

AAA - MT_02__Semester_2225_MAX_POINTS_AT_TH... X

Attempt 1 of 1

Written Apr 13, 2023 12:44 PM - Apr 13, 2023 2:15 PM

Released Nov 30, 2022 4:18 PM

Attempt Feedback

Pradeep,

A good attempt.

I think you could take notes during lectures if you do not already do so. Moreover, I feel you should read the questions with more concentration. Your answers are mostly what you feel by trying to imagine the a real-world situation and analyzing how it would work out. Instead understanding more of what the Professor is explaining would help. He has tried and tested what he says.

I know you can do much better! All the best!

-Zubin

Attempt Score 122.5 / 147

Overall Grade (Highest Attempt) 122.5 / 147

Question 1 5 / 5 points

Dr. Kinsman is happy to drop almost all quizzes, and make the project optional. In this way, students who are good at programming can program the

project. Or, students who are good at taking exams can skip the project and focus on the exam.

Which of Dr. Kinsman's personal characteristics allows him to let students use their best strengths in this way?

Answer: Kindness **(/(?i)Compassion/)**

Question 2

5 / 5 points

I understand the RIT honor code for exams. All of the work here will represent my own understanding of the materials, and I will not receive help from anyone else during the exam.

I understand that this is an open-note, open-book, open-internet exam, but I may not communicate with anyone else, in any way, while taking the exam.

The answers to questions reflect the content of the course, and not the content of the internet.

Violation of the honor code may result in disciplinary actions.

YES

Question 3

10 / 10 points

The study of computer science involves the study of many theories about programs: Do they terminate? Are they provably correct? We study many algorithms.

What is the most important issue when getting an algorithm to work for data mining? Select either the algorithms or the features used? Support your argument.

Choosing the right features is more important than the selecting the algorithm. Choosing the wrong features will always yield in wrong results but the result of choosing the wrong features (which is wrong result) will

occur after consumption of resources. Given an opportunity selecting the features that are right is more important.

The correct answer is not displayed for Written Response type questions.

Question 4

0 / 2 points

DIST - recipe clustering.

Ryan wants to find clusters of recipes that use the same ingredients (not necessarily the same amount of each ingredient). What distance metric does he use to measure the distance between two recipes?

- The Supremum norm
- The L1 norm
- The Euclidean distance.
- The L2 norm
- The L0 norm
- The Manhattan distance

▼ Hide question 4 feedback

It would be The L0 norm

Question 5

0 / 2 points

Dist - Correlation Coef.

What is a good statement about the Correlation Coefficient between two variables:

S and T?

- If S and T are statistically independent, the mutual information is zero.
- It reveals how much information in T is revealed by knowing S.

- It is used to identify and remove outliers.
- The Correlation Coefficient between S and T goes up when the variable s is above its average and the variable t also above its average, or if they both tend to be below their averages at the same time.

Question 6

2 / 2 points

In PCA, what does the first principle component tell you ?

Answer: FIRST eigen vector WITH
Largest variance ✖ **(/Direction of max variation?/, /(?)
i).*max variation.*/)**

Question 7

0 / 5 points

What is an advantage of having a decision tree where the decisions are the same at each level of the tree?

Answer: more stable, less
likely to overfit ✖ **(/(?i)can be implemented with a look up
table/, /(?)implementation speed/)**

▼ Hide Feedback

It can be easily implemented with a look up table which can store the decisions made at each level. Thereby increasing implementation speed.

Question 8

2 / 2 points

In this course, in terms of Data Mining, what is a "Clustering" of data?

- A clustering is a set of cells which are infected.
- A cluster is a set of disk drives, used in servers.
- A clustering is a set of clusters, where each cluster is a set of data records.
- A clustering is a set of similar symptoms seen in a region.

Question 9

2 / 2 points

Using K-Means, what would change the shape of the resulting individual clusters?

- The selection of the distance metric between the points
- The number of iterations.
- The relative weights between attributes.
- The value of K.

▼ Hide question 9 feedback

This is the best answer. Remember the discussion of the Unit Circles -- and how they change shape for a distance metric.

