1) (7 points) Suppose that a query $x$ to be classified will come from class $i$ with probability $\pi_i$, where

$$\pi_1 = \frac{1}{2}, \quad \pi_2 = \frac{1}{3}, \quad \pi_3 = \frac{1}{6},$$

and that if $x$ comes from class $i$ it will be an observation from a binomial $(4, p_i)$ random variable, where $p_i = i/10$. (So if $x$ is a class 2 observation, which will be the case with probability $\pi_2 = 1/3$, then $x$ is the value resulting from a binomial $(4, 2/10)$ random variable.) Determine the Bayes classifier, expressing it as a piecewise function having the form

$$C(x) = \begin{cases} 
1, & x \in S_1, \\
2, & x \in S_2, \\
3, & x \in S_3,
\end{cases}$$

where each $S_i$ is a subset of $\{0, 1, 2, 3, 4\}$. Also give adequate supporting work to justify your answer. (Hint: Use the result given 7 lines from the bottom on p. 2-21 of the class notes.)

2) The purpose of this exercise is to make sure that you’ve started learning about R by reading Sec. 2.3 of the text, and trying at least some of the things introduced there using the software. For this exercise, just turn in answers for parts (b) and (c) as well as the plot requested in part (d) (and don’t submit anything for part (a)). You don’t have to submit any R output to justify your answers, but you should use R to obtain the answers yourself (asking for help as needed), and you should not simply copy answers obtained by someone else.)

(a) (0 points) Within R, enter ?USArrests to obtain information about the USArrests data set that is part of the base R distribution (and so there is no need to use the library() function).

(b) (1 point) Suppose that we want to cluster the cases using all of the variables. What are the values of $n$ and $p$ (using the text’s notation). (You can answer this by looking at the data and/or using the dim() function.) (Note: I’m not requesting that you actually do any clustering. (That may occur in a future assignment.))

(c) (1 point) Considering the Murder variable, what value is obtained if we subtract the smallest number of murder arrests (per 100,000) from the largest number of murder arrests? (You can answer this using the range() function, the summary() function, or the min() and max() functions.)

(d) (1 point) Use R to plot Assault against UrbanPop. Have it label the horizontal axis Percent Urban Population, label the vertical axis Assault Arrests (per 100,000), and give the graph the title Data For 50 States. (See the very bottom of p. 45 of the text for an example of creating and labeling a plot.) Print this plot, and turn in it. (Although you could use the information on the top portion of p. 46 to create a pdf, on my desktop computer (with a printer attached) I could simply create the plot using the plot() function, left click of the displayed plot, and select Print).