Homework 2
ECE/CS 498 DS Spring 2020
Issued: 02/24/20
Due: 03/02/20

Name: Chuhao Feng
NetID: chuhaof2
Registration status: Registered

Note: Only Problem 1,3 will be graded. Problem 2 is optional.

You can either type your solutions or write your solutions by hand. You should make sure your manuscript is legible if you choose to write your solutions.

Problem 1

Task 1 K-Means:

Suppose there are 6 data points in 2D space. The coordinates of these data points are given in the following table:

$\vec{x}_1$	$\vec{x}_2$	$\vec{x}_3$	$\vec{x}_4$	$\vec{x}_5$	$\vec{x}_6$
(1,1)	(1,2)	(2,2)	(6,6)	(6,7)	(7,7)

Our task is to cluster them using K-means algorithm. **We set K=2 and use Euclidean distance**. In the beginning, the centroids are initialized to (3,2), (5,7).

1. Fill in the tables below to complete the first iteration:
   Step 1: Assign data points to the nearest centroid. Fill each data point in one of the columns below.

	Centroid (3,2)	Centroid (5,7)
Assigned Data points	(1,1), (1,2), (2,2)	(6,6), (6,7), (7,7)

   Step 2: Update centroid. Calculate new centroids base on the data assignment in Step 1.

New Centroid I	New Centroid II
$(\frac{4}{3}, \frac{5}{3})$	$(\frac{19}{3}, \frac{20}{3})$

2. In addition, state at least two convergence criteria for K-Means, choose one criterion, and judge whether the K-Means algorithm has converged after the above step.

   Convergence criteria for K-Means:
   1. No (or minimal) re-assignment of data points to different clusters
   2. No (or minimal) change of centroids

   Since the centroids are changed after the above step, the K-Means algorithm has not converged after the above step, according to the second criteria listed above.

Grade:__________

Homework 2 **Name: ____________________**
ECE/CS 498 DS Spring 2020
Issued: 02/24/20 **NetID: ___________________**
Due: 03/02/20

3. Suppose you have a dataset which has some outliers that will affect the K-means' clustering result if included. To achieve a better clustering result, you must reduce the effects of these outliers on the clusters. However, suppose you don't want to remove these outliers from the dataset when doing clustering, but you only want to make one subtle change to the K-Means algorithm. **Describe how you can achieve this, and why**. Hint: the new algorithm has the name beginning with K. Hint, Hint: You may want to update centroid by using other statistics instead of mean.

I think we can use K-Medians algorithm instead. Instead of calculating the mean to update centroids, we can derive the median value to update centroids. By using median values to update centroids, we can minimize the effect of outliers on the clustering while maintaining a similar clustering result to the K-Means algorithm, because mean value and median value are close to each other and that outliers have much less effect on median values than mean values.

Task 2 1-D GMM:

Consider applying EM to train a Gaussian Mixture Model (GMM) to cluster the data in the above task into two clusters. First, we want to apply 1-D GMM. Therefore, we project these data to the horizontal axis by ignoring the second dimension. This leads **to 6 points at 1,1,2,6,6,7**. The initial Gaussian Components are $a{\sim}N(3,1)$,$b{\sim}N(4,1)$.

1.For the point x=1, calculate $P(x = 1|a)$, which is the probability density of observing 1 when sampled from distribution $a{\sim}N(3,1)$. Show your work.

$\because a \sim N(3, 1)$

$\therefore \mu_a = 3,\ \sigma_a = 1$

$\because p(x = 1|a) = \frac{1}{\sqrt{2\pi\sigma_a^2}} exp(-\frac{(1-\mu_a)^2}{2\sigma_a^2})$

$\therefore p(x = 1|a) = 0.053991$

2. Calculate $P(a|x = 1)$, which is the posterior probability of distribution $\boldsymbol{a}$ given sample x=1. Show your work. Assume that prior $P(a) = P(b) = 0.5$.

