

Name and Student ID:

Machine Learning, BLG527E, March 21, 2018, Midterm Exam.

	1 (20)	2 (20)	3 (20)	4 (20)	5 (20)	Total
Predicted						
Actual						

Name:

Number:

Signature:

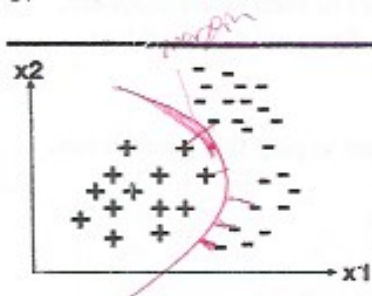
Duration: 110 minutes.

Write your answers neatly in the space provided for them. Write your name on each sheet.
Books and notes are closed. Good Luck!

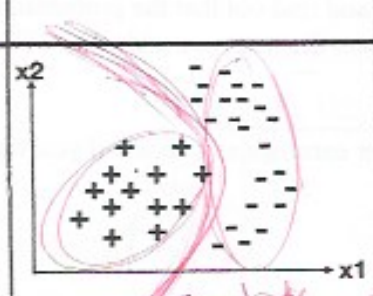
QUESTIONS

Q1) [20 points, 5 points each] What is (use at most three sentences per question, you can use drawings, formulas, etc. also):

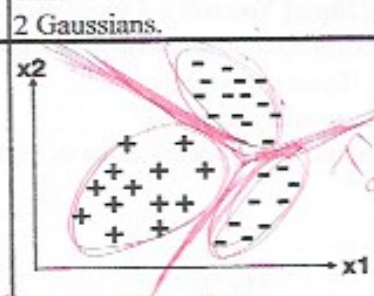
a) Draw a maximum margin classifier with a differentiable discriminant function for the dataset below. Assume that the number of instances from each class is the same, $N_+ = N_-$.



b) Draw the discriminant function if you were to use multivariate classification, assuming different class covariance matrices.

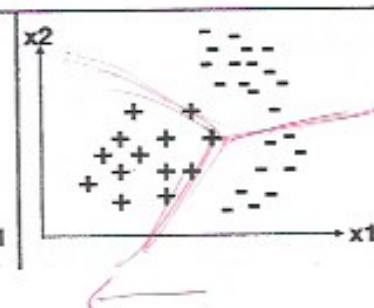
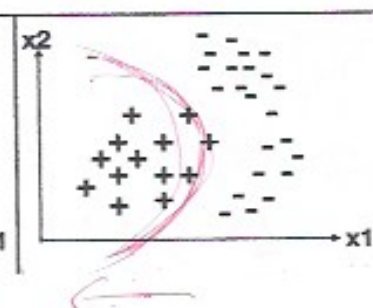
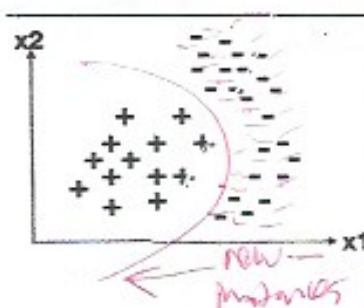


c) Draw the discriminant function if you were to use multivariate classification, assuming different class covariance matrices and also assuming that the negative class is made of a mixture of 2 Gaussians.



$$g(x) = x^T \theta x + w^T x + b$$

d) Assume now that $10 N_+ = N_-$, i.e. you have 10 times as many negative instances as positive instances. Draw the new decision boundaries for each of the cases above.

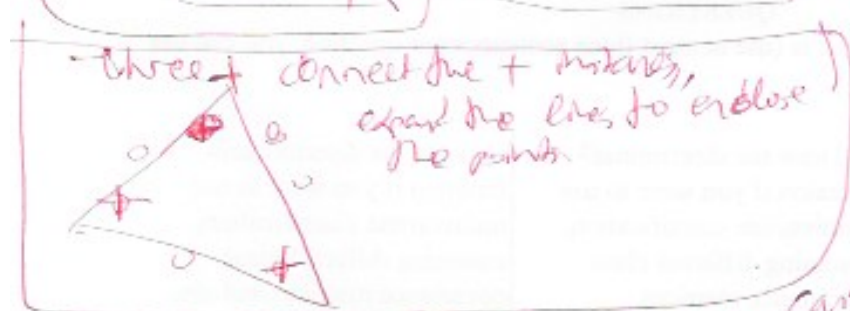
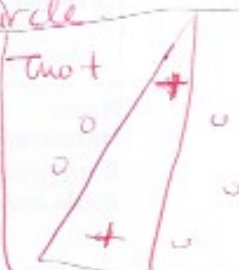
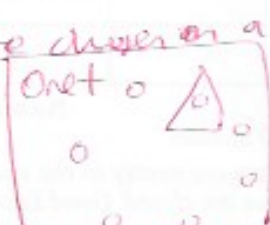
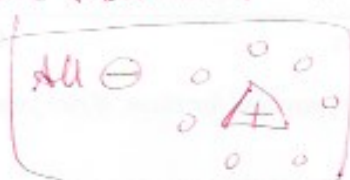
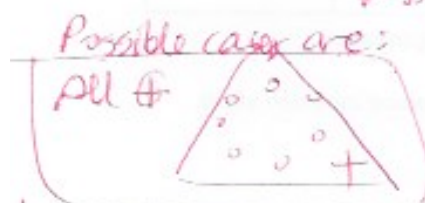


