Group #_______
Names and IDs___________________

Sectional Written Homework #2: (**75 points**):

1. **(10 points)** Given the observed data below,

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Show your stepwise calculation for assigning the class label for a new animal with the following attribute values, using Naïve Bayes.

Give Birth	Can fly	Live in water	Have Legs	Class (Mammal or non-mammal)
No	yes	yes	no	?

(**No score** will be given, if you only answer "mammal" or "non-mammal")
Your answer:

Mammal Class: What chance that a new animal has all these traits while also being a Mammal?
P(Mammal)= 7/20
P(X| Mammal) = P(Give Birth(No) | Mammal)*P(Can Fly(Yes) | Mammal)*P(Live in Water(Yes) | Mammal)*P(Have Legs(No) | Mammal) = 1/7 * 1/7 * 2/7 * 2/7

P(Mammal | X) = P(X| Mammal)*P(Mammal) = 1/7 * 1/7 * 2/7 * 2/7 * 7/20 = 0.00058309037

Non-Mammal Class: What chance that a new animal has all these traits while not being a Mammal?
P(Non-Mammal)= 13/20

P(X| Non- Mammal)= P(Give Birth(No) | Non- Mammal)*P(Can Fly(Yes) | Non- Mammal)*P(Live in Water(Yes) | Non- Mammal)*P(Have Legs(No) | Non- Mammal) = 12/13 * 3/13 * 3/13 * 4/13

P(Non-Mammal | X) = P(X| Non-Mammal)*P(Non- Mammal) = 12/13 * 3/13 * 3/13 * 4/13 * 13/20 = 0.00983158852

While both Probabilities are low, Non-Mammal is more likely, therefore this animal is classified as Non-Mammal.

2. **(10 points)** Given the observed data and the reference table below,

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

naive Bayes classifier:

- (Refund = Yes | No) = 3/7
- (Refund = No | No) = 4/7
- (Refund = Yes | Yes) = 0
- (Refund = No | Yes) = 1
- (Marital Status = Single | No) = 2/7
- (Marital Status = Divorced | No) = 1/7
- (Marital Status = Married | No) = 4/7
- (Marital Status = Single | Yes) = 2/7
- (Marital Status = Divorced | Yes) = 1/7
- (Marital Status = Married | Yes) = 0

For Taxable Income:
If Class = No: sample mean = 110
sample variance = 2975
If Class = Yes: sample mean = 90
sample variance = 25

Show your stepwise calculation for assigning the class label for a new customer with the following attribute values, using Naïve Bayes.

Refund	Marital Status	Taxable Income	Evade Class (No or Yes)
Yes	Single	200K	?

(No score will be given, if you only answer "Yes" or "No")
Hint: For Taxable income, it follows the normal distribution.

$$P(x_j | C_i) = \frac{1}{\sqrt{2\pi\sigma_{ji}^2}} e^{-\frac{(x_j-\mu_{ji})^2}{2\sigma_{ji}^2}}$$

Your answer:

Yes Class:
P(Evade(Yes))= 3/10
P(Refund(Yes) | Evade(Yes))= 0
P(Marital Status(Single) | Evade(Yes))= 2/3
P(Taxable Income(200K) | Evade(Yes))= 1/sqrt(pi*50) * e^(-(110^2 / 50)) = 6.3485631057*10^−107

P(X | Evade(Yes))= P(Refund(Yes) | Evade(Yes))* P(Marital Status(Single) | Evade(Yes))* P(Taxable Income(200K) | Evade(Yes)) =0

P(Evade(Yes) | X)=P(X | Evade(Yes)) * P(Evade(Yes))=0

No Class:
P(Evade(No))= 7/10
P(Refund(Yes) | Evade(No))= 3/7
P(Marital Status(Single) | Evade(No))= 2/7
P(Taxable Income(200K) | Evade(No))= 1/sqrt(pi*5950) * e^(-(90^2 / 5950)) = 0.00187474481027

P(X | Evade(No))= P(Refund(Yes) | Evade(No))* P(Marital Status(Single) | Evade(No))* P(Taxable Income(200K) | Evade(No))= 3/7 * 2/7 * 0.00187474481027

P(Evade(No) | X)=P(X | Evade(No)) * P(Evade(No))= 3/7 * 2/7 * 0.00187474481027* 7/10 = 0.000160692412309

The probability for Evade(Yes) turns out to be 0 with Evade(No) being larger and this new customer should be in "No class".

3. **(15 points; 5 points *3)**

1) Is the total variance of a dataset equal to the variance explained by components identified in PCA?

Your answer:

Yes

2) Based on the loading matrix from the USarrests data, which variables will be counted into PC1 and which one will be counted into PC2?

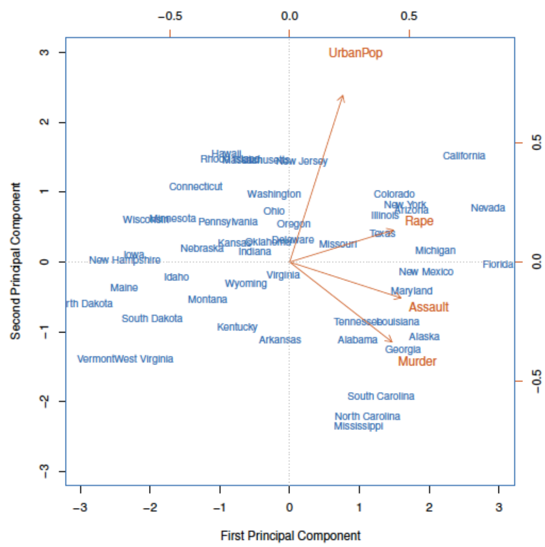
	PC1	PC2
Murder	0.5358995	-0.4181809
Assault	0.5831836	-0.1879856
UrbanPop	0.2781909	0.8728062
Rape	0.5434321	0.1673186

Your answer:

PC1: Murder, Assault, Rape

PC2: UrbanPop

3) What are the principal components scores shown on this bi-plot fusing USarrest data? What do the arrows indicate?