Question 10

0.5 / 1 point

On the internet, you look up the linear algebra formula for The Kinsman Kunundrum, which is used to quantify how complicated a problem is. You find two different websites. They both claim to have the answer, but the math on one is backwards and transposed from the math on the other one. What has happened here?

(Select all that apply.)

- There issue is that Linear Algebra is hard to understand.
- There is no such thing as the Kinsman Kunundrum.
- This is what Dr. Kinsman calls, "Data Science's Dirty Little Secret".
- Electrical Engineers tend to store data records in columns, while Computer Scientists tend to store data records in rows.

▼ Hide question 10 feedback

Yes. Dr. Kinsman coined this phrase to help students learn about the issue.

Question 11**1 / 1 point**

Of the provided options, select two, or more, valid reasons we might want to use PCA?

- For dimensional reduction
- All of these are true
- For feature selection -- the most important features for class separation can sometimes be detected in the eigenvectors.
- To label the data as different clusters.
- It separates the data into classes.

Question 12**1 / 1 point**

You are given data that measures the amount of 12 different grocery items, which 30 people purchased. How many eigenvectors are there?

- 8
- 18
- 30
- 2
- 12

Question 13**1 / 1 point**

Given the equation:

$$\vec{M} \vec{v} = \lambda \vec{v}$$

Describe M and Lambda:

- M is a Matrix & Lambda is a Scalar amount
- M and Lambda are equal
- M is a Data Matrix & Lambda is a vector amount
- M is a Matrix & Lambda is a Projection amount
- M is great than Lambda

Question 14**1 / 1 point**

Consider a data point $D = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$,

and a projection vector $\begin{bmatrix} 2 & 1 \end{bmatrix}$

What is the dot product of these two?

- None of them
- $(2 * 1) + (3 * 2)$
- $(2 * 2) + (1 * 3)$
- $(2 * 2) + (3 * 2) + (2 * 1) + (3 * 1)$
- $(2 * 1) + (3 * 2) + (2 * 3) + (3 * 1)$

Question 15**5 / 5 points**

List three algorithms or applications we learned (or discussed) that are embarrassingly parallelizable?

- 1) Agglomeration
- 2) K MeANS
- 3) PCA

The correct answer is not displayed for Written Response type questions.

Question 16**4 / 4 points**

Given the following Cross Correlation Table, which attribute is least likely to help you classify the target variable?

Table B of Cross Correlation Coefs:

	Attr1	Attr2	Attr3	Attr4	Attr5	Target
Attr1	1	0.57	-0.01	0.63	-0.77	-0.98
Attr2	0.57	1	0	0.36	-0.44	-0.48
Attr3	-0.01	0	1	0.68	0.64	0.11
Attr4	0.63	0.36	0.68	1	-0.05	-0.61
Attr5	-0.77	-0.44	0.64	-0.05	1	0.81
Target	-0.98	-0.48	0.11	-0.61	0.81	1

- 1) Attr1
- 2) Attr2
- 3) Attr3
- 4) Attr4
- 5) Attr5
- 6) Attr6

Question 17

4 / 4 points

Given the following Cross Correlation Table, which attribute is most likely to help you classify the target variable?

Table B of Cross Correlation Coefs:

	Attr1	Attr2	Attr3	Attr4	Attr5	Target
Attr1	1	0.57	-0.01	0.63	-0.77	-0.98
Attr2	0.57	1	0	0.36	-0.44	-0.48
Attr3	-0.01	0	1	0.68	0.64	0.11
Attr4	0.63	0.36	0.68	1	-0.05	-0.61
Attr5	-0.77	-0.44	0.64	-0.05	1	0.81
Target	-0.98	-0.48	0.11	-0.61	0.81	1

1) Attr1

2) Attr2

3) Attr3

4) Attr4

5) Attr5

6) Attr6

Question 18

2 / 2 points

Snowfolks go by many different different names. In fact there is a meme which states, "Bigfoot is often confused with Sasquatch, Yeti never complains."

In terms of natural language processing, the words Bigfoot, Sasquatch, and Yeti all map to the same idea. What is this processing called?

[This is an example of "Concept Mapping". Try that answer.]

