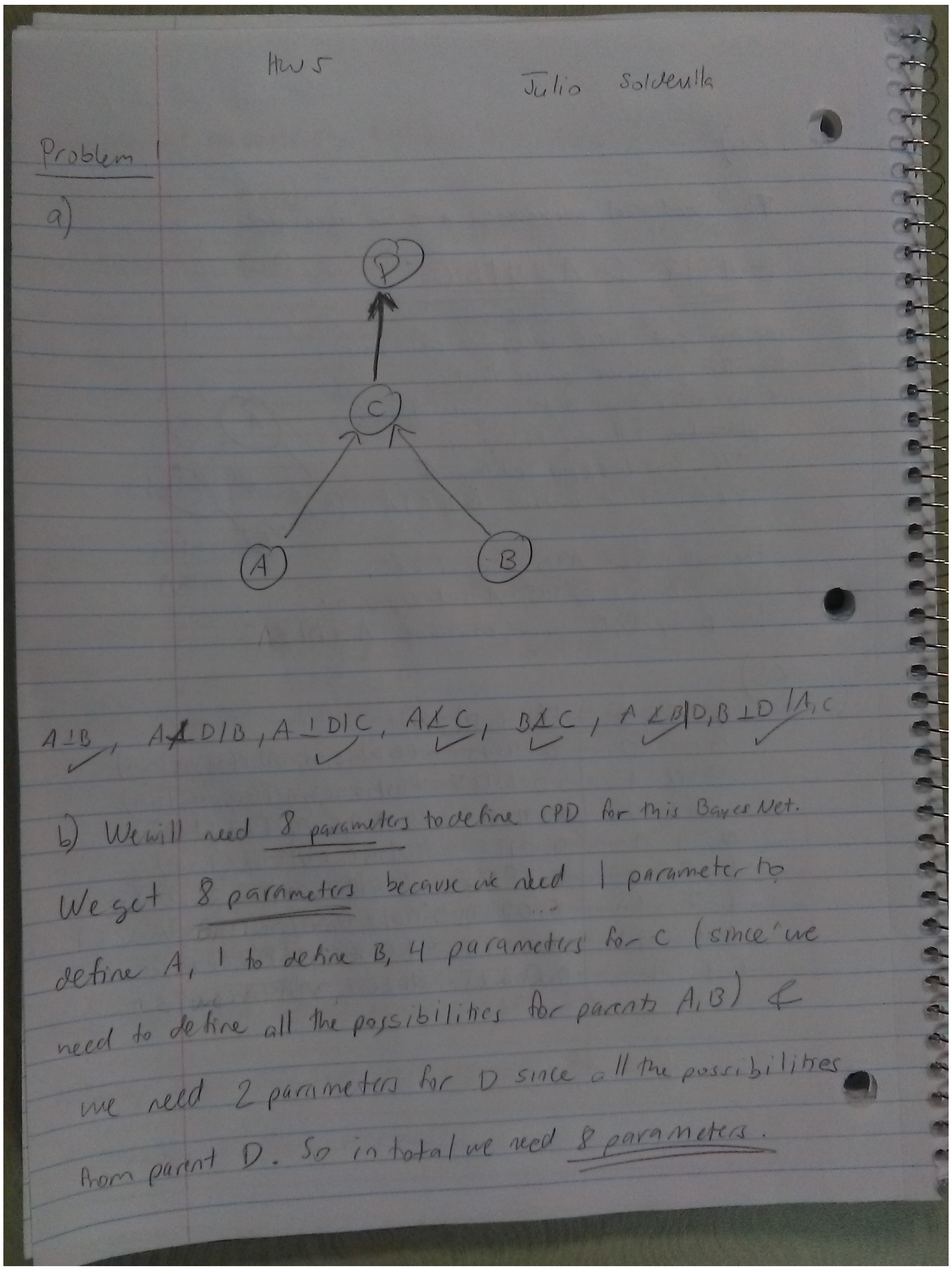
Julio Soldevilla
EECS 545 Winter 2018 — Problem Set 5

Problem 1 Problem 1


HW 5 Julio Soldevilla

Problem 1

a)