Your answer:

The principle component scores show where each state lands in the reduced dimension space created by PCA, states with high scores on PC2 have high urbanization and states with high scores on PC1 have more crime. The arrows indicate the loadings for each variable.

4. **(10 points; 2 points *5)**

1) How to deal with random initialization issues in K-means?

Your answer:

Deal with initialization issues in K-means by:

- Choose the first center as one of the examples, second which is the farthest from the first, third,which is the farthest from both, and so on.
- Trying multiple initializations and choosing the best result
- Using other smarter initialization schemes like the K-means++ algorithm

2) What algorithm can be used to deal with outliers, if k-means is sensitive to outliers?
Your answer:
K-medians

3) What are the assumptions for K-means?
Your answer:
The assumptions are:
- Cluster is spherical
- The spread or variance or density of the individual clusters is similar

4) What algorithm can we use to prevent local minima resulting from K-means?
Your answer:
Using K-means ++

5) How to choose the optimal number of K clusters?
Your answer:
Choose the optimal number of K clusters by:
   a) Compute k-means clustering using a range of k, e.g. k=1, 2,…, 5 clusters.
   b) For each k, calculate the cost, J, the total within-cluster sum of square using the cost funciton
   c) Plot the curve of J based on the number of clusters k.
   d) Find the inflection point ("Elbow") in the plot to be the optimal number of clusters.

Along with elbow plot, you can also use silhouette analysis to help find the optimal number of clusters.
(Note: if they add gap stats, that's ok; as long as they answer one of them, elbow plot, silhouette analysis and gap stats, it will be ok)

5. **(10 points) Write the K-means pseudo code for choosing 2-clusters for a sample of 100 cases with 2 attributes.**
Your answer:

a) Input: 100 cases (n) *2 attributes (x); Assume K=2
b) Initialize: Select 2 points as the initial centroids, using either random partition or Forgy initialization.
c) Repeat:
   Assign each of X_n to its closest centroid by computing its Euclidean distance to the centroids
   Recompute the new centroids by averaging data points in each cluster.
   until cluster centroids do not change anymore

6. **(10 points) Write the pseudo code for agglomerative hierarchical clustering.**
Your answer:

1. Begin with n observations and a measure (such as Euclidean distance) of all the $\binom{n}{2} = n(n-1)/2$ pairwise dissimilarities. Treat each observation as its own cluster.

2. For $i = n, n-1, \ldots, 2$:

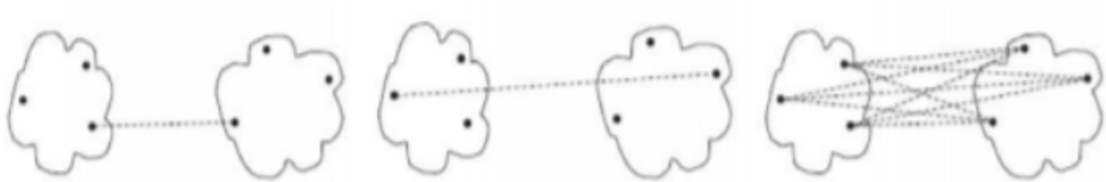
   (a) Examine all pairwise inter-cluster dissimilarities among the i clusters and identify the pair of clusters that are least dissimilar (that is, most similar). Fuse these two clusters. The dissimilarity between these two clusters indicates the height in the dendrogram at which the fusion should be placed.

   (b) Compute the new pairwise inter-cluster dissimilarities among the $i-1$ remaining clusters.

Simplified version :

- Start with each point in its own cluster.
- Identify the closest two clusters and merge them.
- Repeat.
- Ends when all points are in a single cluster.

7. **(5 points) What are the 3 dissimilarity measures in hierarchical clustering?**
   Your answer:
   As long as three names are given, earn points.
   Min-link/Single Link:
   - The minimum distance between data points of each cluster
     Clusters can get very large
   - Max-link/complete-link
     The maximum distance between data points of each cluster (Max-link)
     Small, round clusters
   - Average-link:
     The mean distance between data points of each cluster (Average-link)
     A compromise

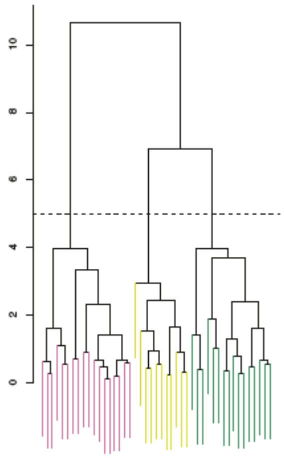


(a) MIN (single link.) (b) MAX (complete link.) (c) Group average.

- Ward’s method (Ward): minimum variance criterion; minimizes the total within-cluster variance

(Note: As asked three, any three of the above four would be ok.)

8. (**3 points)** How many clusters do we have if we cut at a height of 5 in this Figure?



Your answer:

3

9. **(2 Points)**Gap statistic and silhouette plots can be used to select the optimal number of clusters in hierarchical clustering? True or False

Your answer:

True