Answer: Concept Mapping ✓

Question 19

2 / 2 points

We analyze shopping behaviors of 800 shoppers, we find the following are the first two eigenvectors:

Milk	Pet Food	Veggies	Cereal	Bread	Rice	Meat	Eggs	Yogurt	Chips	Cola	Fruit
0.0	-0.2	0.3	-0.3	-0.2	0.5	0.0	0.0	0.4	-0.3	-0.5	0.3
0.5	0.2	0.1	0.5	0.4	0.2	-0.4	0.0	0.0	-0.2	-0.2	-0.1

A. Which two attributes are most important for the second Eigenvector?

B. Which attribute is least important, overall?

Answer for blank # 1: Milk,Cereal ✓(50 %)

Answer for blank # 2: Eggs ✓(50 %)

Question 20

2 / 2 points

You work with 33 records, each with 22 attributes. You are performing PCA. How many Eigenvector/Eigenvalue pairs are there?

- A) 11
- B) 22
- C) 33
- D) You need to use N-Fold cross validation
- E) You cannot tell until you look at the absolute values of the eigenvalues.

Question 21**2 / 2 points**

In the following two blanks, describe two different approaches to guessing the number of clusters that are in your data:

Answer for blank # 1: agglomerative clustering ✓(50 %)

Answer for blank # 2: k-means ✓(50 %)

Question 22**2 / 2 points**

We analyze shopping behaviors of 800 shoppers, we find the following are the first two eigenvectors:

Milk	Pet Food	Veggies	Cereal	Bread	Rice	Meat	Eggs	Yogurt	Chips	Cola	Fruit
0.0	-0.2	0.3	-0.3	-0.2	0.5	0.0	0.0	0.4	-0.3	-0.5	0.3
0.5	0.2	0.1	0.5	0.4	0.2	-0.4	0.0	0.0	-0.2	-0.2	-0.1

- A. Which two attributes are most important for the first Eigenvector?

B. What is their relationship?

Answer for blank # 1: Rice, Cola ✓(50 %)

Answer for blank # 2: Opposite directions ✓(50 %)

Question 23

2 / 2 points

You work with 33 records, each with 22 attributes. You are performing PCA. Where do the Eigenvectors come from?

- A) You need to use N-Fold cross validation.
- B) You cannot tell until you look at the absolute values of the eigenvalues.
- C) The Hyperbaloid Amalgaloid Learning (HAL) algorithm.
- D) The relative sizes of the eigenvalues.
- E) The covariance matrix, which is 22x22 in this case.

Question 24

2 / 2 points

Which principle of data mining says that you should prefer a SVM over an ANN, if they both produce similar accuracy, because an SVM is simpler?

- Occam's Razor
- The No-Free-Lunch Theorem
- N-Fold Cross Validation
- Random Sampling
- Parzen (Kernel) Density Estimation
- A histogram

- Mercer's Theorem
- The curse of dimensionality

Question 25**2 / 2 points**

What says that you **cannot** be sure if a deep-learning neural network is the best classifier to use?

- Occam's Razor
- The No-Free-Lunch Theorem
- N-Fold Cross Validation
- Random Sampling
- Parzen (Kernel) Density Estimation
- A histogram
- Mercer's Theorem
- The curse of dimensionality

Question 26**2 / 2 points**

Which is used to decide the best classifier to use, for sure?

- Occam's Razor
- The No-Free-Lunch Theorem
- N-Fold Cross Validation
- Random Sampling
- Parzen (Kernel) Density Estimation
- A histogram
- Mercer's Theorem
- The curse of dimensionality

Question 27**2 / 2 points**

Which of the following classifiers **DOES NOT** iterate when training?

- A 3-Layered ANN
- A decision tree
- A support vector machine
- ➔ A Naive Bayes Classifier
- A threshold classifier
- A random forest
- A set or random projections
- ➔ A Fisher Linear Differentiator or Linear Discriminant Analysis

Question 28

2 / 2 points

Which of the following classifiers iterate to set parameters when training?

- ➔ A 3-Layered ANN
- ➔ A decision tree using adaboost
- ➔ A support vector machine
- A Naive Bayes Classifier
- ➔ A threshold classifier
- A random forest
- A set or random projections
- A Fisher Linear Differentiator or Linear Discriminant Analysis

Question 29

4 / 4 points

What two factors slow down the speed of a k-NN classifier?

