15-781 Midterm, Fall 2001

YOUR ANDREW USERID IN CAPITAL LETTERS:

YOUR NAME:

Andrew Moore

Solutions created in haste. Email AWM@CS.CMU.EDU if/when I've made mistakes.

- 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

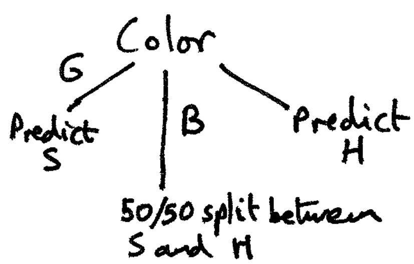
(a) *(2 points)* What is H(Emotion|Wig=Y)? 1

(b) *(2 points)* What is H(Emotion|Ears=3)? 0

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

Color (by inspection: it nicely breaks apart emotion, but the others hardly affect entropy in either branch)

(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.

$12\frac{1}{2}$ %

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?

50%

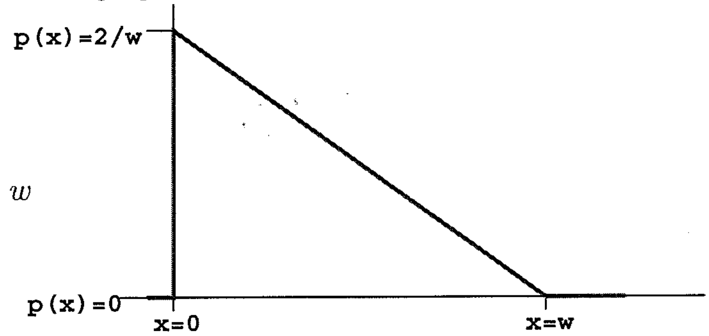
(g) *(3 points)* Construct an example dataset that achieves this maximum percentage training set error. (It must have two or fewer inputs and five or fewer records).

X_1	Y
0	0
0	1
1	0
1	1

2 Probability Density Functions (19 points)

Consider the PDF shown in the following figure and equations

$$p(x) = 0 \quad \text{if } x < 0$$
$$p(x) = \frac{2}{w} - \frac{2x}{w^2} \quad \text{if } 0 \le x \le w$$
$$p(x) = 0 \quad \text{if } w < x$$



(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 $P(x = 1|w = 2)$? 0 (no probabilty mass)

(c) *(4 points)* What is $p(x = 1|w = 2)$? $\frac{2}{2} - \frac{2}{4} = \frac{1}{2}$

(d) *(4 points)* What is $p(x = 0|w = 1)$? $\frac{2}{1} - \frac{0}{4} = 2$

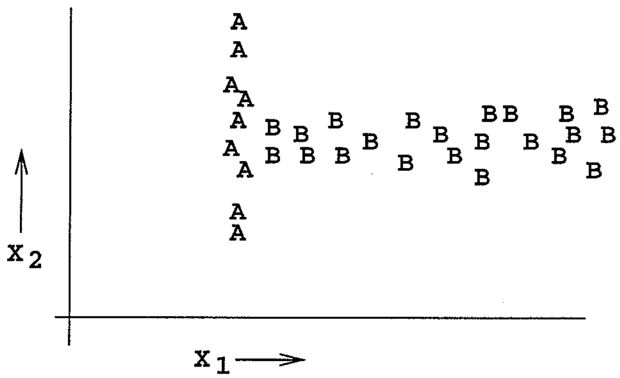
(e) *(4 points)* Suppose you don't know the value of w but you observe one sample from the distribution: $x = 3$. What is the maximum likelihood estimate of w?

$$p(x=3|w) = \frac{2}{w} - \frac{6}{w^2} \quad \text{if } w \ge 3$$
$$= 0 \quad \text{otherwise}$$

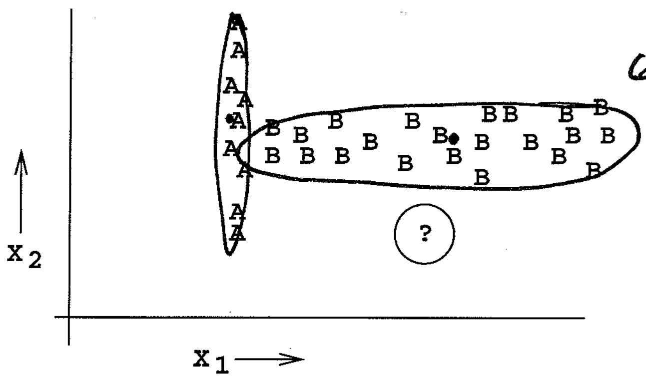
$$\text{MLE } w = \underset{w \ge 3}{\text{argmax}}\, P(x=3|w) = w \text{ s.t. } \frac{\partial}{\partial w} P(x=3|w) = 0$$

$$= w \text{ s.t. } \frac{-2}{w^2} + \frac{12}{w^3} = 0 = w \text{ s.t. } 2 = \frac{12}{w} = 6$$

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.



