CS412 Machine Learning
HW 2 – Probabilities – Bayesian Learning
100pts

- **Please TYPE your answer or write legibly by hand (pts off if it is hard to read).**
- **Use this document to type in your answers** (rather than writing on a separate sheet of paper), so as to keep questions, answers and and grades together to facilitate grading.
- **SHOW all your work for partial/full credit.**
- **Allocated spaces should be enough for your answers** (unnecessarily long and irrelevant answers may loose points)

1) 20 pt - Suppose that we have 3 colored boxes r (red), b (blue) and g (green).
Box r contains 9 apples, 5 oranges and 3 limes;
Box b contains 8 apples, 4 oranges and 1 limes;
Box g contains 5 apples, 2 oranges and 6 limes.

Assume a process **where we pick a box first and then pick a fruit from the selected box**. A box is chosen at random according to the following probability of being selected: p(r) = p(b) = 0.3 and p(g)=0.4 and a piece of fruit is selected from the **chosen** box randomly.

a) 10 pt – What is the **probability of selecting an orange**?

P(Orange) = P(r) * P(O|r) + P(b) * P(O|b) + P(g) * P(O|g) =

$$= \frac{3}{10} * \frac{5}{17} + \frac{3}{10} * \frac{4}{13} + \frac{4}{10} * \frac{2}{13} \cong \mathbf{0.24}$$

b) 10pt - If we **observe that the selected fruit is an orange**, what is the probability that it came from the red box?

$$P(r|O) = \frac{P(O|r) * P(r)}{P(O)} = \frac{\frac{5}{17} * \frac{3}{10}}{0.24} \cong \mathbf{0.36}$$

2) 40 pt - For a 2-dimensional input space, we are given the following class conditional probability densities. Assume that x_1 and x_2 are conditionally independent given class names.

$p(x_1|C_1) = 1/3$ for $0 <= x_1 <= 4$
0 elsewhere

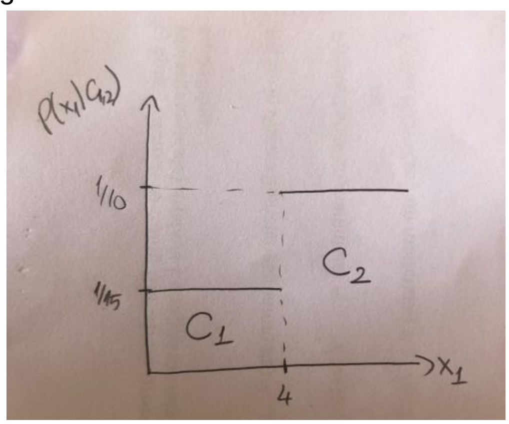
$p(x_2|C_1) = 1/4$ for $0 <= x_2 <= 4$
0 elsewhere

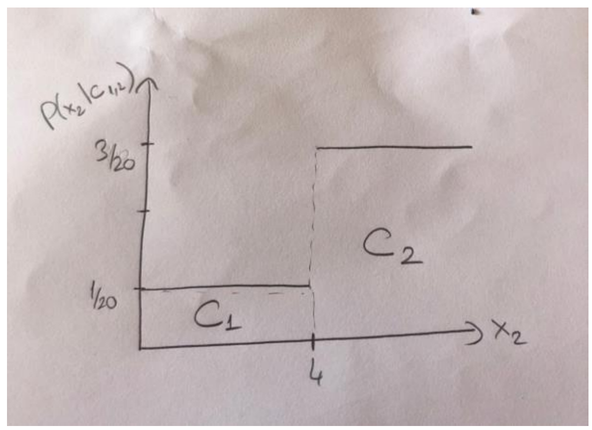
$p(x_1|C_2) = 1/2$ for $4 < x_1$
0 elsewhere

$p(x_2|C_2) = 3/4$ for $4 < x_2$
0 elsewhere

Assume $P(C_1)=P(C_2)=0.2$ and $P(C_3)=0.6$.

a) 12pt – Draw the corresponding pdfs for x_1 and x_2 , being as precise as possible (e.g. **label axes and important points on the axes**). You can draw by hand, take a picture and include here as image.





