15-781 Midterm, Fall 2001

YOUR ANDREW USERID IN CAPITAL LETTERS:

YOUR NAME:

- There are 6 questions.
- Questions 1-5 are worth 19 points each.
- Question 6 is potentially time-consuming and worth only 5 points. Only attempt it if you are certain there aren't any missing parts or errors in your answers to the other questions.
- The maximum possible total score is 100.

1 Decision Trees (19 points)

The following dataset will be used to learn a decision tree for predicting whether a person is happy (H) or sad (S) based on the color of their shoes, whether they wear a wig and the number of ears they have.

Color	Wig	Num. Ears	(Output) Emotion
G	Y	2	S
G	N	2	S
G	N	2	S
B	N	2	S
B	N	2	H
R	N	2	H
R	N	2	H
R	N	2	H
R	Y	3	H

(a) *(2 points)* What is H(Emotion|Wig=Y)?

(b) *(2 points)* What is H(Emotion|Ears=3)?

(c) *(3 points)* Which attribute would the decision-tree building algorithm choose to use for the root of the tree (assume no pruning).

(d) *(3 points)* Draw the full decision tree that would be learned for this data (assume no pruning).

(e) *(3 points)* What would be the training set error for this dataset? Express your answer as the percentage of records that would be misclassified.

The next two parts do not use the previous example, but are still about decision tree classifiers.

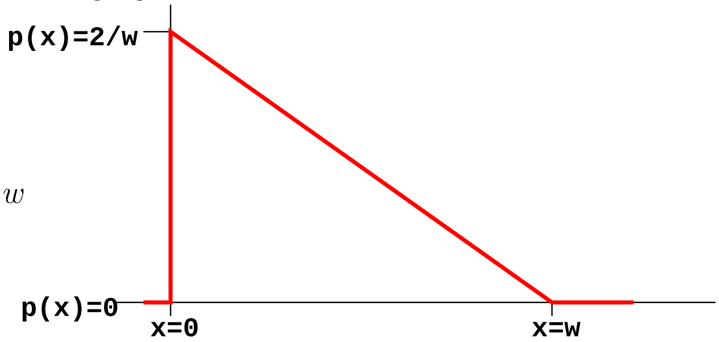
(f) *(3 points)* Assuming that the output attribute can take two values (i.e. has arity 2) what is the maximum training set error (expressed as a percentage) that any dataset could possibly have?

(g) *(3 points)* Construct an example dataset that achieves this maximum percentage training set error. (It must have two or fewer inputs and five or fewer records).

2 Probability Density Functions (19 points)

Consider the PDF shown in the following figure and equations

$$
\begin{aligned}
p(x) &= 0 && \text{if } x < 0 \\
p(x) &= \frac{2}{w} - \frac{2x}{w^2} && \text{if } 0 \le x \le w \\
p(x) &= 0 && \text{if } w < x
\end{aligned}
$$



(a) *(3 points)* Which one of the following expressions is true? (note—exactly one is true). Write your answer (a choice between 1 to 12) here:

(1) $E[X] = \int_{x=-\infty}^{\infty} (\frac{2}{w} - \frac{2x}{w^2})dx$

(2) $E[X] = \int_{x=-\infty}^{\infty} x(\frac{2}{w} - \frac{2x}{w^2})dx$

(3) $E[X] = \int_{x=-\infty}^{\infty} w(\frac{2}{w} - \frac{2x}{w^2})dx$

(4) $E[X] = \int_{x=0}^{w} (\frac{2}{w} - \frac{2x}{w^2})dx$

(5) $E[X] = \int_{x=0}^{w} x(\frac{2}{w} - \frac{2x}{w^2})dx$

(6) $E[X] = \int_{x=0}^{w} w(\frac{2}{w} - \frac{2x}{w^2})dx$

(7) $E[X] = \int_{w=-\infty}^{\infty} (\frac{2}{w} - \frac{2x}{w^2})dw$

(8) $E[X] = \int_{w=-\infty}^{\infty} x(\frac{2}{w} - \frac{2x}{w^2})dw$

(9) $E[X] = \int_{w=-\infty}^{\infty} w(\frac{2}{w} - \frac{2x}{w^2})dw$

(10) $E[X] = \int_{w=0}^{x} (\frac{2}{w} - \frac{2x}{w^2})dw$

(11) $E[X] = \int_{w=0}^{x} x(\frac{2}{w} - \frac{2x}{w^2})dw$

(12) $E[X] = \int_{w=0}^{x} w(\frac{2}{w} - \frac{2x}{w^2})dw$

(b) *(4 points)* What is $P(x = 1|w = 2)$?

(c) *(4 points)* What is $p(x = 1|w = 2)$?

(d) *(4 points)* What is $p(x = 0|w = 1)$?

