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 8 apples, 1 oranges and 1 limes;
Box b contains 5 apples, 5 oranges and 0 limes;
Box g contains 3 apples, 3 oranges and 4 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) = 0.2 and p(g) = 0.6 and a piece of fruit is selected from the **chosen** box randomly.

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

$P(\text{lime}) = 0.2 \times \frac{1}{10} + 0.2 \times \frac{0}{10} + 0.6 \times \frac{4}{10}$

$= \frac{0.2}{10} + \frac{0.6 \times 4}{10} = 0.26$ //

b) 10pt - If we **observe that the selected fruit is a lime**, what is the probability that it came from the green box?

$P(\text{Green} \mid \text{LIME}) = \frac{P(\text{LIME} \mid g) * P(g)}{P(\text{LIME})} = \frac{0.4 \times 0.6}{0.26} = 0.923$ //

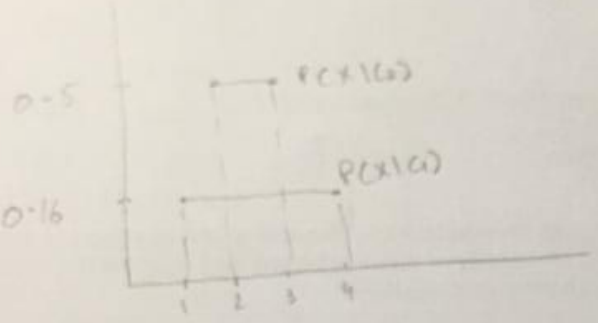
2) 40 pt - For a 1-dimensional input space, assume we are given the following class conditional probability densities as follows:

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

$p(x|C_2) = 1$ for $2 <= x <= 3$
0 elsewhere

Assume $P(C_1)=P(C_2)=0.5$

a) 15pt – Draw the corresponding pdfs, 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 a given x (just looking at the graph – no formula), just complete the sentence(s):

if x is in the region ..[1,2] ∪ [3,4].. , I will classify it as ..C1..;

if x is in the region ..[2,3].. , I will classify it as ..C2..;

c) 5pts - Draw the decision regions on the above figure.

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 $P(C_1)=0.9$ and $P(C_2)=0.1$?

  Yes, because in interval [2,3] C1 would dominate C2.

- How about if it was the reverse $P(C_1)=0.1$ and $P(C_2)=0.9$?

  Decision Strategy wouldn't change because in same intervals same attributes will still dominate.

3) 40pts – NAIVE BAYES

a) 15pts – 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)

b) 20pts - Using the PlayTennis data given below (and in the lecture slides), **how would you classify x=(Overcast, Mild, Normal, Strong), 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(Yes| x) = $P(x|Yes) * P(Yes) = (4/9 \times 4/9 \times 6/9 \times 3/9) \times (9/14) = 0.282$ //

P(No| x) = $P(x|No) * P(No) = (2/5 \times 0 \times 1/5 \times 3/5) \times (5/14) = 0$ //

Decision: .....YES.....

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

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

P(Outlook=Overcast | Yes) = 4/9.

P(Outlook = Overcast | No) = 0

P(Humidity = Normal | Yes) = 6/9

P(Humidity = Normal | No) = 1/5

P(Temperature = Mild | Yes) = 4/9

P(Temperature = Mild | No) = 2/5

P(Wind = Strong | Yes) = 3/9

P(Wind = Strong | No) = 3/5

c) 5pts - Without re-doing the whole process, give the for P(Overcast|No) and P(Temperature|Yes) using Laplace smoothing:

P(Outlook=Overcast|No) = ...1/8.....

P(Humidity=Normal|Yes) = ...9/11.....