$\because a \sim N(3, 1),\ b \sim N(4, 1)$

$\therefore \mu_a = 3,\ \sigma_a = 1,\ \ \mu_b = 4,\ \sigma_b = 1$

$\because P(x = 1|a) = \frac{1}{\sqrt{2\pi\sigma_a^2}} exp(-\frac{(1-\mu_a)^2}{2\sigma_a^2})$

$\because P(x = 1|b) = \frac{1}{\sqrt{2\pi\sigma_b^2}} exp(-\frac{(1-\mu_b)^2}{2\sigma_b^2})$

$\therefore P(x = 1|a) = 0.053991,\ P(x = 1|b) = 0.004432$

$\because P(a|x = 1) = \frac{P(x=1|a)P(a)}{P(x=1|a)P(a)+P(x=1|b)P(b)} = 0.924139$

3. Follow your steps in the above question, calculate posterior $P(a|x_i)$ and $P(b|x_i)$ for all the data points. You might want to write a Python program to solve this question.

data point x_i	Posterior $P(a \mid x_i)$	Posterior $P(b \mid x_i)$
1	0.9241418199787564	0.07585818002124355
1	0.9241418199787564	0.07585818002124355
2	0.8175744761936437	0.18242552380635635
6	0.07585818002124355	0.9241418199787564
6	0.07585818002124355	0.9241418199787564
7	0.02931223075135632	0.9706877692486436

Grade:__________

4. Based on your results in the previous questions, calculate new means and variances for the two new Gaussian Components.

$$\therefore \mu_a = \frac{a_1x_1+a_2x_2+a_3x_3+a_4x_4+a_5x_5+a_6x_6}{a_1+a_2+a_3+a_4+a_5+a_6} = 1.615419523594012$$

$$\therefore \mu_b = \frac{b_1x_1+b_2x_2+b_3x_3+b_4x_4+b_5x_5+b_6x_6}{b_1+b_2+b_3+b_4+b_5+b_6} = 5.8358460105669945$$

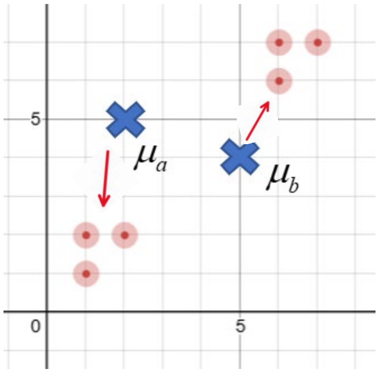
$$\therefore \sigma_a^2 = \frac{a_1(x_1-\mu_a)^2+a_2(x_2-\mu_a)^2+a_3(x_3-\mu_a)^2+a_4(x_4-\mu_a)^2+a_5(x_5-\mu_a)^2+a_6(x_6-\mu_a)^2}{a_1+a_2+a_3+a_4+a_5+a_6} = 1.6114060717172303$$

$$\therefore \sigma_b^2 = \frac{b_1(x_1-\mu_b)^2+b_2(x_2-\mu_b)^2+b_3(x_3-\mu_b)^2+b_4(x_4-\mu_b)^2+b_5(x_5-\mu_b)^2+b_6(x_6-\mu_b)^2}{b_1+b_2+b_3+b_4+b_5+b_6} = 2.409504187152901$$

Task 3 2-D GMM:

Now consider clustering the original data given in Task 2 using a 2-D GMM. We assign two initial Gaussian components to be $a \sim N((2,5), I)$, $b \sim N((5,4), I)$, where I means Identity matrix, which implies that the covariance between features is 0 and the variance of each feature is 1.

1.Draw on the figure the directions in which μ_a and μ_b will move in the next iteration. You can choose to copy or redraw the figure in your solution.