(e) *(4 points)* Suppose you don't know the value of w but you observe one sample from the distribution: $x = 3$. What is the maximum likelihood estimate of w?

3 Gaussian Bayes Classifiers (19 points)

(a) *(2 points)* Suppose you have the following training set with one real-valued input X and a categorical output Y that has two values.

X	Y
0	A
2	A
3	B
4	B
5	B
6	B
7	B

You must learn the Maximum Likelihood Gaussian Bayes Classifier from this data. Write your answers in this table:

$\mu_A =$	$\sigma^2_A =$	$P(Y = A) =$
$\mu_B =$	$\sigma^2_B =$	$P(Y = B) =$

I considered asking you to compute $p(X = 2|Y = A)$ using the parameters you had learned. But I decided that was too fiddly. So in the remainder of the question you can give your answers in terms of α and β, where:

$$\begin{aligned} \alpha &= p(X = 2|Y = A) \\ \beta &= p(X = 2|Y = B) \end{aligned}$$

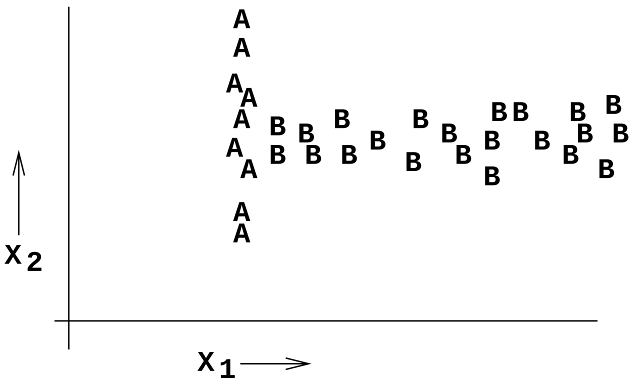
(b) *(2 points)* What is $p(X = 2 \wedge Y = A)$ (answer in terms of α)?

(c) *(2 points)* What is $p(X = 2 \wedge Y = B)$ (answer in terms of β)?

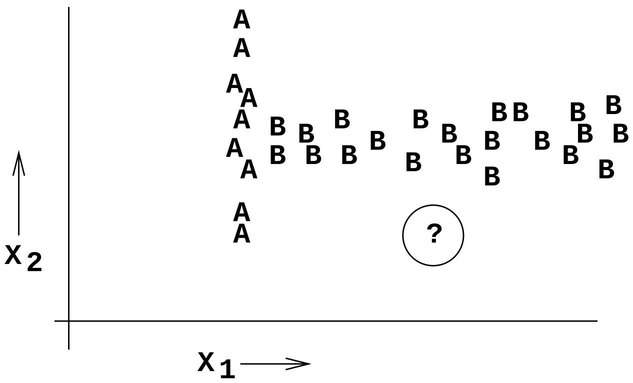
(d) *(2 points)* What is $p(X = 2)$ (answer in terms of α and β)?

(e) *(2 points)* What is $P(Y = A|X = 2)$ (answer in terms of α and β)?

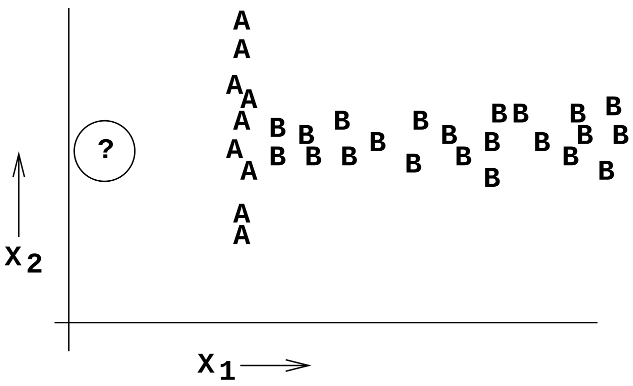
In a completely different Bayes Classifier example, suppose you trained a Bayes Classifier using General Gaussians on the following data that has two real-valued inputs (X_1 and X_2) and one two-valued categorical output Y.



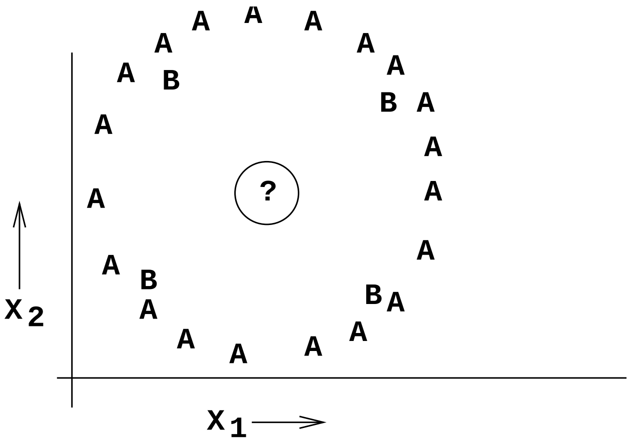
(f) *(3 points)* What class would the Bayes Classifier predict for inputs at the location shown by a question mark in the following figure? (Note: I strongly advise you to answer this by common sense and "eyeballing"—don't waste precious time calculating the Bayes Classifier.