all discriminant surfaces will move towards + class

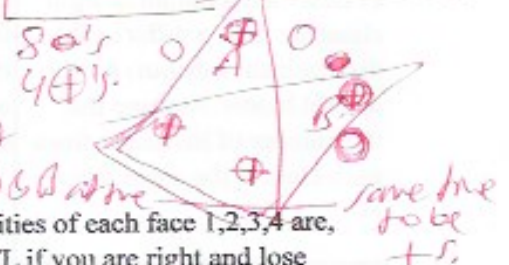
Name and Student ID:

Q2)[10pts] Show that the VC dimension of a triangle in 2D space is 7. Choose data points in such a way that, all labelings of 7 points can be classified and some labelings of 8 points can not be classified.

Assume that $\triangleleft$ or $\triangleleft^+$ are possible with the classifier.
Assume that all points are chosen on a circle.



One $\triangleleft$ $\triangleleft^+$
Two $\triangleleft$ $\triangleleft^+$
Three $\triangleleft$ $\triangleleft^+$
Four $\triangleleft$ $\triangleleft^+$
Five $\triangleleft$ $\triangleleft^+$
Six $\triangleleft$ $\triangleleft^+$
Seven $\triangleleft$ $\triangleleft^+$



Q3)[20pts] You roll a 4 faced dice and find out that the probabilities of each face 1,2,3,4 are, 0.1,0.2,0.3,0.4 respectively. You guess the next toss and win 10TL if you are right and lose 5TL if you are wrong.

What would you guess? 4 all the time.

Compute the expected value of your earnings and decide if you want to play this game or not.

$$E(\text{earnings}) = 0.1 \times (-5) + 0.2 \times (-5) + 0.3 \times (-5) + 0.4 \times 10$$

$$= (0.1 + 0.2 + 0.3) \times (-5) + 4$$

$$= -3 + 4 = 1 > 0$$

213

Since $E(\text{earnings}) > 0$, I would play this game.

Name and Student ID:

Q4)[20pts] What are the differences and similarities between PCA (Principal Component Analysis), LDA (Linear Discriminant Analysis)?

PCA

LDA

Diff: On unlabeled data
As many components as the # of features
Sim: Dimensionality reduction
Does not require class labels

On labeled data
As many components as $K-1$ (# classes)
Dimensionality reduction
Does require class labels

What are the differences and similarities between mRMR (minimum Redundancy Maximum Relevance) algorithm and forward feature selection?

mRMR

Forward Feature Selection

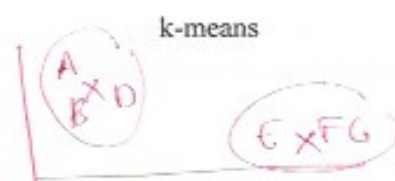
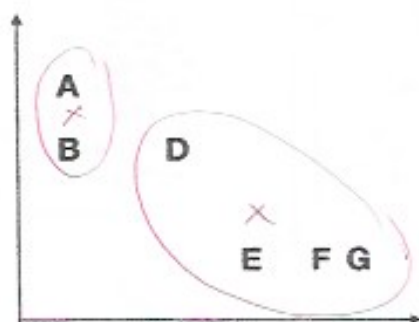
Sim: Forward feature sel.
No need to train models

Forward feature sel.
✓

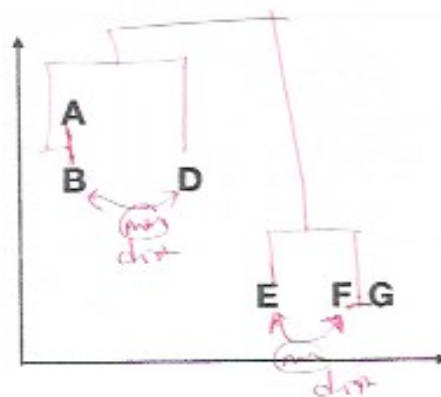
Diff: No need to train models
Faster

Need to train models
Slower

Q5)[20pts] Cluster the dataset on the left using k-means algorithm, $k=2$ and starting with cluster centers of B and D. Cluster the dataset on the right using hierarchical clustering with single link (minimum) distance.



x: center



hierarchical



Dendrogram