1) submit an expression for p(x,y) using the terms mentioned in the pdf.

the idea is: for construction of the data set we used P(C=1) P(C=2)  f(x|C=1) f(x|C=2) p(y=0| x,C=1) p(y=0| x,C=2)

so how do these terms contribute to how p(x,y) looks like ?

suggestion: submit the derivation, too, so that you do not lose all points if you got something wrong.

2) no need to submit any coding - this time. i will show a solution monday in class.

3) many students confuse C and y.

what are these things about ?

we want to predict y based on x

I give another example:

y =1 treatment success

y=0 treatment failure

x= (a vector of liver measurements) = (albumin,globulin, SGOT)

you can imagine, that if you would plot the liver measurements of many patients, you would see groups of patients with very similar values grouped together - clusters of x values

and C(x) just serves to create data with such a cluster structure!

you can imagine that for every cluster p(y=1 | C ) ... the probability of treatment success given in what cluster x is, can be different across clusters

we modeled that general idea by having p(y=0|x,C=1) being different from p(y=0 | x, C=2 )

Best, Alex