(g) *(3 points)* And what class would it predict for the following location?



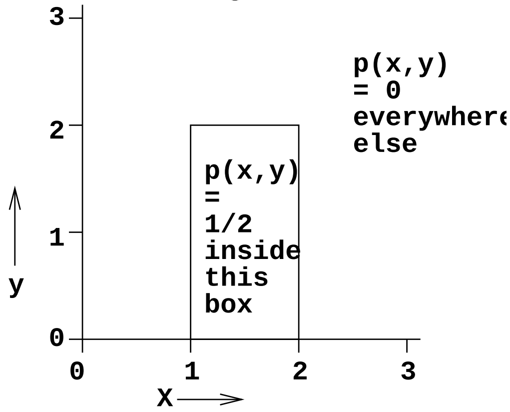
(h) *(3 points)* Finally, consider the following figure. If you trained a new Bayes Classifier on this data, what class would be predicted for the query location?



4 Bivariate PDFs (19 points)

Consider the following PDF defined by these equations and sketched in this figure:

$$
\begin{aligned}
p(x,y) = &\ \tfrac{1}{2} \quad \text{if } x \geq 1 \wedge x \leq 2 \wedge y \geq 0 \wedge y \leq 2 \\
p(x,y) = &\ 0 \quad \text{Otherwise}
\end{aligned}
$$

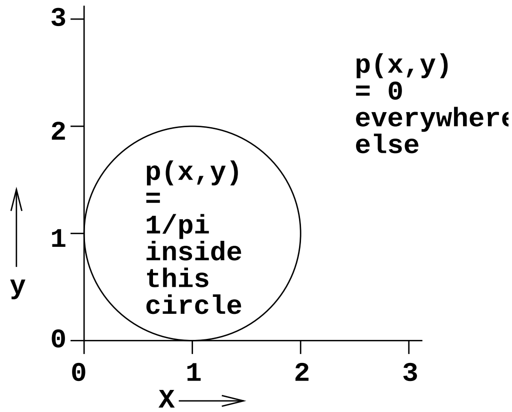


(a) *(7 points)* What is $p(x)$? (your answer should be a function of x that integrates to 1)

(b) *(6 points)* Is X independent of Y?

Now consider a different joint distribution over (x,y):

$$
\begin{aligned}
p(x,y) = &\ \tfrac{1}{\pi} \quad \text{if } (x-1)^2 + (y-1)^2 \leq 1 \\
p(x,y) = &\ 0 \quad \text{Otherwise}
\end{aligned}
$$



(c) *(6 points)* Is X independent of Y?

5 Bayes Rule (19 points)

(a) *(4 points)* I give you the following fact:

$$P(A|B) = 2/3$$

Do you have enough information to compute $P(B|A)$? If not, write "not enough info". If so, compute the value of $P(B|A)$.

(b) *(5 points)* Instead, I give you the following facts:

$$\begin{aligned} P(A|B) &= 2/3 \\ P(A|{\sim}B) &= 1/3 \end{aligned}$$

Do you now have enough information to compute $P(B|A)$? If not, write "not enough info". If so, compute the value of $P(B|A)$.

(c) *(5 points)* Instead, I give you the following facts:

$$\begin{aligned} P(A|B) &= 2/3 \\ P(A|{\sim}B) &= 1/3 \\ P(B) &= 1/3 \end{aligned}$$

Do you now have enough information to compute $P(B|A)$? If not, write "not enough info". If so, compute the value of $P(B|A)$.

(d) *(5 points)* Instead, I give you the following facts:

$$\begin{aligned} P(A|B) &= 2/3 \\ P(A|{\sim}B) &= 1/3 \\ P(B) &= 1/3 \\ P(A) &= 4/9 \end{aligned}$$

Do you now have enough information to compute $P(B|A)$? If not, write "not enough info". If so, compute the value of $P(B|A)$.

6 Drunk Squirrels (5 points)

- A drunk squirrel is dropped onto a 1-dimensional branch of an oak tree at location s. s is drawn from a Gaussian: $s \sim N(\mu_s = 0, \sigma_s^2 = 2^2)$.
- The squirrel makes a step. It moves to the right by distance d, where $d \sim N(0, 1)$. (If d is negative, it moves to the left of course). If we write f as the final location of the squirrel, we see $f \sim N(s, 1)$.
- d is independent of s.

The squirrel ends up at location location $f = 2$. What is the most likely location s that the squirrel landed on the branch initially?