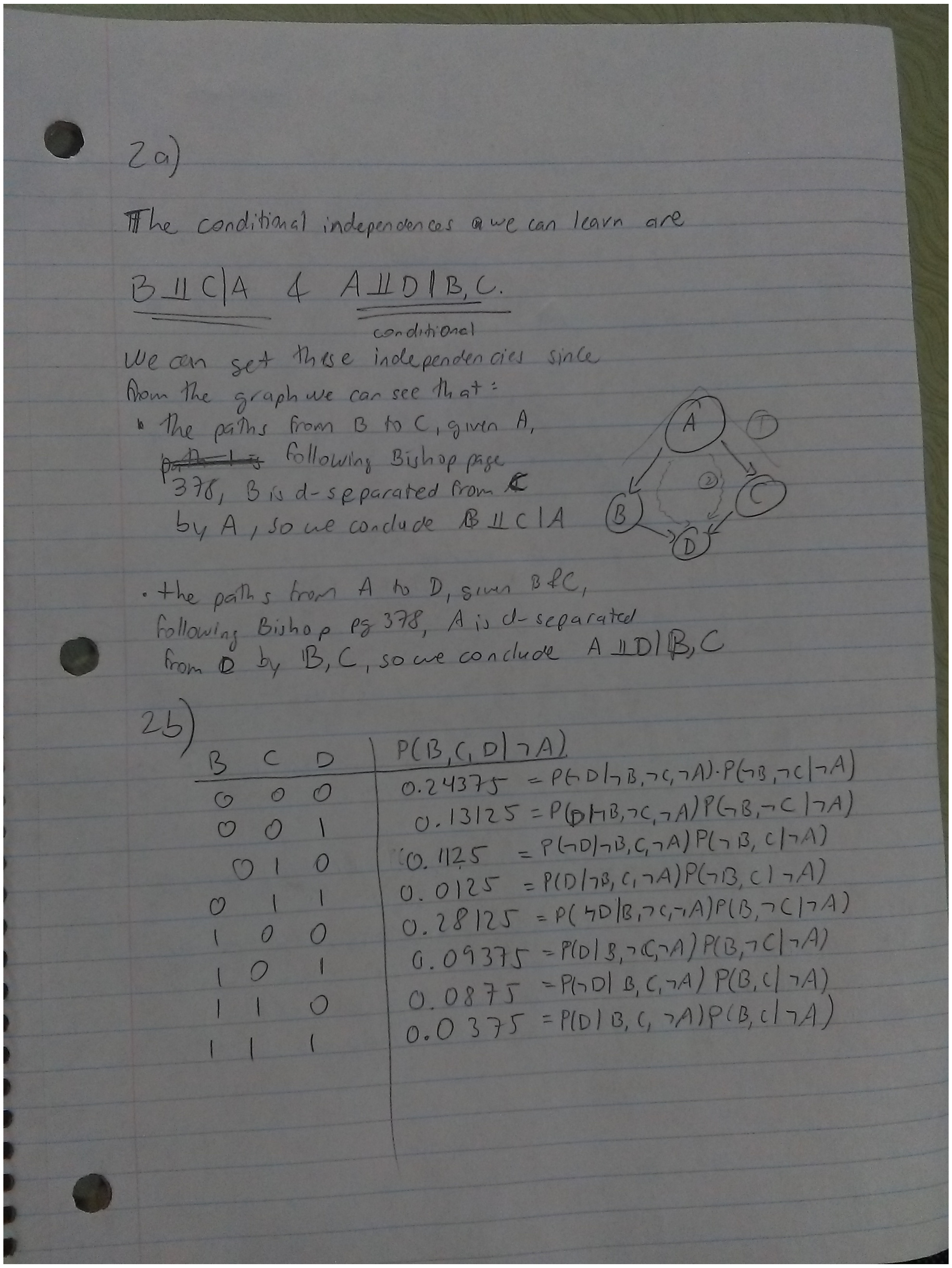
$A \perp B$, $A \not\perp D \mid B$, $A \perp D \mid C$, $A \not\perp C$, $B \not\perp C$, $A \not\perp B \mid D$, $B \perp D \mid A, C$

b) We will need 8 parameters to define CPD for this Bayes Net. We get 8 parameters because we need 1 parameter to define A, 1 to define B, 4 parameters for C (since we need to define all the possibilities for parents A, B) & we need 2 parameters for D since all the possibilities from parent D. So in total we need 8 parameters.

Figure 1: **Problem 1 part a and b:** Image showing the work for part a and b of problem 1

Proof: ■

Problem 2 Problem 2



2a)

~~T~~The conditional independences we can learn are

$B \perp\!\!\!\perp C \mid A$ & $A \perp\!\!\!\perp D \mid B, C$.

conditional

We can get these independencies since from the graph we can see that:

- The paths from B to C, given A, ~~path~~ following Bishop page 378, B is d-separated from C by A, so we conclude $B \perp\!\!\!\perp C \mid A$

- the paths from A to D, given B & C, following Bishop pg 378, A is d-separated from D by B, C, so we conclude $A \perp\!\!\!\perp D \mid B, C$