Answer: Size of the data, distance metric,
Dimensionality of the data metric,

✖ **(/(?)The Yagaguchi
cross ratio/)**

▼ Hide question 29 feedback

The choice of distance metric has no effect on the speed of the K-NN classifier. It influences the shape and size of the clusters formed as a result of applying it on the given data.

Question 30

2 / 4 points

When using a K-NN Classifier, what design decisions do you need to make?

Answer: K Value, Size of the dataset X **(/(?i)distance metric.*value of K/)**

▼ Hide question 30 feedback

Value of k is correct, the other would be the distance metric since it influences the shape and size of the cluster. You cannot determine the size of the dataset in design phase.

Question 31

0 / 2 points

Fisher made some assumptions about the classification problem. These assumptions allowed him to use the Rayleigh Quotient.

Dr. Kinsman says that this is an example of some kind of simplification. What is this an example of?

assumption simplification

The correct answer is not displayed for Written Response type questions.

▼ Hide question 31 feedback

It is an example of transformational thinking.

He transformed the problem to a simpler problem which led to the Rayleigh Quotient.

Question 32

0 / 2 points

In a two class problem, for which the answer is a linear decision boundary, what is the dual of the decision boundary?

set of weights that can **X** **(/(?i)a projection vector orthogonal to**
Answer: be used to classify new **the decision boundary/, /(?i).*orthogonal.*/,**
data points **/(?i).*vector.*/**

▼ Hide question 32 feedback

It would be the projection vector which is always orthogonal to the decision boundary.

Question 33

0 / 2 points

How many times does the Fisher Linear Discriminant iterate?

Answer: 0 **X** **(/1/, /once/, /only once/)**

▼ Hide question 33 feedback

It would be once.

Try to understand, running it once would be mean 1 iteration.

I understand that you meant it does not iterate but it means it runs once, which is the above.

Question 34

2 / 2 points

What is the difference between correlation and causation?

Correlation can be positive, negative and zero depending on the direction and strength of relation.

Causation is the relationship between two variables where one variable directly influences the other variable.

The correct answer is not displayed for Written Response type questions.

Question 35**2 / 2 points**

What is the important design decision that you make for implementing K-Nearest Neighbors classification?

Answer: distance metric and k value ✓

Question 36**2 / 2 points**

Given a dendrogram, where does it naturally split the data?

To determine the natural split in the data using a dendrogram, you can visually inspect the dendrogram and look for a point where there is a significant increase in the distance between branches or where the height of the tree is particularly high.

The correct answer is not displayed for Written Response type questions.

Question 37**5 / 5 points**

Graphically, what regions form to reveal the solution to the 1-NN (i.e. the one nearest neighbor classifier) problem? What shapes result?

Answer: voroni
 polygon ✗ **(/(?i)voronoi polygons/, /(?)i).*voronoi.*/, /(?)i)voronoi/)**

Question 38**5 / 5 points**

I work for Amazon. I want to form clusters of places to drop off deliveries. I plan to use Agglomeration to help with the route planning?

What agglomerative design decisions will I need to make?

What kind of things might I penalize to make the route most effective?

design decisions :

- 1) Distance matrix to understand the nearest storage space where all the products are available.

- 2) Linkage (Shortest linkage is the best)
- 3) Evaluation (number of miles that a product is travelling)

Penalizing :

- 1) Orders from one person should be clustered and delivered if not penalize.

The correct answer is not displayed for Written Response type questions.

▼ Hide question 38 feedback

Another thing to penalize could be right turns. A company actually did this.
Try looking it up! :)

Question 39

4 / 5 points

For your agglomeration homework, you found that there were four main clusters in the shopping data. In the associated space, sketch a dendrogram that starts with 10 data points, and merges them together, indicating that there are four main clusters.

[You can sketch a chart somewhere else, and upload an image if necessary.]

- No text entered -

✉ [dendo.jpeg](#) (70.39 KB)

The correct answer is not displayed for Written Response type questions.

▼ Hide question 39 feedback

In your dendrogram the biggest jump in between branches is where there are two branches or clusters.

Question 40

0 / 5 points

When clustering, why does the choice of a distance metric matter? What difference does it make?

