ENGN 2520 Final exam
Due Thursday May 10 by 4pm

Problem 1

Consider a classification problem in $\mathbb{R}^D$ with two classes C_1 and C_2. Suppose we model $p(x|C_i)$ with a Gaussian distribution with class dependent mean u_i and class independent covariance Σ. The covariance is shared between both classes but is not necessarily diagonal.

(a) Derive the maximum likelihood estimate for the parameters u_1, u_2, Σ given two sets of independent examples, $S_1 \subset \mathbb{R}^D$ from C_1 and $S_2 \subset \mathbb{R}^D$ from C_2.

(b) Suppose we classify an example x by selecting the class C_i with maximum posterior probability $p(C_i|x)$. Show that the decision boundary of the resulting classifier is a hyperplane in $\mathbb{R}^D$.

Problem 2

Let H denote the set of ellipses in the plane. Each ellipse defines a binary classifier that assigns label $+1$ to points inside the shape and label -1 to points outside the shape.

What is the VC dimension of H? Justify your answer.

Problem 3

Consider the naive Bayes classifier for images of digits in Homework 2. Let c be a random variable specifying the class of an image, and x_i be a random variable specifying the value of the i-th pixel.

(a) In the naive Bayes model we assume the pixel values are independent conditional on the image class. Draw a bayesian network corresponding to the naive Bayes model.

(b) To model dependencies between pixels we could augment the model with a random variable s specifying a "style" for the digit. Here we assume s takes one of K possible values. For example, the value of s when the class is "7" could indicate if we have an "european" or an "american" 7. Now suppose we assume the pixel values are independent conditional on the image class and style. Draw a bayesian network corresponding to this model.

Problem 4

Consider a weighted regression problem defined as follows.

Let $\phi : \mathbb{R}^D \to \mathbb{R}^{\mathbb{K}}$ be a feature map.

We have a set of training examples $\{(x_1, y_1), \ldots, (x_N, y_N)\}$, and weights $a_1, \ldots, a_N$ with $a_i > 0$. We would like to select a weight vector w such that $w^T \phi(x_i) \approx y_i$. The weights a_i specify the importance of the i-th training example, leading to the following objective

$$E(w) = \frac{1}{2} \sum_{i=1}^{N} a_i (w^T \phi(x_i) - y_i)^2.$$

Find an expression for the solution w^* that minimizes $E(w)$.

Note that when all $r_i = 1$ this is equivalent to least squares regression (section 3.1.1 in the textbook).