2. For data point (2,2), calculate $P((2,2)|a)$, which is the probability density of observing (2,2) when sampled from distribution multivariate Gaussian distribution a. Show your work.

$$\because a \sim N((2,5), I)$$

$$\therefore \vec{\mu_a} = (2,5),\ \sigma_{a1} = 1,\ \sigma_{a2} = 1$$

$$\therefore P((2,2)|a) = \frac{1}{\sqrt{2\pi\sigma_{a1}^2}} exp(-\frac{1}{2}\frac{(2-2)^2}{\sigma_{a1}^2}) \times \frac{1}{\sqrt{2\pi\sigma_{a2}^2}} exp(-\frac{1}{2}\frac{(2-5)^2}{\sigma_{a2}^2}) = 0.001768$$

3. Calculate $P(a|(2,2))$, which is the posterior probability of distribution *a* given sample (2,2). Show your work, and assume $P(a) = P(b) = 0.5$.

$$\because a \sim N((2,5), I),\ b \sim N((5,4), I)$$

$$\therefore \vec{\mu_a} = (2,5),\ \sigma_{a1} = 1,\ \sigma_{a2} = 1$$

$$\therefore \vec{\mu_b} = (5,4),\ \sigma_{b1} = 1,\ \sigma_{b2} = 1$$

$$\therefore P((2,2)|a) = \frac{1}{\sqrt{2\pi\sigma_{a1}^2}} exp(-\frac{1}{2}\frac{(2-2)^2}{\sigma_{a1}^2}) \times \frac{1}{\sqrt{2\pi\sigma_{a2}^2}} exp(-\frac{1}{2}\frac{(2-5)^2}{\sigma_{a2}^2}) = 0.001768$$

$$\therefore P((2,2)|b) = \frac{1}{\sqrt{2\pi\sigma_{b1}^2}} exp(-\frac{1}{2}\frac{(2-5)^2}{\sigma_{b1}^2}) \times \frac{1}{\sqrt{2\pi\sigma_{b2}^2}} exp(-\frac{1}{2}\frac{(2-4)^2}{\sigma_{b2}^2}) = 0.000239$$

$$\therefore P(a|(2,2)) = \frac{P((2,2)|a)P(a)}{P((2,2)|a)P(a)+P((2,2)|b)P(b)} = 0.880797$$

4. Follow your steps in the above question, calculate posterior $P(a|x_i)$ and $P(b|x_i)$ for all the data points. You might want to write a Python program for this question.

data point x_i	Posterior $P(a\,|\,x_i)$	Posterior $P(b\,|\,x_i)$
(1,1)	0.9820137900379085	0.01798620996209156
(1,2)	0.9933071490757152	0.006692850924284857
(2,2)	0.8807970779778824	0.11920292202211756
(6,6)	0.002472623156634774	0.9975273768433652
(6,7)	0.006692850924284857	0.9933071490757152
(7,7)	0.00033535013046647805	0.9996646498695335

5. Based on your result in the previous questions, calculate new means and covariance matrices for two new Gaussian Components.

$$\because P(a) = \frac{1}{6}\sum_{i=1}^{6} P(a|\vec{x_i}),\ P(b) = \frac{1}{6}\sum_{i=1}^{6} P(b|\vec{x_i})$$

$$\because \mu_{a,1} = \sum_{i=1}^{6}(\frac{P(a|\vec{x_i})}{6P(a)})x_{i,1},\ \mu_{a,2} = \sum_{i=1}^{6}(\frac{P(a|\vec{x_i})}{6P(a)})x_{i,2}$$

$$\because \mu_{b,1} = \sum_{i=1}^{6}(\frac{P(b|\vec{x_i})}{6P(b)})x_{i,1},\ \mu_{b,2} = \sum_{i=1}^{6}(\frac{P(b|\vec{x_i})}{6P(b)})x_{i,2}$$

$$\therefore \mu_a = (\mu_{a,1}, \mu_{a,2}) = (1.324061, 1.673026)$$

$$\therefore \mu_b = (\mu_{b,1}, \mu_{b,2}) = (6.127444, 6.446486)$$

$$\because (\textstyle\sum_c)_{j,k} = \sum_{i=1}^{N}(\frac{P(c|\vec{x_i})}{N \cdot P(c)})(x_{i,j} - \mu_{c,j})(x_{i,k} - \mu_{c,k})$$

$$\because M_a = \begin{bmatrix} (\sum_a)_{1,1} & (\sum_a)_{1,2} \\ (\sum_a)_{2,1} & (\sum_a)_{2,2} \end{bmatrix},\ M_b = \begin{bmatrix} (\sum_b)_{1,1} & (\sum_b)_{1,2} \\ (\sum_b)_{2,1} & (\sum_b)_{2,2} \end{bmatrix}$$

$$\therefore M_a = \begin{bmatrix} 0.286525 & 0.185117 \\ 0.185117 & 0.310897 \end{bmatrix},\ M_b = \begin{bmatrix} 1.108027 & 1.056691 \\ 1.056691 & 1.222608 \end{bmatrix}$$

