University of Massachusetts Dartmouth
Department of Computer and Information Science
CIS 490 Machine Learning – Exam II (Spring 2022)

Tuesday, April 26, 2022

Printed Full Name: Anubhav Shankar

Student ID: 01951462

DO NOT TURN THE PAGE OVER UNTIL YOU ARE INSTRUCTED TO DO SO

Please read the following instructions:

1. You have 75 minutes to complete the examination.
2. This examination is OPEN Notes
3. Type your answer in space provided on the examination sheets, any work not on the examination sheets will not be graded.
4. Type your answers legibly.
5. Submit your answer according to the instruction for grading by the end of the examination.
6. DO NOT communicate any of your classmates during the examination.

Honor Policy: copying in whole or in part of the examination will be considered to be an act of scholastic dishonesty. Students who violate university rules on scholastic dishonesty are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the university. Since such dishonesty harms individuals, all students, and the integrity of the university, policies on scholastic dishonesty will be strictly enforced.

I have read the above instructions and I will act in accordance with all of them.

Anubhav Shankar	04/26/2022
Student Signature	Date

Type your name and date to agree the policy before you start!

This examination contains three sections. The whole midterm examination carries 100 points.

Section I. Single-Choice Questions (20 points, 2 points per question; only ONE choice is correct). Please write your answers in the table provided below.

Question	1	2	3	4	5	6	7	8	9	10
Answer	a	a	b	b	c	b	d	d	d	a

1. If k-means is sensitive to outliers, what algorithm can be used to deal with outliers?
   a. K-medians
   b. K-means++
   c. K-means
   d. KNN

2. How can we prevent local minima resulting from K-means?
   e. Using K-means ++
   a. Using K-medians
   b. Using KNN
   c. Using K-mode

3. What is the tree size?
   a. The number of nodes
   b. The number of terminal nodes
   c. The number of subtrees
   d. The number of parent nodes

4. Which statistic does PCA looks at in the high-dimensional data?
   a. Mean
   b. Variance
   c. Correlation
   d. Median

5. What are the assumptions for K-means?
   a. The clusters are spherical
   b. The variance of clusters is similar
   c. All the above

6. What criteria can we use to select the optimal number of clusters in hierarchical clustering?
   a. Silhouette plots
   b. Gap statistic
   c. Cophenetic correlation
   d. All the above

7. Cross-validation (CV) can be used for?
   a. Only choosing the optimal tuning parameter in L1 regression
   b. Only choosing the optimal tuning parameter in L2 regression
   c. Only choosing the optimal tuning parameter in the pruning process of CART
   d. Choosing the optimal tuning parameter in regularized regression and CART

8. The feature detection layers of Convolutional neural network perform
   a. Only convolution
   b. Only pooling
   c. Only ReLU
   d. Convolution, pooling or ReLU

9. Which dissimilarity measures are <u>not</u> used in hierarchical clustering?
   a. Max-link
   b. Min-link
   c. Average-link
   d. None of the above

10. Which algorithm is widely used in normalized spectral clustering?
    a. Ng-Jordan-Weiss algorithm
    b. Greedy
    c. Top-down and greedy approach
    d. Top-down

Section II. True or False questions (20 points, 2 points per question).

Questions	True	False
1. Theoretically, the total variance of a dataset is equal to the variance explained by components identified in PCA	True	
2. For classification tree, we examine the MSE for accuracy		False
3. DNN typically contains 2 layers		False
4. Scaling would change the clustering results	True	
5. For hierarchical clustering, we draw conclusions about the similarity of two observations based on their proximity along the horizontal axis		False
6. Different similarity criteria can lead to different clustering results	True	
7. PCA is for clustering		False
8. K-means is for dimension reduction		False
9. K-means can be used when the clusters are non-convex		False

10. When Y is a continuous variable, multiple regression, regularized regression and regression trees can be considered.	True	

Section III. Short problems (60 points)

1. **(5 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)
Yes	No	No	Yes	?

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

Your answer:

Important -> Naïve Bayes assumes independence among attributes xj when given a class.

