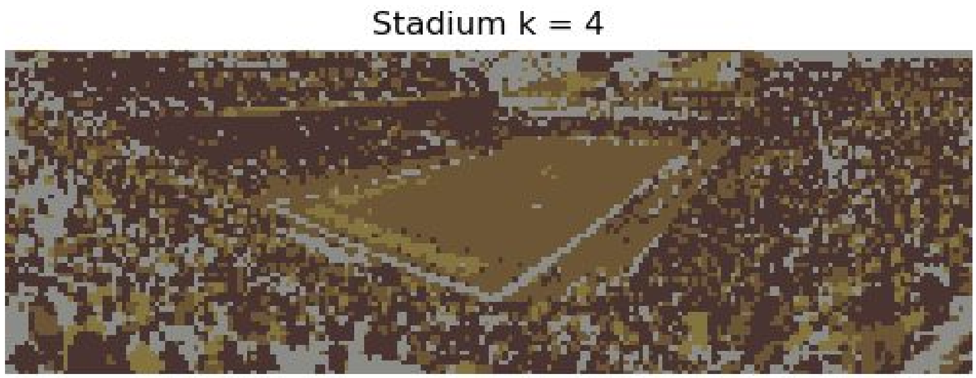
$$\nabla_{p_{ij}} \sum_t \sum_i h_i^t \log\left[\frac{m!}{\prod_i x_i!} \prod_j^n p_{ij}^{x_j}\right] - \lambda\left(\sum_{j=1}^n p_{ij} - 1\right) = 0$$

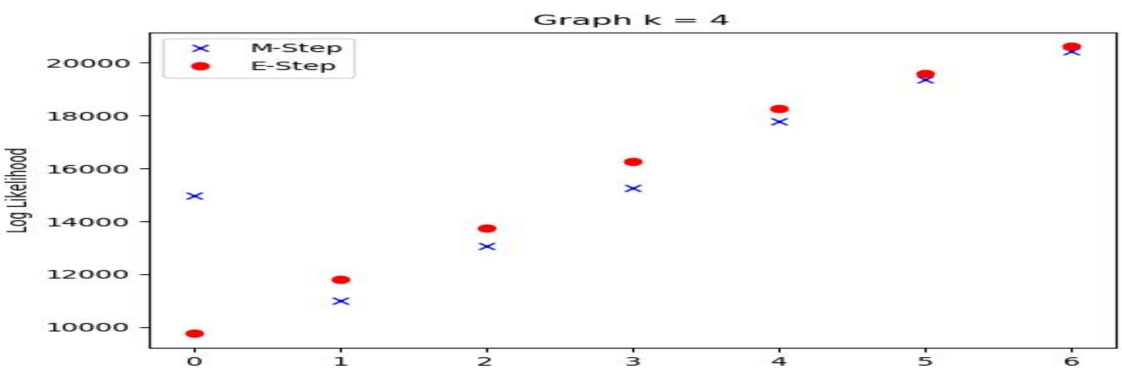
Solve for π_i & p_{ij}

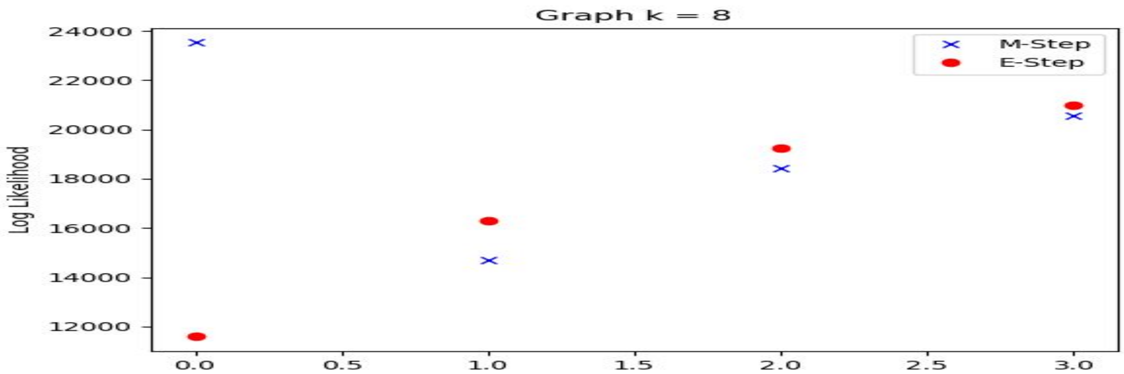
Note:
I've never learned about lagrange... I will need to learn how to do this, but for now I am out of time.