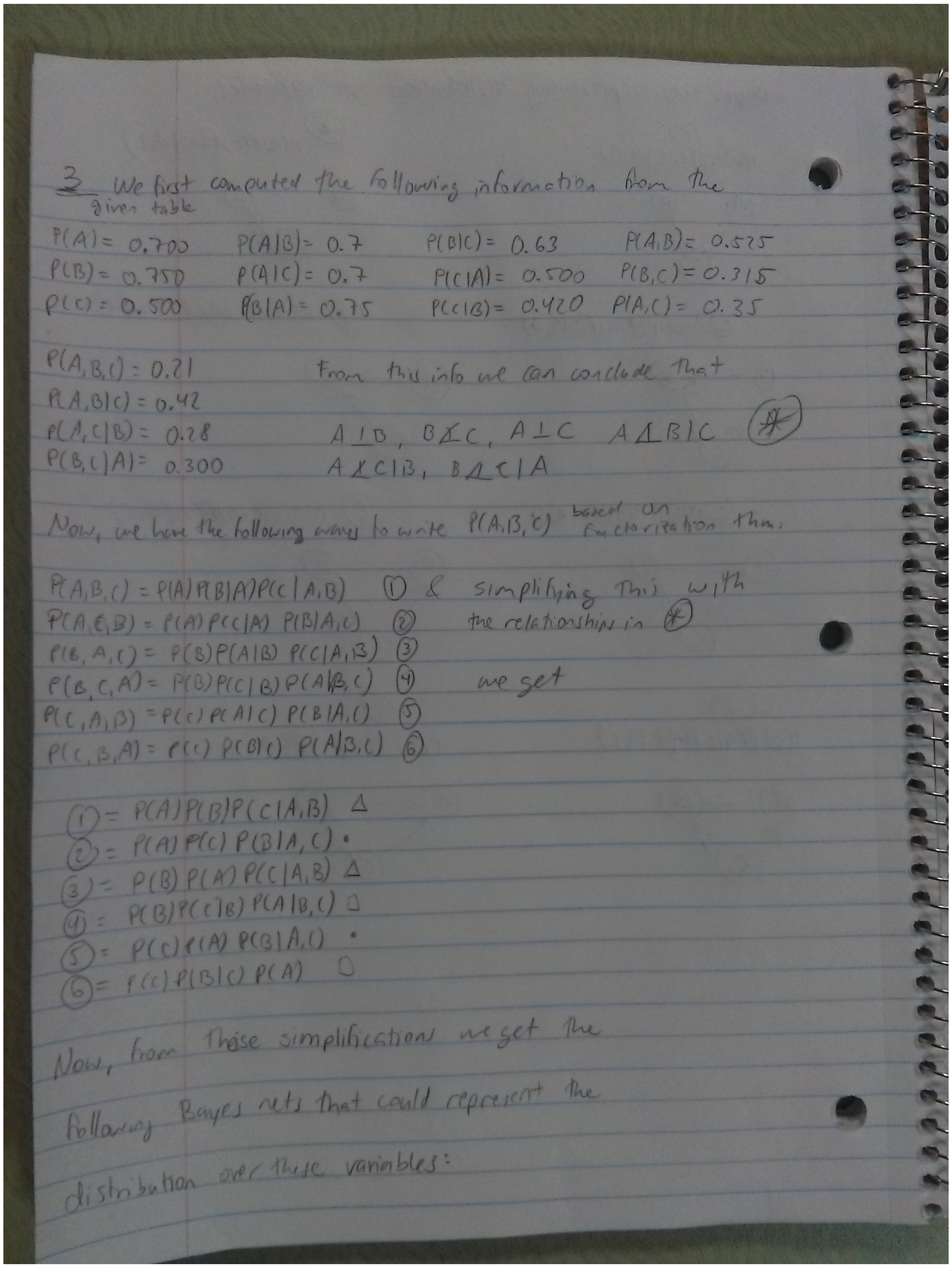
2b)

B	C	D	$P(B, C, D \mid \neg A)$
0	0	0	$0.24375 = P(\neg D \mid \neg B, \neg C, \neg A) \cdot P(\neg B, \neg C \mid \neg A)$
0	0	1	$0.13125 = P(D \mid \neg B, \neg C, \neg A) P(\neg B, \neg C \mid \neg A)$
0	1	0	$0.1125 = P(\neg D \mid \neg B, C, \neg A) P(\neg B, C \mid \neg A)$
0	1	1	$0.0125 = P(D \mid \neg B, C, \neg A) P(\neg B, C \mid \neg A)$
1	0	0	$0.28125 = P(\neg D \mid B, \neg C, \neg A) P(B, \neg C \mid \neg A)$
1	0	1	$0.09375 = P(D \mid B, \neg C, \neg A) P(B, \neg C \mid \neg A)$
1	1	0	$0.0875 = P(\neg D \mid B, C, \neg A) P(B, C \mid \neg A)$
1	1	1	$0.0375 = P(D \mid B, C, \neg A) P(B, C \mid \neg A)$

Figure 2: **Problem 2 part a and b:** Image showing the work for part a and b of problem 2

Proof: ■

Problem 3 Problem 3



3 We first computed the following information from the given table

$P(A) = 0.700$ $\quad P(A|B) = 0.7$ $\quad P(B|C) = 0.63$ $\quad P(A,B) = 0.525$

$P(B) = 0.750$ $\quad P(A|C) = 0.7$ $\quad P(C|A) = 0.500$ $\quad P(B,C) = 0.315$

$P(C) = 0.500$ $\quad P(B|A) = 0.75$ $\quad P(C|B) = 0.420$ $\quad P(A,C) = 0.35$

$P(A,B,C) = 0.21$

$P(A,B|C) = 0.42$

$P(A,C|B) = 0.28$

$P(B,C|A) = 0.300$

From this info we can conclude that

$A \perp B$, $B \not\perp C$, $A \perp C$ $\quad A \not\perp B | C$ $\quad$ (*)

$A \not\perp C | B$, $B \not\perp C | A$

Now, we have the following ways to write $P(A,B,C)$ based on factorization thm:

$P(A,B,C) = P(A)P(B|A)P(C|A,B)$ ① & simplifying this with

$P(A,C,B) = P(A)P(C|A)\,P(B|A,C)$ ② the relationships in (*)

$P(B,A,C) = P(B)P(A|B)\,P(C|A,B)$ ③

$P(B,C,A) = P(B)P(C|B)P(A|B,C)$ ④ we get

$P(C,A,B) = P(C)P(A|C)\,P(B|A,C)$ ⑤

$P(C,B,A) = P(C)\,P(B|C)\,P(A|B,C)$ ⑥

① $= P(A)P(B)P(C|A,B)$ Δ

② $= P(A)P(C)\,P(B|A,C)$ $\bullet$

③ $= P(B)\,P(A)\,P(C|A,B)$ Δ

④ $= P(B)P(C|B)\,P(A|B,C)$ $\square$

⑤ $= P(C)P(A)\,P(B|A,C)$ $\bullet$

⑥ $= P(C)P(B|C)\,P(A)$ $\square$

Now, from these simplifications we get the following Bayes nets that could represent the distribution over these variables:

Figure 3: **Problem 3** Image showing the work for problem 3

Proof: ■

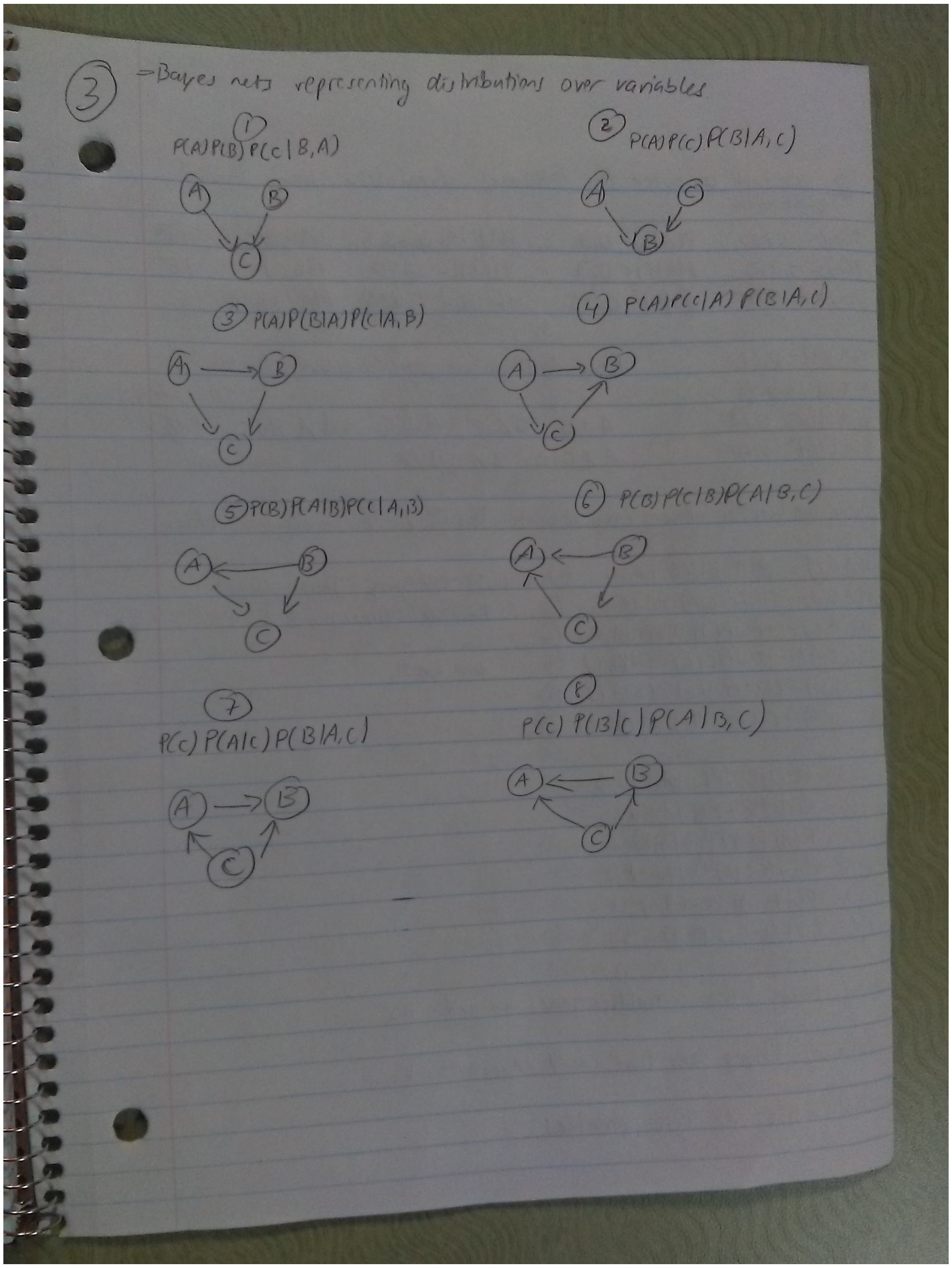


Figure 4: **Problem 3** Image showing the work for problem 3

Problem 4 Problem 4

Proof:

1. For the first part of this problem, we see, in the following graph, the picture of the scatter plot of the data.

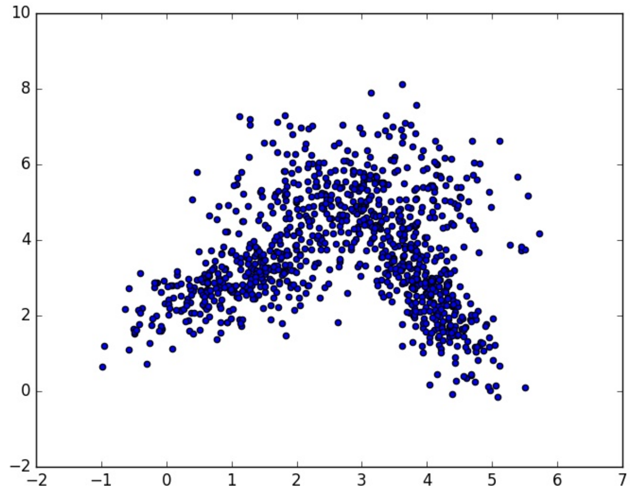


Figure 5: **Problem 4a** Image showing the scatter plot for problem 4

Now, when we have $K = 2$, we get the following Gaussian components in the mixture: with mean $\mu = \left\{ \begin{bmatrix} 1.49386365 \\ 3.60593399 \end{bmatrix} \begin{bmatrix} 3.74590611 \\ 3.68496211 \end{bmatrix} \right\}$ and the covariance matrices $\Sigma = \left\{ \begin{bmatrix} 0.98615571 & 0.90483046 \\ 0.90483046 & 1.61666384 \end{bmatrix}, \begin{bmatrix} 0.56267209 & -0.71849557 \\ -0.71849557 & 2.81343664 \end{bmatrix} \right\}$ and we get the following scatter plot with the clusters.

