**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.

1. 10 pt – What is the **probability of selecting an orange**?

**(0.3\*5/17)+( 0.3\*4/13)+(0.4\*2/13) = 0.242081448**

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

P(X=red box | Y=orange)= P( Y=orange|X=red box ) \* P(X=red box) / P(Y=orange)

= **(5/17) \* 0.3 / 0.242081448 = 0,364485981253909**

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

p(|C1) = 1/3 for 0 <= <= 4

0 elsewhere

p(|C1) = 1/4 for 0 <= <= 4

p(|C2) = 1/2 for 4 <

0 elsewhere

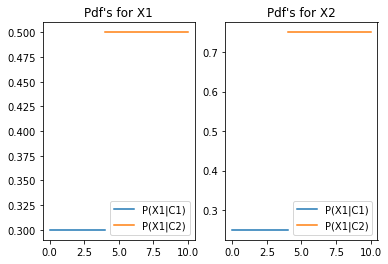
0 elsewhere

p(|C2) = 3/4 for 4 <

0 elsewhere

Assume P(C1)=P(C2)=0.2 and P(C3)=0.6.

1. **12pt – Draw the corresponding pdfs for and**  **, 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.



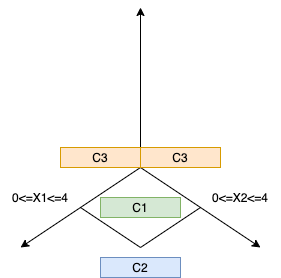
1. **10pts -** **Develop a classification strategy for given feature values** (just looking at the graph – no formula), just complete the sentence(s):

if is in the region ..0,4.., and is in the region ..0,4.. , I will classify it as ..C1..;

if is in the region ..4<.., and is in the region ..4<.. , 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(C1)=0.8 and P(C2)=P(C3)=0.1?

Decisions would not be affected since there is no change in the decision intervals of X1 and X2.The intervals are not superpositioned,so change in the probabilities of classes does not affect.

* How about if it was the reverse P(C1)=P(C3)=0.1 and P(C2)=0.8?

Decisions would not be affected since there is no change in the decision intervals of X1 and X2.The intervals are not superpositioned,so change in the probabilities of classes does not affect.

**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, C ) = P(X | Y) True / **False**
* P(X , C | Y ) = P(X | Y ) True / **False**
* 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 P(C)>0

**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(mammal| x) = 0

P(mammal) = 7/20

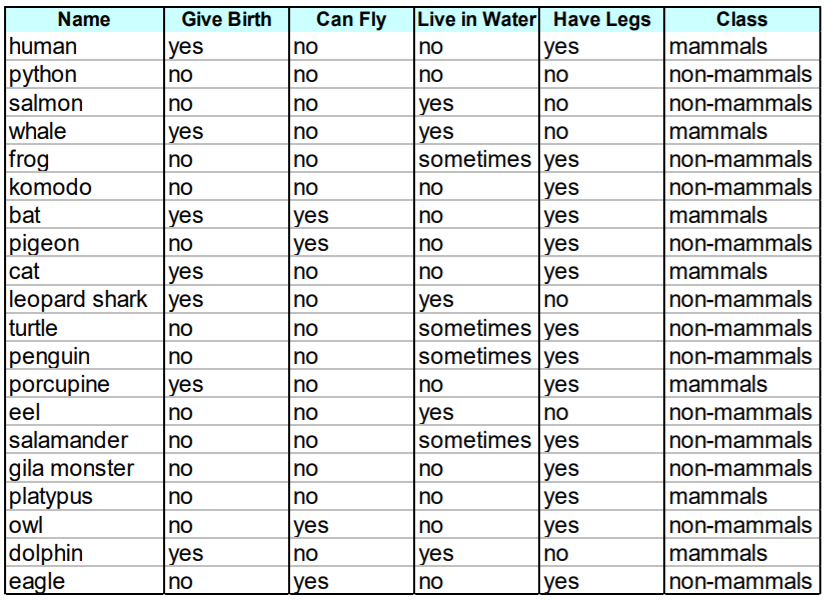
6/7 \* 0/7 \* 6/7 \* 2/7 = 0

P(non-mammals| x) =

P(non-mammal) = 13/20

4/13 \* 4/13 \* 10/13 \* 1/13 \* 13/20 ~= 0.003

**Decision: non-mammal**



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

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

**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) = (0 + 1) / (7+2)

P(Have Legs=No|mammals) = (2+1)/(7+2)

P(Give Birth=Yes|non-mammals) = (1+1)/(13+2)