Problem2:

a/b) Below are the pictures generated from the EMG function. Note both the E-Step and M-Step are shown. The complete log likelihood function changes less each after each iteration because we are finding the optimized parameters. This results in converging values.

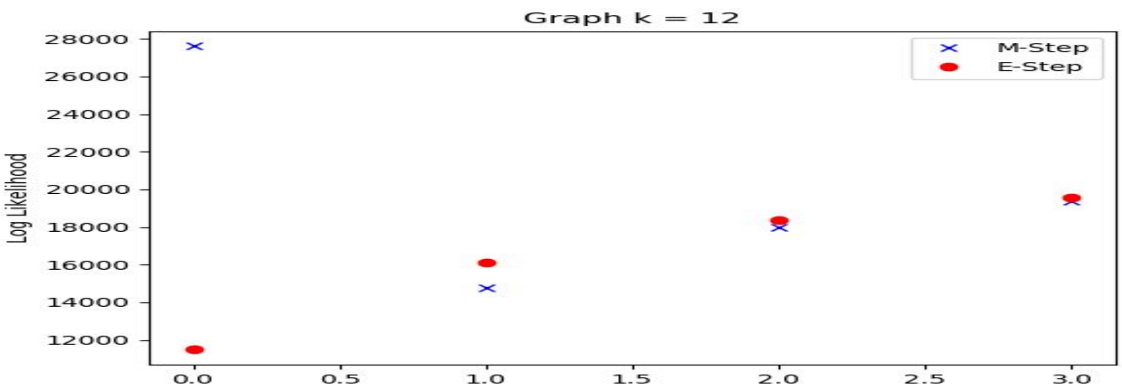





Graph k = 8
M-Step
E-Step
Log Likelihood
24000
22000
20000
18000
16000
14000
12000
0.0
0.5
1.0
1.5
2.0
2.5
3.0

Stadium k = 8

Stadium k = 12



c) KMeans and EM behave differently because KMeans focuses on minimizing the reconstruction error while EM focuses on maximizing the expected likelihood of its parameters occurring with each set of data. This results in a difference in the goal of each of the algorithms, and thus allow the algorithm to choose different values for their means to satisfy their condition.

Note: When implementing EMG on Goldy without regularization, we get a singular matrix which causes a failure. I catch this error and print the following error message:

- "Error: Singular Matrix detected. Trying again with regularization term."

I then retry the EMG algorithm with the flag set to True (as question "e" suggests). More on the results of that in section "e".

d) I solved the problem (seen below). I then added an arbitrary and small regularization term (0.01) to the covariance calculation in the function called m_step(). This regularization prevents a singular Matrix from occurring. Note: A singular matrix is not good because the multivariate gaussian distribution relies on the inverse of the covariance matrix. When a matrix is singular, the inverse does not exist (because the determinant is 0). The written solution is as follows:

Cameron Wonchoba
wonch002
5239321

Problem 4.

$$\frac{d \displaystyle\sum_{t=1}^{N} -\frac{d}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(x^t-\mu)^T \Sigma^{-1}(x^t-\mu) + \frac{-\lambda}{2}\sum_{i=1}^{K}\sum_{j=1}^{d}(\Sigma_i^{-1})_{jj}}{d\,\Sigma_i^{-1}}$$

$$0 \equiv \frac{N}{2}\Sigma - \frac{1}{2}\sum_{t=1}^{N}(x^t-\mu)(x^t-\mu)^T + \frac{-\lambda I}{2}$$

