

## Homework 3

*Due: 25 Sep 2007*

### Problem 3.1

Some online sites try to decide whether you're more likely to like or to dislike a song, before you even listen to it. They can take into account various types of information, like the opinions of professional critics or your friends who have already heard it, as well as how similar it is to music you already have an opinion on. Consider the following example:

The new album *The things I want to say to you* by the band Anapestic Trochee got a neutral review in TKS, but your friends Chris and Loren didn't like it. You've also heard AT's *All of the times I imagine your happiness*, and enjoyed it.

Will you like *The things I want to say to you*?

The site has information about how frequently you've agreed with TKS, Chris, and Loren in the past, and how often you've liked multiple albums by the same band. Express using the notation of conditional probability how you would build a naïve Bayes classifier to determine from this data whether it's more probable that you will like, dislike, or be ambivalent about the book.

Identify the independence assumptions being made in the model, and comment on whether they seem reasonable to you.

### Problem 3.2

Linked from the course website is an article from a few weeks ago about some new recommendations on lung cancer treatment. ("More May Be Offered Lung Cancer Surgery") Read it with an eye to probabilities, especially conditional probabilities; and write down at least six probabilities you can learn from reading the articles, clearly specifying using probability notation what the conditionals are.

Here's one for free: the probability of getting diagnosed during the first two stages, given that you're diagnosed at all, is 20%:  $p(E|D) = .2$  ( $E$  = "early",

i.e. stage 1 or 2, diagnosis;  $D =$  receives a lung cancer diagnosis during their lifetime).

Can you combine any of these probabilities? If so, what additional non-explicitly-written facts can you learn from them? If not, what prevents them from being combined?