CIS 635 Data Mining Name Sajal

Homework 4

Hand in sheet

Write in or copy and paste the answers for the following:

Part 1

		virus = n	virus = y
1.	priors	6/10	4/10
2.	likelihoods		
a)	gen = F	4/6	2/4
	gen = M	2/6	2/4
b)	group = 1	0/6	3/4
	group = 2	2/6	1/4
	group = 3	4/6	0/4
3.	likelihoods (numeric)		
a)	pulse mean	71.833	63.750
	pulse sd	4.956	9.945
b)	test mean	66.166	45.750
	test sd	5.913	5.377

Part 2

1. Less likely to have the virus

2. Male

3. categorize (fill in the numbers for virus=n and virus=y)

	numerator for n	numerator for y	class label
a)	1.923811e-05	3.680086e-05	p(virus=? | gender=f, ageGroup=2, pulse=66, test=50)
b)	2.540663e-07	6.958789e-07	p(virus=? | gender=m, ageGroup=1, pulse=90, test=81)
c)	7.510184e-08	2.108054e-05	p(virus=? | gender=m, ageGroup=3, pulse=90, test=55)

4. Lower the pulse rate.

Part 3

1. attach the code for the analysis you did in parts a through d

```
     # read the data
df = read.table("hw03dataSenate.txt", header = TRUE)

# make pol, law, mil, soc, edu, med, bus, agr, fin and approval, factors
colFactNames = c("pol", "law", "mil", "soc", "edu", "med", "bus", "agr",
"fin", "approval")
for (col in colnames(df)) {
    if(!is.element(col, colFactNames)) {
        next
    }
    df[,col] = as.factor(df[,col])
}

# build a model using approval as the class and the rest of the attributes
model = naiveBayes(approval~., df)

# display the model
print(model)
```

2. answers for part e

e) agr pol