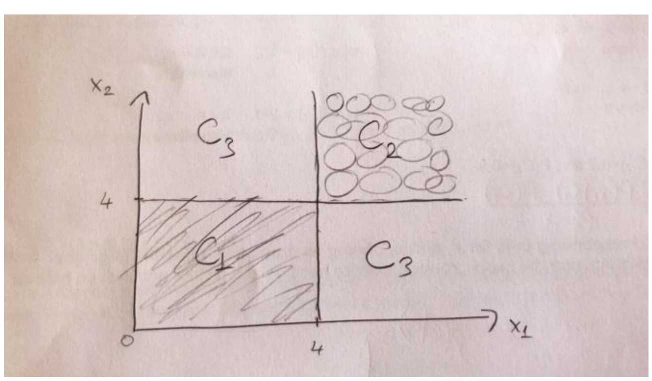
b) 10pts - Develop a classification strategy for given feature values (just looking at the graph – no formula), just complete the sentence(s):

if x_1 is in the region **$0 \leq x_1 \leq 4$**, and x_2 is in the region **$0 \leq x_2 \leq 4$**, I will classify it as **C1**;

if x_1 is in the region **$4 \leq x_1$**, and x_2 is in the region **$4 \leq x_2$** , I will classify it as **C2**;

Otherwise, I will classify it as **C3**.

c) 8pts - Draw the decision regions.



d) 10pts – Give a one line qualitative answer (no precise numbers/thresholds...) **& reasoning for each case below (**how your decision changes or whether it doesn't).

- Would your decision strategy change if $P(C_1)=0.8$ and $P(C_2)=P(C_3)=0.1$?

My decision strategy wouldn't change since the decision areas won't intersect. Therefore, only the posterior probabilities would change. For example, I would still choose C1 when $0 \leq x_1 \leq 4$ and $0 \leq x_2 \leq 4$.

- How about if it was the reverse $P(C_1)=P(C_3)=0.1$ and $P(C_2)=0.8$?

Again my strategy wouldn't change. The posterior probabilities would change reversely this time.

3) 40pts – NAIVE BAYES

a) 10pts – Given that two random variables X and Y are conditionally independent given C, circle True or False (2pts for each correct answer; -1pts each wrong answer):

- $P(X|Y) = P(X)$ True / **False** → $P(X|Y) = P(X|C)$
- $P(X | Y, C) = P(X | Y)$ True / **False** → $P(X|Y,C) = P(X|C)$
- $P(X, C | Y) = P(X | Y)$ True / **False** → $P(X,C|Y) = P(X|Y) * P(C|Y)$
- $P(X, Y | C) = P(X | C) P(Y | C)$ **True** / False
- $P(X, Y, C) = P(X | C) P(Y | C) P(C)$ **True** / False

b) 24pts - Using the Mammal dataset given below, **how would you classify the animal that give birth, cannot fly, sometimes live in water, and has not legs, using Naive Bayes classifier *without any smoothing*. Show your work** (e.g. indicate class conditional attribute probabilities under the given table in the next page and just transfer them here).

$$P(\text{mammal}|\ x) = \frac{P(x|mammal)*P(mammal)}{P(x)} = \frac{\left(\frac{6}{7}*\frac{6}{7}*\frac{0}{7}*\frac{2}{7}\right)*\frac{7}{20}}{\frac{7}{20}*\frac{16}{20}*\frac{4}{20}*\frac{6}{20}} = 0$$

$$P(\text{non-mammals}|\ x) = \frac{P(x|nonmammal)*P(nonmammal)}{P(x)} = \frac{\left(\frac{1}{13}*\frac{10}{13}*\frac{4}{13}*\frac{4}{13}\right)*\frac{13}{20}}{\frac{7}{20}*\frac{16}{20}*\frac{4}{20}*\frac{6}{20}} = 0.216$$

Decision: Non-Mammal

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Write here the estimated probabilities (you should only write those related to the question for simplicity):

P(Give Birth=Yes | mammal) = 6/7

P(Give Birth=Yes | non-mammal) = 1/13

P(Can Fly=No | mammal) = 6/7

P(Can Fly=No | non-mammal) = 10/13

P(Live in Water =Sometimes | mammal) = 0/7

P(Live in Water =Sometimes | non-mammal) = 4/13

P(Have Legs=No | mammal) = 2/7

P(Have Legs=No | non-mammal) = 4/13

P(Give Birt=Yes) = 7/20

P(Can Fly=No) = 16/20

P(Live in Water=Sometimes) = 4/20

P(Have Legs=No) = 6/20

P(mammal) = 7/20

P(non-mammal) = 13/20

P(x) = P(Give Birth=Yes) * P(Can Fly=No) * P(Live in Water=Sometimes) * P(Have Legs=No)

c) 6pts - Without re-doing the whole process, calculate the probabilities for P(Live in Water|mammals), P(Have Legs|mammals) and P(Give Birth|non-mammals) using Laplace smoothing:

P(Live in Water=Sometimes|mammals) = **1/10**

P(Have Legs=No|mammals) = **3/9**

P(Give Birth=Yes|non-mammals) = **2/15**