Department of Mathematics and Computer Science
University of Southern Denmark, Odense

February 27, 2013

Marco Chiarandini

DM825 - Introduction to Machine Learning

Sheet 11, Spring 2013

Exercise 1 – Probability theory

Prove the following rule:

$$p(x_i|x_{-i}) = \frac{p(x_1, \ldots, x_N)}{\int p(x_1, \ldots, x_N) dx_i}$$

where $x_{-i} = \{x_1, \ldots, x_N\} \setminus x_i$.

Exercise 2 – Naive Bayes

Consider the binary classification problem of spam email in which a binary label $Y \in \{0, 1\}$ is to be predicted from a feature vector $X = (X_1, X_2, \ldots, X_n)$, where $X_i = 1$ if the word i is present in the email and 0 otherwise. Consider a naive Bayes model, in which the components X_i are assumed mutually conditionally independent given the class label Y.

a Draw a directed graphical model corresponding to the naive Bayes model.

b Find a mathematical expression for the posterior class probability $p(Y = 1|x)$, in terms of the prior class probability $p(Y = 1)$ and the class-conditional densities $p(x_i|y)$.

c Make now explicit the hyperparameters of the Bernoulli distributions for Y and X_i. Call them, μ and θ_i, respectively. Assume a beta distribution for the prior of these hyperparameters and show how to learn the hyperparameters from a set of training data $(y^j, \vec{x}^j)_{j=1}^m$ using a Bayesian approach. Compare this solution with the one developed in class via maximum likelihood.

Exercise 3 – Directed Graphical Models

Consider the graph in Figure left.

- Write down the standard factorization for the given graph.
- For what pairs (i, j) does the statement X_i is independent of X_j hold? (Don't assume any conditioning in this part.)
- Suppose that we condition on $\{X_2, X_9\}$, shown shaded in the graph. What is the largest set A for which the statement X_1 is conditionally independent of X_A given $\{X_2, X_9\}$ holds?
- What is the largest set B for which X_8 is conditionally independent of X_B given $\{X_2, X_9\}$ holds?
- Suppose that I wanted to draw a sample from the marginal distribution $p(x_5) = \Pr[X_5 = x_5]$. (Don't assume that X_2 and X_9 are observed.) Describe an efficient algorithm to do so without actually computing the marginal.

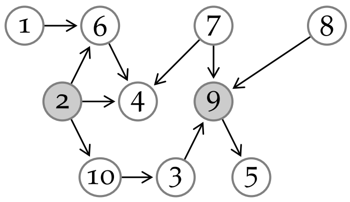

Figure 1: A directed graph.