Now, when we have $K = 3$, we get the following Gaussian components in the mixture: with mean $\mu = \left\{ \begin{bmatrix} 1.08271327 \\ 2.8828234 \end{bmatrix} \begin{bmatrix} 2.97193003 \\ 5.04472742 \end{bmatrix} \begin{bmatrix} 4.04333322 \\ 2.50060083 \end{bmatrix} \right\}$ and the covariance matrices $\Sigma = \left\{ \begin{bmatrix} 0.64506251 & 0.38116334 \\ 0.38116334 & 0.5568736 \end{bmatrix}, \begin{bmatrix} 1.01331798 & -0.02915437 \\ -0.02915437 & 1.01501022 \end{bmatrix}, \begin{bmatrix} 0.21979711 & -0.37026767 \\ -0.37026767 & 1.14190251 \end{bmatrix} \right\}$ and we get the following scatter plot with the clusters.

Now, when we have $K = 5$, we get the following Gaussian components in the mixture: with mean $\mu = \left\{ \begin{bmatrix} 1.87818858 \\ 2.68555211 \end{bmatrix} \begin{bmatrix} 2.98754896 \\ 5.0617341 \end{bmatrix} \begin{bmatrix} 4.07182365 \\ 2.21736995 \end{bmatrix}, \begin{bmatrix} 3.95411976 \\ 3.045238 \end{bmatrix}, \begin{bmatrix} 1.01229367 \\ 2.90454076 \end{bmatrix} \right\}$ and the covariance matrices $\Sigma =$

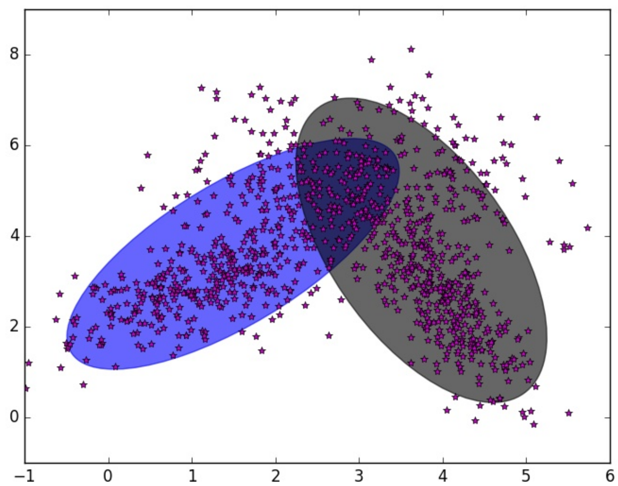


Figure 6: **Problem 4a** Image showing the scatter plot and result of EM algorithm for 2 clusters.

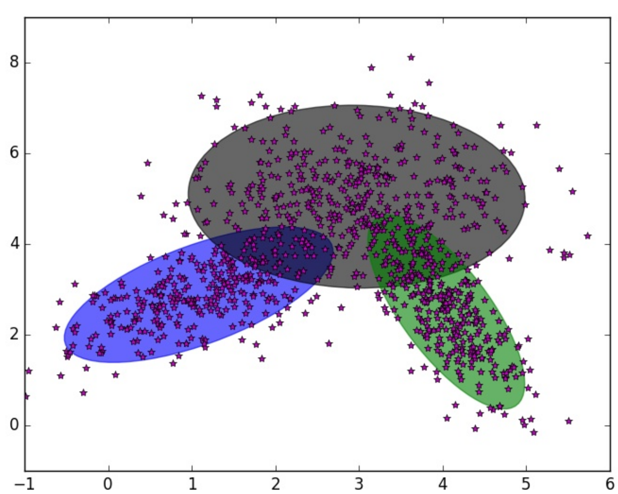


Figure 7: **Problem 4a** Image showing the scatter plot and result of EM algorithm for 3 clusters.

$\left\{ \begin{bmatrix} 0.12519243 & -0.00347678 \\ -0.00347678 & 0.30986251 \end{bmatrix}, \begin{bmatrix} 1.01766795 & -0.04941322 \\ -0.04941322 & 0.98578055 \end{bmatrix}, \begin{bmatrix} 0.2104126 & -0.32793317 \\ -0.32793317 & 1.00038797 \end{bmatrix}, \right\}$
$\left\{ \begin{bmatrix} 0.26025229 & -0.45255172 \\ -0.45255172 & 1.03764694 \end{bmatrix}, \begin{bmatrix} 0.61327566 & 0.42382975 \\ 0.42382975 & 0.58797768 \end{bmatrix} \right\}$ and we get the following scatter plot with the clusters.

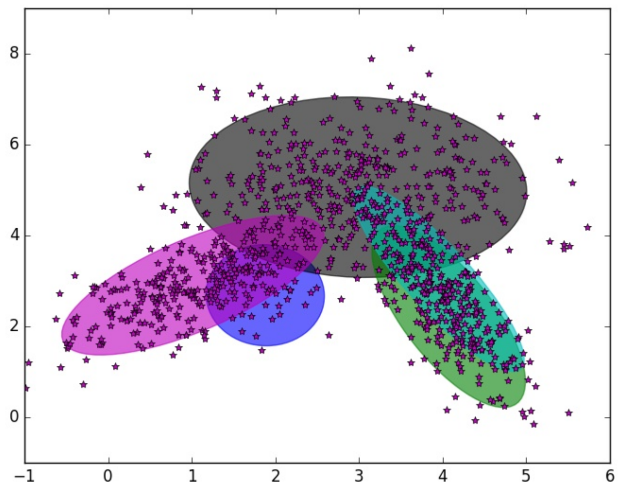


Figure 8: **Problem 4a** Image showing the scatter plot and result of EM algorithm for 5 clusters.