Grade:__________

Problem 2 (Optional)

Weather (sunny or rainy) and Location (town or highway) have the potential to cause disengagements of autonomous vehicles. These disengagements could lead to accidents. Given the Bayes Net in Figure 1, answer the following questions:

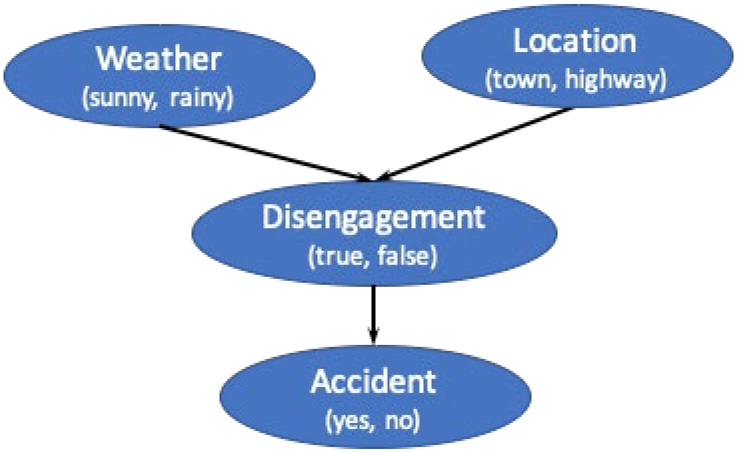


Figure 1

Weather	**Probability**
sunny	0.7
rainy	0.3
Location	**Probability**
town	0.8
highway	0.2

Disengagement Conditional Probability Table (CPT)			
Weather	**Location**	**Disengagement=true**	**Disengagement=false**
sunny	town	0.05	0.95
sunny	highway	0.01	0.99
rainy	town	0.15	0.85
rainy	highway	0.05	0.95

Accident CPT		
Disengagement	**Accident=yes**	**Accident=no**
true	0.4	0.6
false	0.01	0.99

Homework 2 **Name: ____________________**
ECE/CS 498 DS Spring 2020
Issued: 02/24/20 **NetID: ___________________**
Due: 03/02/20

A. How many parameters are needed to define the conditional probability distribution of the Bayes Net given in Figure 1.

B. Construct the **joint probability distribution** of Weather, Location, and Disengagement

Weather	**Location**	**Disengagement=true**	**Disengagement=false**
sunny	town		
sunny	highway		
rainy	town		
rainy	highway		

C. Calculate the probability of the following hypotheses
Let

a. A = Accident
b. D = Disengagement
c. W = Weather
d. L = Location

Hypothesis		**Probability**
H0	$P(W = sunny, L = town \mid A = yes)$	
H1	$P(W = sunny, L = highway \mid A = yes)$	
H2	$P(W = rainy, L = town \mid A = yes)$	
H3	$P(W = rainy, L = highway \mid A = yes)$	

