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?
2. K-medians
3. K-means++
4. K-means
5. KNN
6. How can we prevent local minima resulting from K-means?
7. Using K-means ++
8. Using K-medians
9. Using KNN
10. Using K-mode
11. What is the tree size?
12. The number of nodes
13. The number of terminal nodes
14. The number of subtrees
15. The number of parent nodes
16. Which statistic does PCA looks at in the high-dimensional data?
17. Mean
18. Variance
19. Correlation
20. Median
21. What are the assumptions for K-means?
22. The clusters are spherical
23. The variance of clusters is similar
24. All the above
25. What criteria can we use to select the optimal number of clusters in hierarchical clustering?
26. Silhouette plots
27. Gap statistic
28. Cophenetic correlation
29. All the above
30. Cross-validation (CV) can be used for?
31. Only choosing the optimal tuning parameter in L1 regression
32. Only choosing the optimal tuning parameter in L2 regression
33. Only choosing the optimal tuning parameter in the pruning process of CART
34. Choosing the optimal tuning parameter in regularized regression and CART
35. The feature detection layers of Convolutional neural network perform
36. Only convolution
37. Only pooling
38. Only ReLU
39. Convolution, pooling or ReLU
40. Which dissimilarity measures are not used in hierarchical clustering?
41. Max-link
42. Min-link
43. Average-link
44. None of the above
45. Which algorithm is widely used in normalized spectral clustering?
46. Ng-Jordan-Weiss algorithm
47. Greedy
48. Top-down and greedy approach
49. 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 |  |
| 1. For classification tree, we examine the MSE for accuracy |  | False |
| 1. DNN typically contains 2 layers |  | False |
| 1. Scaling would change the clustering results | True |  |
| 1. For hierarchical clustering, we draw conclusions about the similarity of two observations based on their proximity along the horizontal axis |  | False |
| 1. Different similarity criteria can lead to different clustering results | True |  |
| 1. PCA is for clustering |  | False |
| 1. K-means is for dimension reduction |  | False |
| 1. K-means can be used when the clusters are non-convex |  | False |
| 1. 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,



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.**

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





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.



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 µji = 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 µji = 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.**

1. **(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.
2. Which machine learning method would this company use to help their decision, unsupervised or supervised?

Your Answer: The researcher should use Unsupervised Learning methods.

1. 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.

1. 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.

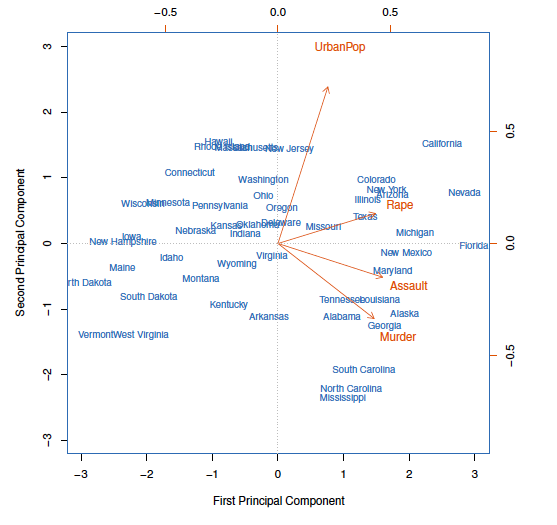
1. **(15 points, 5 points\*3) PCA**
2. Based on the loading matrix from the USarrests data, which variables will be counted into PC1 and which one will be counted into PC2?



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.

1. 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.

1. 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.

1. **(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

1. **(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:

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.