Now, when we have $K = 10$, we get the following Gaussian components in the mixture: with mean $\mu =$
$\left\{ \begin{bmatrix} 3.6025199 \\ 4.83936943 \end{bmatrix} \begin{bmatrix} 4.71296034 \\ 0.1986524 \end{bmatrix} \begin{bmatrix} 4.356082 \\ 1.6782708 \end{bmatrix}, \begin{bmatrix} 4.12471251 \\ 2.85796 \end{bmatrix}, \begin{bmatrix} 2.14465053 \\ 5.08194636 \end{bmatrix}, \begin{bmatrix} 3.85398589 \\ 2.10096814 \end{bmatrix}, \right\}$
$\left\{ \begin{bmatrix} 3.75209835 \\ 3.20487639 \end{bmatrix}, \begin{bmatrix} 1.30307196 \\ 3.29183234 \end{bmatrix}, \begin{bmatrix} 1.14010914 \\ 2.78515828 \end{bmatrix}, \begin{bmatrix} 3.4274561 \\ 5.20456023 \end{bmatrix} \right\}$ and the covariance matrices
$\Sigma =$
$\left\{ \begin{bmatrix} 3.75429517e-01 & 4.41361567e-01 \\ 4.41361567e-01 & 5.19969327e-01 \end{bmatrix}, \begin{bmatrix} 1.48798492e-01 & -2.85454354e-02 \\ -2.85454354e-02 & 3.49132323e-02 \end{bmatrix} \right\}$
$\left\{ \begin{bmatrix} 1.22727443e-01 & -6.60134578e-02 \\ -6.60134578e-02 & 2.50033836e-01 \end{bmatrix}, \begin{bmatrix} 1.99033468e-02 & 2.84515021e-02 \\ 2.84515021e-02 & 6.58684882e-02 \end{bmatrix}, \right\}$
$\left\{ \begin{bmatrix} 6.25215337e-01 & 2.66846671e-01 \\ 2.66846671e-01 & 1.14271145e+00 \end{bmatrix}, \begin{bmatrix} 1.35135357e-02 & 1.58503668e-03 \\ 1.58503668e-03 & 2.50685466e-04 \end{bmatrix}, \right\}$
$\left\{ \begin{bmatrix} 2.08282963e-01 & -3.02666524e-01 \\ -3.02666524e-01 & 8.50492621e-01 \end{bmatrix}, \begin{bmatrix} 1.00560907e+00 & 9.90227009e-01 \\ 9.90227009e-01 & 1.14660462e+00 \end{bmatrix}, \right\}$
$\left\{ \begin{bmatrix} 6.58828033e-01 & 2.24299843e-01 \\ 2.24299843e-01 & 3.91589498e-01 \end{bmatrix}, \begin{bmatrix} 8.85017840e-01 & -1.96596216e-01 \\ -1.96596216e-01 & 8.51715398e-01 \end{bmatrix} \right\}$
and we get the following scatter plot with the clusters.

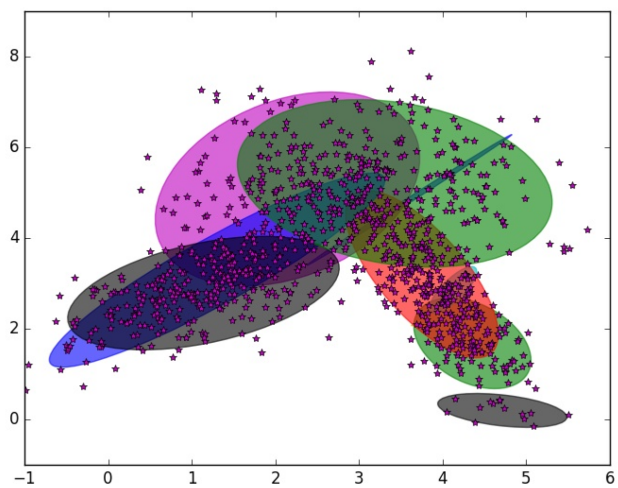


Figure 9: **Problem 4a** Image showing the scatter plot and result of EM algorithm for 10 clusters.

2. Now, in this problem, we perform the EM algorithm and plot the results after 1,5, 10, 20 and 50 steps. We get the following graphs:

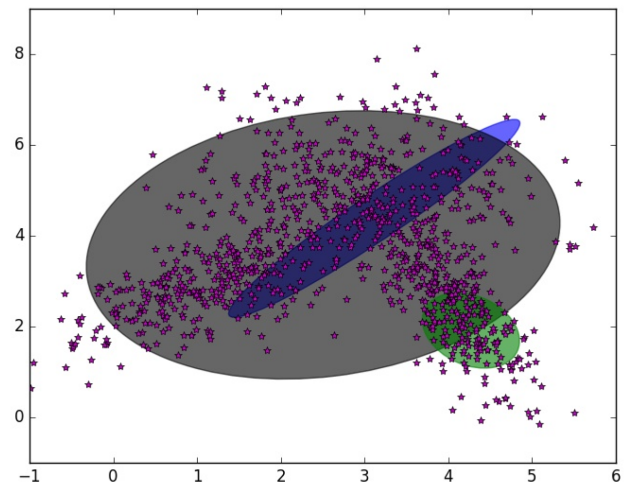


Figure 10: **Problem 4b** Image showing the EM algorithm result after 1 step of the algorithm