Solve for Σ

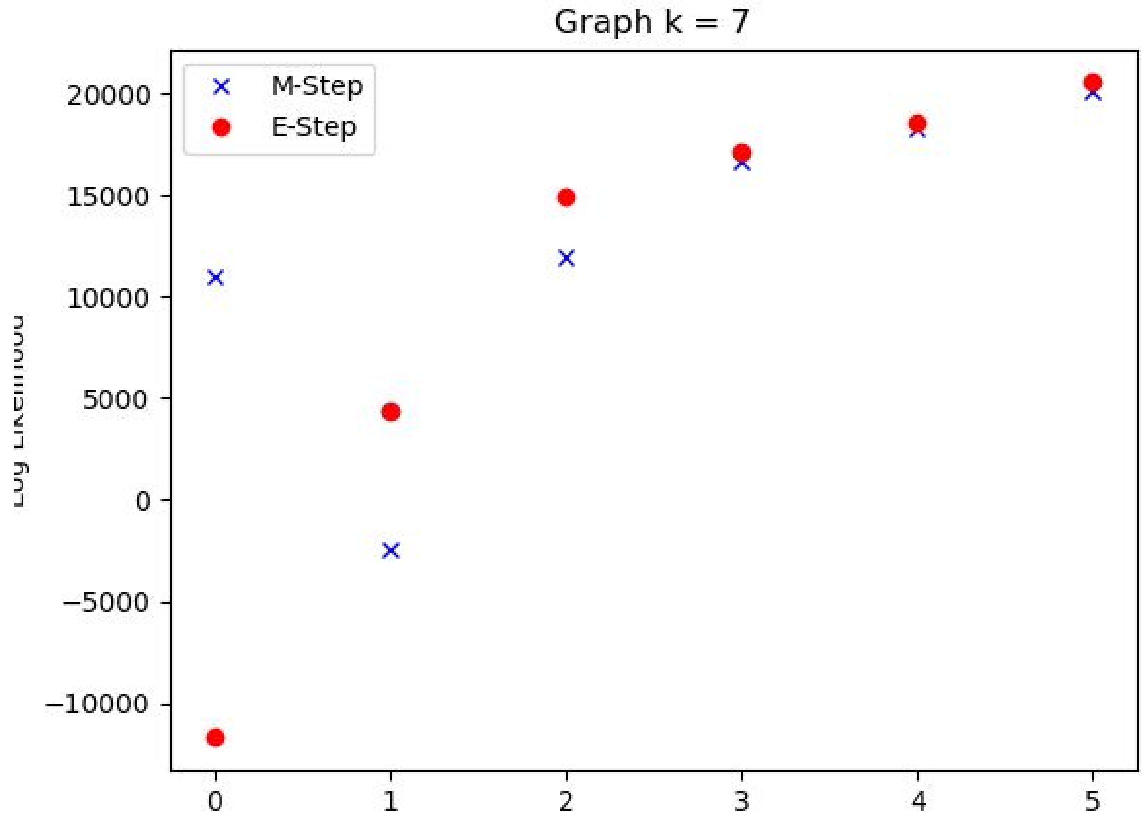
$$\boxed{\Sigma = \frac{\displaystyle\sum_{t=1}^{N}(x^t-\mu)(x^t-\mu)^T}{N} + \lambda I}$$

Just add a random term λ to covariance...

Note $\lambda > 0$.

e) The new model was implemented. My Goldy picture is slightly odd because the covariance matrix is being tinkered with (regularization). Consequently, the means must offset this tinkering to ensure the parameters are optimized for a given distribution. This means that the colors are slightly offset from their true values. Below are the results from EMG and KMeans for Goldy:

Stadium k = 7


Graph k = 7
M-Step
E-Step
Log Likelihood
20000
15000
10000
5000
0
−5000
−10000
0
1
2
3
4
5