If you are clustering data with continuous features, the Euclidean distance metric may be a good choice because it captures the overall difference in magnitude between feature values. On the other hand, if you are clustering data with categorical features, a distance metric that is designed to handle categorical data is not the right fit. Choosing Euclidean distance in place of some other distance may result in inaccurate clustering as well.

The correct answer is not displayed for Written Response type questions.

▼ Hide question 40 feedback

It influences the size and shape of clusters formed.

L1: Square-shaped clusters

L2: Circular

Question 41

5 / 5 points

I want to form clusters of Wordle Words using Agglomeration.

What design decisions will I need to make?

Why would the standard method of Agglomeration be difficult?

Design decisions to make :

- 1) How to find the distance between two words (similarity in NLP words)
- 2) Linkage (farther, average, minimum)
- 3) Not important but need to check how the cluster is (cluster evaluation) use impurity method.

Wordle words are discrete and creating a distance matrix for it will be

difficult, separate words mean dealing with high dimensional data around 2309 words so very difficult and computationally expensive.

The correct answer is not displayed for Written Response type questions.

Question 42

4 / 5 points

What is the curse of dimensionality? Why do we care? Why is this important to understand?

How does reducing the number of dimensions help?

the problems that arise when working with high-dimensional data, where the number of features or variables is large relative to the number of observations.

If we continue to work with high dimensional data it will take a significant computational capacity (may take months to solve a problem) also in multidimension the is with sparsity of how the points are distributed.

Reducing the number of dimensions can help to mitigate these issues and make the data easier to analyze and model. Dimensionality reduction techniques, such as PCA can be used to transform high-dimensional data into lower-dimensional representations which reduces the complexity as the projections will not sparse and resolving a problem in lower dimensions is very better (computationally less expensive).

The correct answer is not displayed for Written Response type questions.

▼ Hide question 42 feedback

The curse of dimensionality is that as the number of dimensions increase, the data becomes more sparse.

You did not state the curse of dimensionality. Otherwise you answered the rest of the questions.

Question 43**7 / 7 points**

Suppose you have data with 18 features. You want to be able to view it in three dimensions.

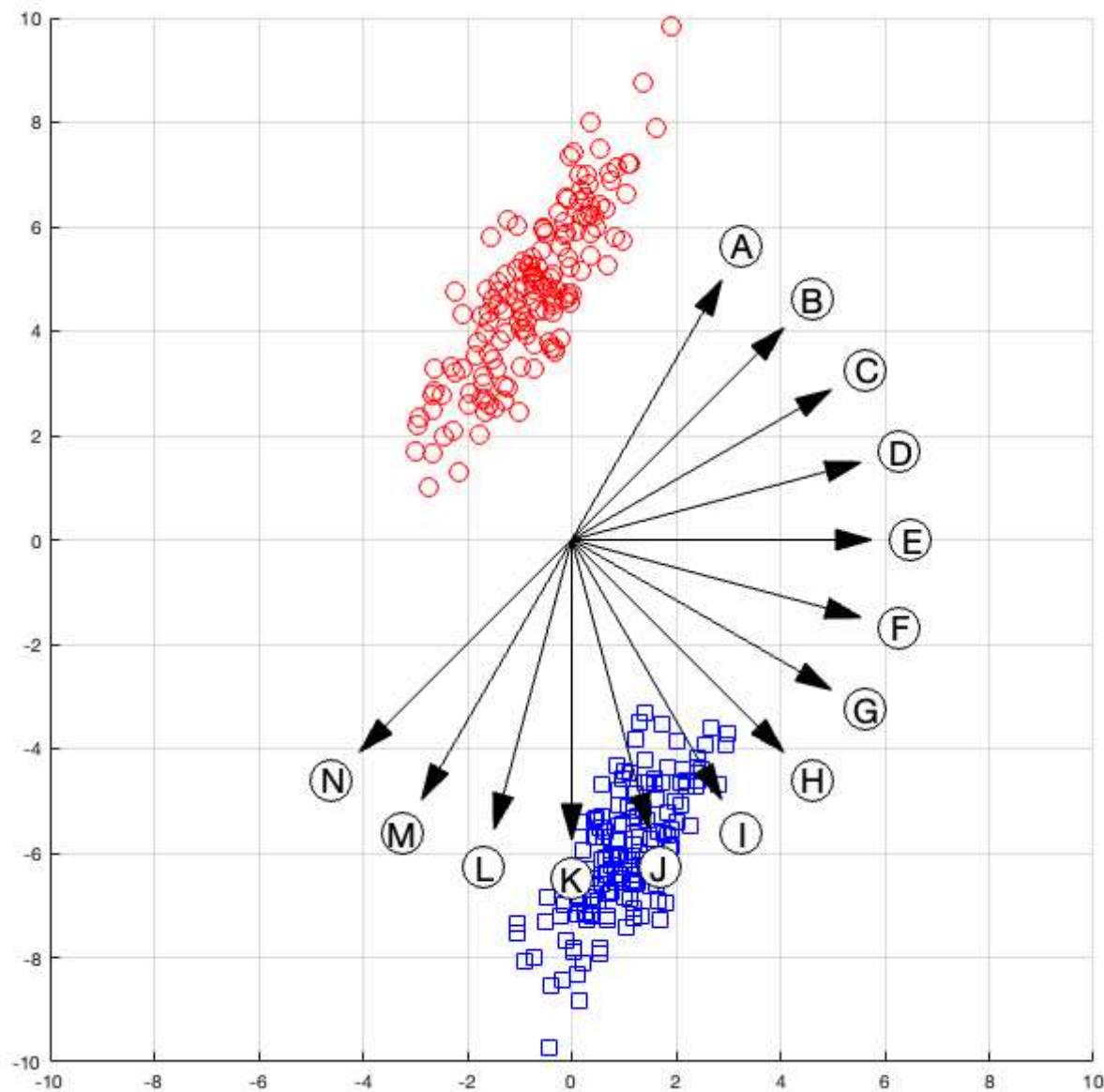
- A. What technique have we studied that would allow you to plot it and work with it in three dimensions?
- B. What are the steps involved in this process?