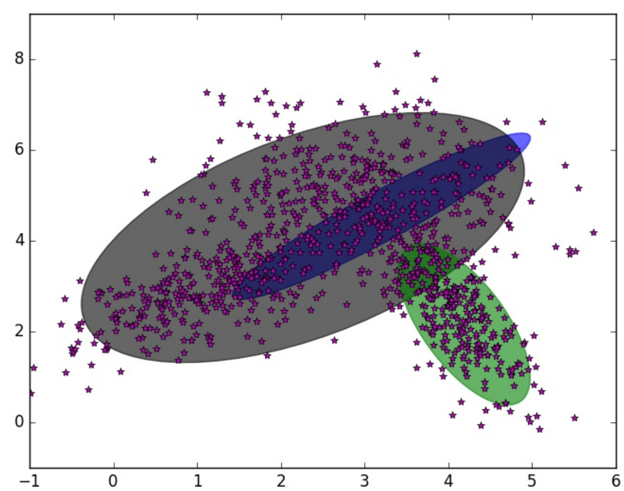


Figure 11: **Problem 4b** Image showing the EM algorithm result after 5 step of the algorithm

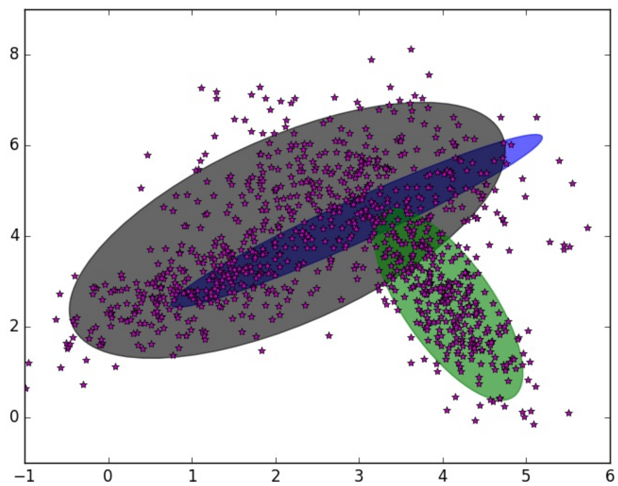


Figure 12: **Problem 4b** Image showing the EM algorithm result after 10 step of the algorithm

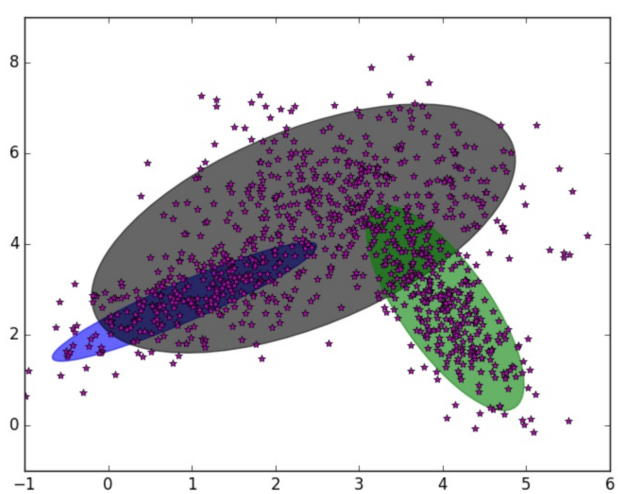


Figure 13: **Problem 4b** Image showing the EM algorithm result after 20 step of the algorithm

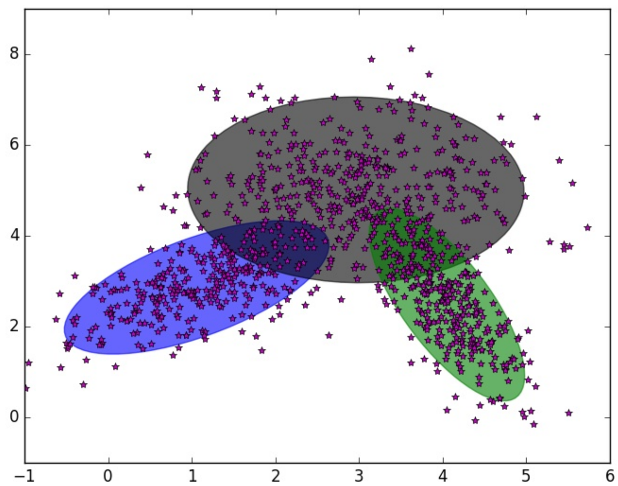


Figure 14: **Problem 4b** Image showing the EM algorithm result after 50 step of the algorithm

3. Finally, after playing around with the initialization parameters (particularly) changing the mean, we end up having that

