ENGN 2520 / CSCI 1950-F Midterm

IMPORTANT:

No collaboration is allowed on the midterm. You should not talk to anyone about the contents of the midterm until after you turn it in. If you need any clarifications you can send email to the TA list.

Problem 1

Suppose we have an input space $X = \mathbb{R}$ and a label space $Y = \{0, 1\}$. Let $p(x, y)$ be a probability density defined as follows.

$$p(y = 0) = \frac{1}{3}$$

$$p(y = 1) = \frac{2}{3}$$

$p(x|y = 0)$ is uniform over $[5, 7]$ and zero outside the interval.

$p(x|y = 1)$ is uniform over $[1, 20]$ and zero outside the interval.

(a) Compute $p(y = 1|x = 6)$ and $p(y = 0|x = 6)$.

(b) What is the Bayes optimal classifier for this example? How does it divide the input space into classes?

Problem 2

Consider a classification problem with two classes C_1 and C_2. The examples to be classified are vectors in $\mathbb{R}^D$. Suppose we model $p(x|C_i)$ with a Gaussian distribution with class specific mean $u_i \in \mathbb{R}^D$ and class independent covariance αI where $\alpha \in \mathbb{R}$.

(a) Derive the maximum likelihood estimate for the parameters u_1, u_2 and α given training data $\{(x_1, y_1), \ldots, (x_n, y_n)\}$.

(b) Suppose we classify an example x by selecting the class with maximum posterior probability $p(C_i|x)$. Show that the decision boundary of the resulting classifier is a hyperplane in $\mathbb{R}^D$. Points in one side of the hyperplane are classified as C_1 and points in the other side are classified as C_2.

(c) We say a dataset is separated by a hyperplane H if the examples from C_1 are in one side of H and examples from C_2 are in the other side of H. Describe a training set that can be separated by a hyperplane H but the decision boundary defined by the ML estimation procedure from above does not separate the data.

Problem 3

Suppose we have 4 binary random variables CLOUDY, RAIN, SPRINKLER, WETGRASS.

CLOUDY=True if the sky is cloudy today.

RAIN=True if it rains today.

SPRINKLER=True if the sprinkler comes on today.

WETGRASS=True if the grass is wet today.

(a) Suppose the sprinkler controller has no sensors. Draw a Bayesian network that models the natural relationships between these random variables.

(b) How many total parameters do you need to specify to define the necessary probability distributions associated with the bayesian network?

(c) How many parameters would you need to specify an arbitrary joint distribution over the random variables?

(d) What is the Markov blanket for SPRINKLER?

Problem 4

Let H be a hidden Markov model with state space S and observation space O. Suppose we are given a sequence of observations $(y_1, \ldots, y_n)$ and we would like to find the MAP estimate of the hidden states $(x_1, \ldots, x_n)$. The Viterbi algorithm can be used to compute the MAP estimate in $O(nk^2)$ time where $k = |S|$.

Suppose the transition matrix M has the following special structure. $M(i,i) = a$ and $M(i,j) = b$ when $j \neq i$. Suppose $b < a$. Show how we can modify the Viterbi algorithm to run in $O(nk)$ time in this case.