Name: ____

Problem	Points	Score
1(a)	10	
1(b)	10	
1(c)	10	
2(a)	10	
2(b)	10	
2(c)	10	
3(a)	10	
3(b)	10	
4(a)	10	
4(b)	10	
Total	100	
4(b)	10	

Notes:

(1) The exam is closed books and notes, but you are allowed one sheet of notes.

(2) Please indicate clearly your answer to the problem.

(3) Note that ungrammatical sentences, incoherent statements, or general illegible scratches will get zero credit.

(4) If I can't read or follow your solution, it is wrong, and no partial credit will be awarded.

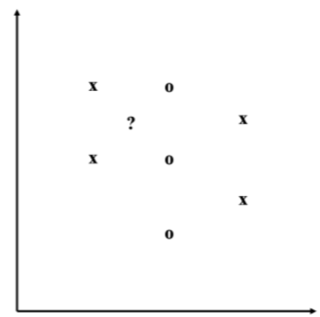
Problem No. 1: Consider a black box that outputs a random scalar value. The box has a three-way switch labeled “class 1” (left), “class 2” (middle) and “class unknown” (right). You throw the switch in the left position and obtain two samples: $[-1, 1]$. You move the switch to the middle position and draw two samples: $[1/2, 2]$. You throw the switch into the third position, collect some values, and attempt to identify which class they belong to.

(a) (10 pts) Discuss how you would build a machine to classify the data. Be very specific. Don’t simply say things like “I would use PCA.” Start from basic issues and address all the obstacles you would encounter.

(b) (10 pts) Now you assume the underlying model for the generator is Gaussian. Repeat (a), but again be very specific about the algorithms you use and the assumptions they make.

(c) (10 pts) Next, you put the switch in the left position and measure a new value: $[-0.5]$. How would you use Bayesian techniques to sharpen your estimates of your models in (a) and (b).

Problem No. 2: Consider the two-class problem shown to the right. Assume each class is equiprobably.



(a) (10 pts) Describe how PCA and LDA would classify this data using sketches, showing decision surfaces, etc. Which one works better? Justify your answers.

(b) (10 pts) Compare and contrast QDA and PCA for this data set.

(c) (10 pts) Suppose you were to represent the optimal decision surface separating these classes using just two data points from each class. Which data points would you select and why? (We will later see this is a method called Support Vector Machines.)

Problem No. 3: Consider two probability distributions defined by:

$$p(\omega_1|x_1,x_2) = \begin{Bmatrix} 1 & 0 \le x_1 \le 1, 0 \le x_2 \le 1 \\ 0 & elsewhere \end{Bmatrix}$$

$$p(\omega_2|x_1,x_2) = \begin{Bmatrix} 1 & \alpha \le x_1 \le 1+\alpha, \alpha \le x_2 \le 1+\alpha \\ 0 & elsewhere \end{Bmatrix}$$

(a) (10 pts) Assuming the prior probabilities, $P(\omega_1) = P(\omega_2) = 0.5$, sketch the probability of error, $P(E)$, for a maximum likelihood classifier as a function of α. Label all critical points.

(b) (10 pts) How does the shape of this plot change if $P(\omega_1) = 0.75$ and $P(\omega_2) = 0.25$? Sketch the new shape and label all critical points. Justify your answer.

Problem No. 4: Suppose we have a discrete random variable, X, that takes on one of two values, 0 or 1, with the following probabilities:

$$p(x_i) = \begin{Bmatrix} 1-\alpha & x = 0 \\ \alpha & x = 1 \end{Bmatrix}$$

(a) (10 pts) What is the maximum likelihood estimate of α? Justify your answer.

(b) (Bonus: 10 pts) How would this answer change if you assumed the data was actually drawn from a Gaussian distribution, but for some reason generated the above distribution. In other words, what were the underlying statistical assumptions of your answer to (a).