Mathematically,

p(x1, x2, … , xn | Ci) = p(x1 | Ci) p(x2 | Ci) … p(xn | Ci)

Here, X: attributes -> Give Birth, Can fly, Live in water, Have Legs.

Two classes -> M = Mammal, N = Non-Mammal

P(M) = 7/20 -(1)
P(N) = 13/20 –(2)

Two cases:

a.) P (Class = Mammal|X) α p(x | Class = mammal) p(Class = mammal)

b.) P (Class = Non-Mammal|X) α p(x | Class = Non-Mammal) p(Class = Non-Mammal)

Case 1 -> P (Class = Mammal|X) α p(x | Class = mammal) p(Class = mammal)

p(x **|** *Class = mammal) = p(GiveBirth = Yes* **|** *Class = mammal) * p(CanFly= No* **|** *Class = mammal) *p(LiveinWater=No***|** *Class = mammal) * p(HaveLegs = Yes***|** *Class = mammal)* - (3)

From (1)

n(Mammals) = 7

p(X|M) = 6/7 * 6/7 * 2/7 * 4/7 = 0.1199 -(4)

p(X|M) p(M) = 0.1199 * 7/20 = 0.0419 - (5)

Case 2 -> P (Class = Non-Mammal|X) α p(x | Class = Non-Mammal) p(Class = Non-Mammal)

*p(x | Class = Non-mammal) = p(GiveBirth = Yes | Class = Non-mammal) * p(CanFly= No | Class = Non-mammal) * p(LiveinWater=No| Class = Non-mammal) * p(HaveLegs = Yes| Class = Non-mammal)*   - (6)

From (2)

n(Non-Mammals) = 13

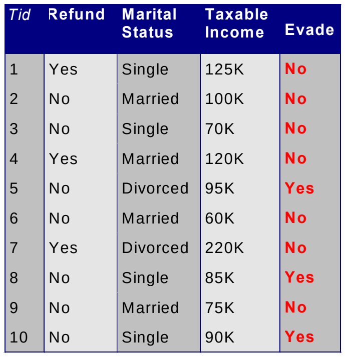
p(X|N) = 1/13 * 2/13 * 6/13 * 9/13 = 0.0037 - (7)

p(X|N) p(N) = 0.0037 * 13/20 = 0.0024 - (8)

From (5) and (8) ->

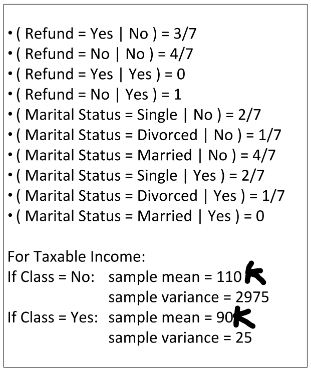
P(Class = M | X) > P(Class = N | X) => This case is a Mammal.

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 = 110K
                 sample variance = 2975
If Class = Yes: sample mean = 90K
                 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)
No	Single	80K	?

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

$$P(x_j \mid C_i) = \frac{1}{\sqrt{2\mathrm{ps}^{2}_{ji}}} e^{-\frac{(x_j - \propto_{ji})^2}{2\mathrm{s}^{2}_{ji}}}$$

Your answer:

Assumption -> Income follows a Gaussian/Normal Distribution.

New case, x = (Refund = No, Status = Single, Taxable Income = 80K)

P(Yes) = 0.3 -(1)

P(No) = 0.7 -(2)

*p(x | Class = No) = p(Refund = No | Class = No) * p(Single | Class = No) * p(Income = 80K | Class = No)*

- ⇨ p(x **|** Class = No) = 4/7 * 2/7 * p(Income = 80K **|** Class = No) - (3)
- ⇨ If class = No, then μ_{ji} = 110; Sample variance = 2975
- ⇨ P(Income = 80K **|** Class = No) = 1/(√2π * 2975) * $e^{(80-110/(2 * 2975))}$ = 0.0072 -(4)
- ⇨ Substituting (4) in (3) -> p(x **|** Class = No) = 0.00118 -(5)

*p(x | Class = Yes) = p(Refund = No | Class = Yes) * p(Single | Class = Yes) * p(Income = 80K | Class = Yes)*