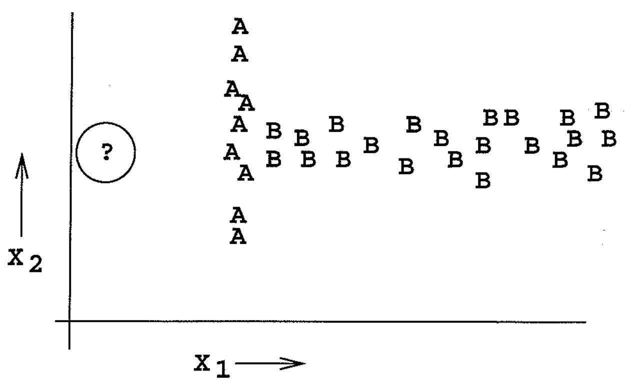
(f) *(3 points)* What class would the Bayes Classifier predict for inputs at the location shown by a question mark in the following figure? (Note: I strongly advise you to answer this by common sense and "eyeballing"—don't waste precious time calculating the Bayes Classifier.



Drawing guess at 2-sigma contours of classes

B: ~~this~~ ? is far closer in Mahal distance to B gaussian

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



B: Only about $3\sigma_{x_1}$ from center of B gaussian, but maybe $10\sigma_{x_1}$ from center of A gaussian.

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:

$\mu_A =$ 1	$\sigma_A^2 =$ 1	$P(Y = A) =$ 2/7
$\mu_B =$ 5	$\sigma_B^2 = \frac{4+1+0+1+4}{5} = 2$	$P(Y = B) =$ 5/7

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)?

$= P(X=2|Y=A)P(Y=A) = \frac{2}{7}\alpha$

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

$= P(X=2|Y=B)P(Y=B) = 5/7\,\beta$

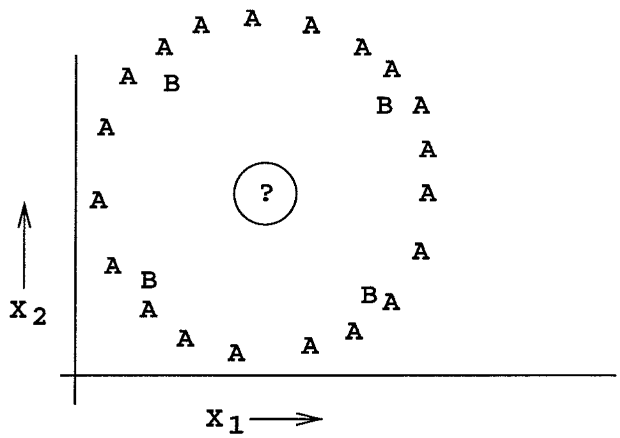
(d) *(2 points)* What is $p(X = 2)$ (answer in terms of α and β)?

$= \frac{1}{7}(2\alpha + 5\beta)$

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

$= \frac{P(Y=A \wedge X=2)}{P(X=2)} = \frac{\frac{2}{7}\alpha}{\frac{1}{7}(2\alpha+5\beta)} = \frac{2\alpha}{2\alpha+5\beta}$

(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?

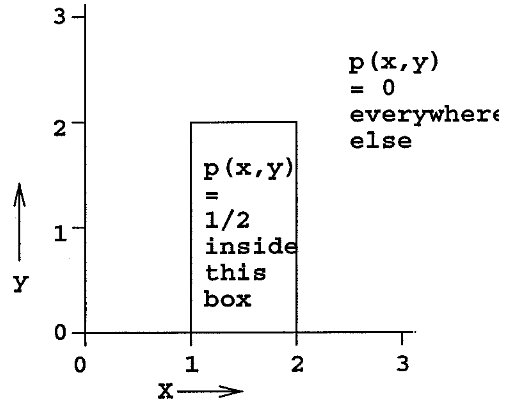


Both classes have same centroid, and A has slightly larger variance in each dimension, but the critical fact is that A's class prior is much higher, so we predict A

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)

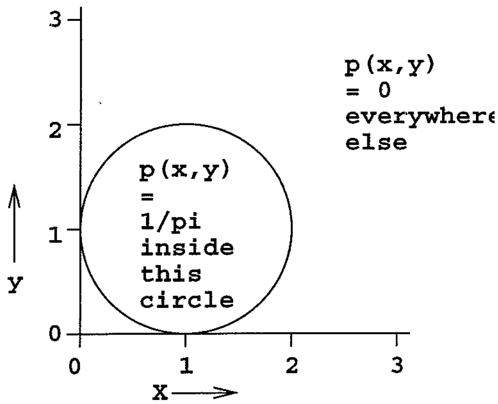
$$
p(x) = \begin{cases} 0 & \text{if } x < 1 \\ 1 & \text{if } 1 \leq x \leq 2 \\ 0 & \text{if } x > 2 \end{cases}
$$

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

YES

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? No. E.G. if $x = 0.01$ then I'm sure y is close to 1. But if $x=1$ then y could be anywhere between 0 and 2

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 $P(B|A)$.

Not Enough Info

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

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

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

Not enough Info

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

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

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

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|{\sim}B)P({\sim}B)} = \frac{2/3 \times 1/3}{2/3 \times \frac{1}{3} + \frac{1}{3} \times 2/3} = \frac{1}{2}$$

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

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

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

Still $\frac{1}{2}$ (of course)

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?

$$\begin{pmatrix} s \\ f \end{pmatrix} \sim N\left(\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \Sigma = \begin{pmatrix} 4 & 4 \\ 4 & 5 \end{pmatrix}\right)$$

$$\mu_{s|f} = \mu_s + \frac{\Sigma_{sf}}{\Sigma_{ff}}(f - \mu_f) = 0 + \frac{4}{5}(f-0) = \frac{4}{5}f$$

~~MAP~~ Most likely s given $f=2$

$$= \underset{s}{\operatorname{argmax}}\ P(s \mid f=2) = \mu_{s|f=2} = \frac{4}{5} \times 2 = 1.6$$