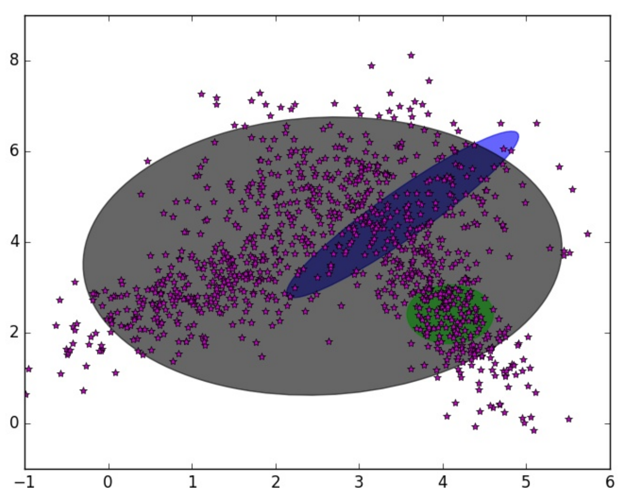


Figure 15: **Problem 4c** Image showing the EM algorithm result after 1 step of the algorithm

From these pictures, we see that after changing the initial mean of the EM algorithm, the algorithm learned different clusters that don't quite align with the given data, we are finding clusters centered at different points of the learning space. Also, (not presented in pictures) if we change the initial covariance matrices drastically, the algorithm would learn perhaps thinner or much fatter regions, corresponding to the change in initial covariance.

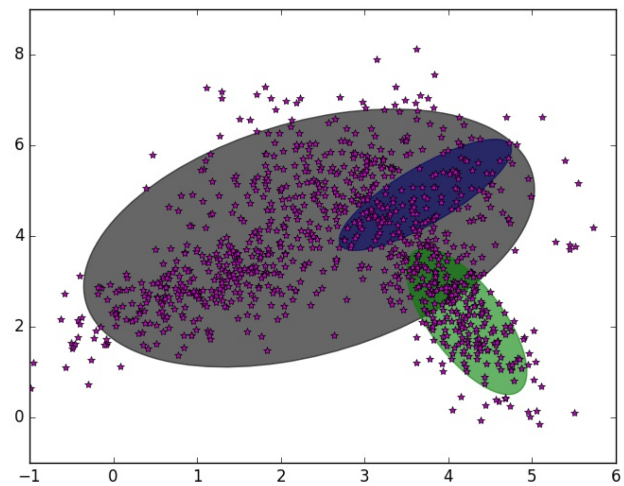


Figure 16: **Problem 4c** Image showing the EM algorithm result after 5 step of the algorithm

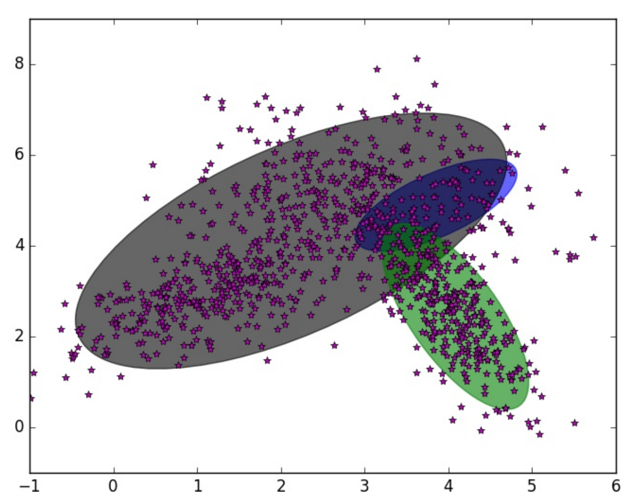


Figure 17: **Problem 4c** Image showing the EM algorithm result after 10 step of the algorithm

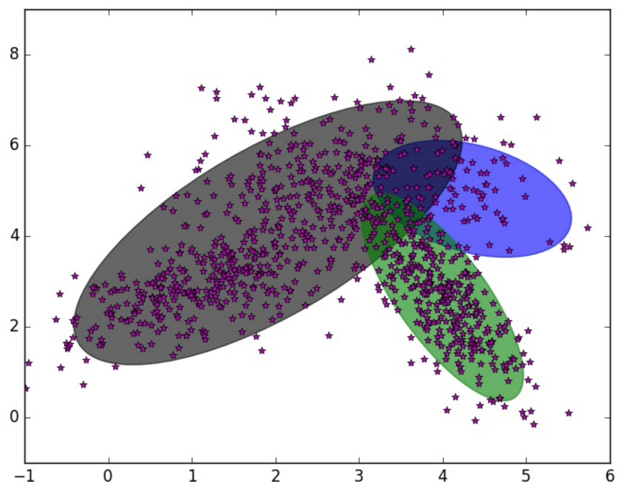


Figure 18: **Problem 4c** Image showing the EM algorithm result after 20 step of the algorithm

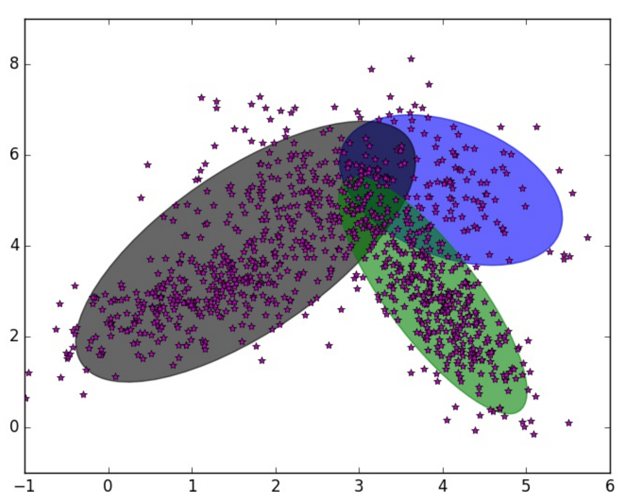


Figure 19: **Problem 4c** Image showing the EM algorithm result after 50 step of the algorithm