Assignment 1: Statistical Foundations

Unsupervised Learning

Zoubin Ghahramani

Due: Thursday 9 Oct, 2003

Note: all assignments for this course are to be handed in to the Gatsby Unit, **not** to the CS department. Please hand in all assignments at the beginning of lecture on the due date to Zoubin. Late assignments will be penalised. If you are unable to come to class, you can also hand in assignments to Alexandra Boss, Room 408, Gatsby Unit.

1. Read "Nuances of Probability Theory" by Tom Minka:
   `http://www.stat.cmu.edu/~minka/papers/nuances.html` . Pick one of the topics and write a short paragraph discussing it (do you agree or disagree, can you think of another example, etc).

2. Read the preface and chapters 1 and 2 of E.T Jaynes *Probability Theory: The Logic of Science*, available on the web at
   `http://omega.math.albany.edu:8008/JaynesBook.html` Pay particular attention to and try to understand: the desiderata (p 112-114) and the derivations of the sum and product rules. Write a short paragraph paraphrasing the arguments in your own words.

3. You will need to be familiar with the following terms from statistics. Find definitions and write them down in your own words. Make sure you mention your source (where you got the definition) and make sure you understand the definition.

   expectation value, unbiased, sufficient statistics, exponential family

   In the coming weeks we will be making extensive use of the following distributions, which you should know. Find definitions and write them down:

   Binomial, Multinomial, Beta, Dirichlet, Gaussian, Gamma

4. Assume you have a data set of independent and identically distributed binary vectors $Y = \{\mathbf{y_1}, \ldots, \mathbf{y_N}\}$ each of which is D-dimensional. Describe a simple statistical model for your data. What is the expression for the likelihood of the data given the parameters of your model? Write down the equations for how you would estimate the parameters of your model from the data.

Bonus In an example as the above, how would you calculate the (relative) probability of the three different models:

(a) all D components are generated from Bernoulli $q = 0.5$

(b) all D components are generated from Bernoulli with unknown (but identical) q

(c) each component is Bernoulli with seperate, unknown q_d