D. Apply the MAP decision rule to the 4 hypotheses above.

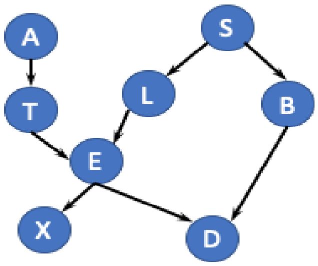
Grade:__________

Homework 2
ECE/CS 498 DS Spring 2020
Issued: 02/24/20
Due: 03/02/20

Name: ____________________

NetID: ___________________

Problem 3



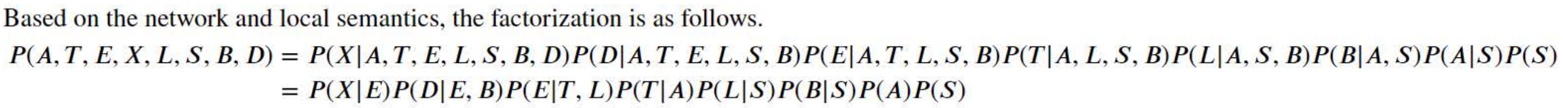
The chest clinic network above concerns the diagnosis of lung disease (tuberculosis, lung cancer, or both, or neither). In this model, a visit to Asia is assumed to increase the probability of tuberculosis. We have the following binary variables:

Variable	
X	positive X-ray
D	dyspnea (shortness of breath)
E	either tuberculosis or lung cancer
T	tuberculosis
L	lung cancer
B	bronchitis
A	a visit to Asia
S	Smoker

A. What are the differences between a joint probability and a conditional probability? Briefly explain.

Generally speaking, the joint probability describes the probability of multiple events happening at the same time. By contrast, the conditional probability describes the probability of events happening given that other events happen. Typically, the joint probability is evaluated across the entire space while the conditional probability is evaluated across a smaller space in which some conditions are given.

B. Write down the factorization of the joint probability P(A,T,E,X,L,S,B,D) based on the network.



Based on the network and local semantics, the factorization is as follows.

$$P(A,T,E,X,L,S,B,D) = P(X|A,T,E,L,S,B,D)P(D|A,T,E,L,S,B)P(E|A,T,L,S,B)P(T|A,L,S,B)P(L|A,S,B)P(B|A,S)P(A|S)P(S)$$
$$= P(X|E)P(D|E,B)P(E|T,L)P(T|A)P(L|S)P(B|S)P(A)P(S)$$

Grade:__________

C. This video introduces a general way to determine independence relationships in Bayes Net: https://www.coursera.org/lecture/probabilistic-graphical-models/flow-of-probabilistic-influence-1eCp1.

Example: Is it true that tuberculosis ⫫ smoking | shortness of breath (given shortness of breath, tuberculosis and smoking are independent)?

Solution: There are two trails from T to S: (T, E, L, S) and (T, E, D, B, S). The trail (T, E, L, S) features a collider node E that is opened by the conditioning variable D. The trail is thus active and we do not need to check the second trail because for independence all trails needed to be blocked. The independence relationship does thus generally not hold.

Are the following conditional independence relationships true or false? Explain why.

1. either tuberculosis or lung cancer (E) ⫫ bronchitis (B) | smoking (S)

   True. There are two trails from E to B: (E, L, S, B) and (E, D, B). The trail (E, L, S, B) is blocked, because the non-collider node S is given. Besides, the trail (E, D, B) is also blocked, because the collider node D is not observed. Therefore, the conditional independence relationship in this case is true.

2. positive x-ray (X) ⫫ smoking (S) | lung cancer (L)

   True. There are two trails from X to S: (X, E, L, S) and (X, E, D, B, S). The trail (X, E, L, S) is blocked, because the non-collider node L is given. Besides, the trail (X, E, D, B, S) is also blocked, because the collider node D is not observed. Therefore, the conditional independence relationship in this case is true.

3. a visit to Asia (A) ⫫ bronchitis(B) | lung cancer (L), shortness of breath (D)

   False. There are two trails from A to B: (A, T, E, L, S, B) and (A, T, E, D, B). The trail (A, T, E, D, B) is active, because the collider node D is observed. There is no need to check the second trail, because for independence all trails needed to be blocked. Therefore, the conditional independence relationship in this case is false.

D. Express the P(D) by marginalizing the joint probability, simplify it using your answer in B

$$P(D) = \sum_{\text{all } A,T,E,X,L,S,B} P(A, T, E, X, L, S, B, D)$$

$$= \sum_{\text{all } A,T,E,X,L,S,B} P(X|E)P(D|E, B)P(E|T, L)P(T|A)P(L|S)P(B|S)P(A)P(S)$$