- ⇨ P(x **|** Class = Yes) = 1 * 2/7 * p(Income = 200K **|** Class = Yes) – (6)
- ⇨ If class = Yes, then μ_{ji} = 90; Sample variance = 25
- ⇨ P(Income = 200K **|** Class = No) = 1/(√2π * 25) * $e^{(80-90/(2 * 25))}$ = 0.0653 -(7)
- ⇨ Substituting (7) in (6) -> p(x **|** Class = Yes) = 0.1866 -(8)

From (1), (2), (5), and (8) ->

p(x | Class = No) * P(No) < p(x | Class = Yes) * P(Yes) = 0.055

Hence, P(No | x) < P(Yes | x), and this case will be classified as Evade.

3. **(15 points, 5 points*3)** A researcher only has attributes X for a variety of dogs, and would like to explore or describe which species they belong to.

   1) Which machine learning method would this company use to help their decision, unsupervised or supervised?
   Your Answer: The researcher should use Unsupervised Learning methods.

   2) Please justify your decision based on your understanding of unsupervised or supervised learning methods in this case study. (3 points)
   Your Answer: Given the nature of the problem, it is clear that this is a classification problem. The researcher has a few attributes and for each, he can calculate the prior probability from the available data for each specie. He can then calculate the probability of a dog belonging to a particular species. For example, if there are ‘n’ dogs and ‘x’ species then he can calculate the probability belonging to a particular species and which ever has the highest probability, the dog can be categorized under that species.

   3) What specific supervised or unsupervised learning methods/models you would like to propose to your supervisor
   Your Answer: Given that this is a classification problem, I’d suggest using a Naïve-Bayes Classifier to execute this endeavor.

4. **(15 points, 5 points*3) PCA**

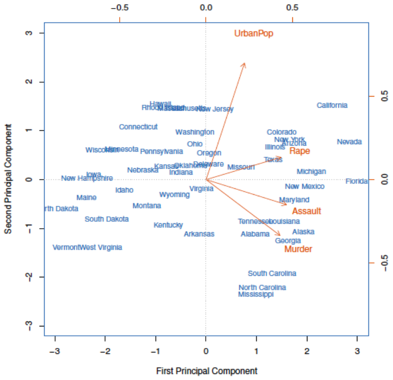
   1) Based on the loading matrix from the USarrests data, which variables will be counted into PC1 and which one will be counted into PC2?

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

   Your Answer:

   Based on the loading matrix, the variables Murder, Assault, and Rape will be counted in PC1 and thus indicate the measure of overall rates of serious crimes. A high correlation between the variables is observed. The variable, UrbanPop, will be counted in PC2 and thus measure the level of urbanization in the state.

   2) What are the principal components scores shown on this bi-plot using USarrest data?



   Your Answer: The PC loading scores for each variable is -> Murder(0.52, -0.41), Assault(0.58, -0.18), UrbanPop(0.27,0.87), Rape(0.54, 0.16). The negative sign indicates the direction.

   3) What do the arrows indicate in the above bi-plot using USarrest data?
   Your Answer: The arrows in the biplot indicate the first two principal components’ loading vectors.

5. **(5 points)** Write the K-means **pseudo code** for choosing 3-clusters for a sample of 200 cases with 3 attributes
Your Answer:

```
Select 3 points as the initial centroids
    Repeat:
        Form 3 clusters by assigning all 100 points to the closest centroid
        Recompute the centroid of each cluster
        Compare each individual's distance to its updated cluster mean and to that of the other cluster
    Until:
        The centroids don't change
```

6. **(10 points)** Write the **pseudo code** for using agglomerative hierarchical clustering to cluster 100 patients with 2 attributes and using a dendrogram to choose 3 clusters for this data.
Your answer: n = 100

Begin with '100' observations and a measure of all the (n * (n – 1))/3 pairwise dissimilarities. Treat each observation as its cluster.

For (i = n, n-1, …., 2) :

Examine all pairwise inter-cluster dissimilarities among the clusters and identify the pair of clusters that are least dissimilar. Fuse the two clusters.

Compute the new pairwise inter-cluster dissimilarities among the remaining clusters.