Principal Component Analysis (PCA) is the technique that can be used to convert the data into 2D/3D.

- 1) Sub mean values
- 2) Find the covariance matrix
- 3) Select the top three principal components (eigen vectors and values)
- 4) Plot these three

The correct answer is not displayed for Written Response type questions.

Question 44**0 / 10 points**



The above diagram shows two elliptical distributions, with several labeled vectors to pick from. In the blank below, give the label of the vector that would be the best Fisher Linear Discriminant (FLD) to project the data onto.

Answer: d (is perpendicular to the line that touches the centers of both clusters)

(/ (?i)g/, /G/)

▼ Hide question 44 feedback

G would be the correct FLD. It is the first best guess. It is drawn from the Center of Mass of one cluster to the other.

Question 45**10 / 10 points**

The associated image shows two different Gaussian distributions: Class A (in Red) and Class B (in Blue).

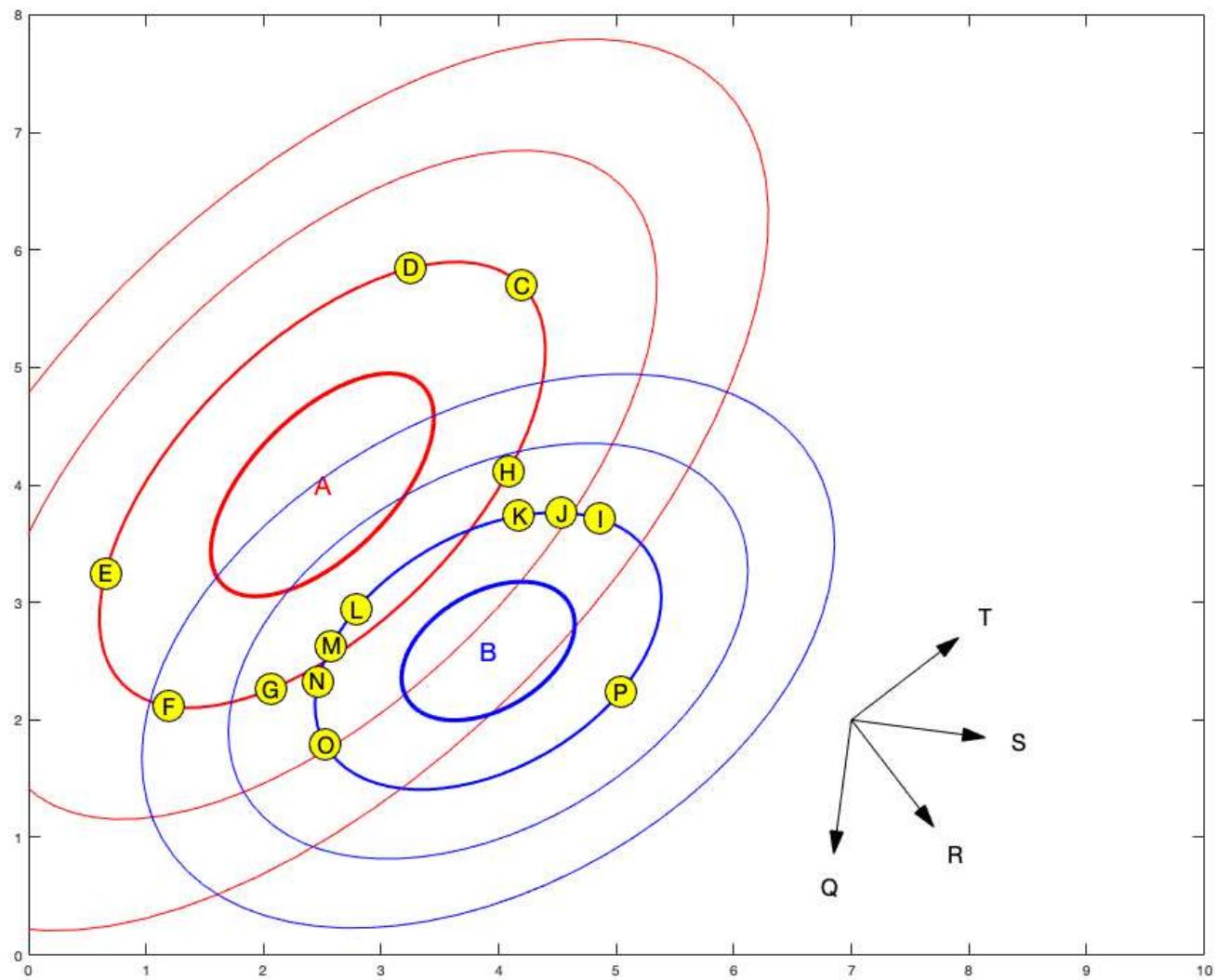
On top of each Gaussian distributions are two sets of Mahalanobis distances, one for each class.

Superimposed over everything are several points, labeled C, D, E, F, ... in yellow circles.

List the points in alphabetical order which would be classified as belonging to Class A.

If the points are 50/50, or near the decision boundary so that they could be Class A or Class B, do not include them. List the letters only, for example:

N O P Q R



Answer: Class A : CDEHGFLM; Class B: IJKNOP ✓

Question 46

5 / 10 points

What techniques do you know for attribute selection or rejection? (Called "feature selection" in general.) Why is this important?

How does this help you do data mining?

Why do we care?

Correlation Matrix :

Correlation matrix can be used for feature selection or rejection. Higher the absolute value of correlation higher will be the inine direction of both the

records.

Covariance Matrix:

Covariance matrix measures the degree and direction of linear relationship between pairs of attributes in the data. If two attributes have a high positive covariance, it means that they tend to increase or decrease together, while a high negative covariance indicates that they tend to move in opposite directions. On the other hand, a low covariance suggests that the two attributes are not related or are weakly related.

Above mentioned two are the exemplary techniques. It is very important to reject a few features in a real world case where the attributes are very high in number this will result in longer computation times and no relevant results.

In data mining, by rejecting no relevant variables will consume a less time in computation and will be helpful with the accuracy and precision of the trained models.

It is expensive to buy the computation power in the realworld. Using computation resources for something that will not yield better results is waste of resources.

The correct answer is not displayed for Written Response type questions.

▼ Hide question 46 feedback

Techniques could be Forward Feature Selection and Backward Feature Elimination. It could be the first choiced eigen vectors found from PCA.

Longer computation times is an issue that is solved by dimensional reduction.

Done