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

You should justify all your answers.

Problem 1

Consider the naive Bayes classifier for images of digits in Homework 3. Let $c$ be a random variable specifying the class of an image, and $x_1, \ldots, x_M$ be random variables specifying the value of the different pixels in an image with $M$ pixels.

(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. The bayesian network should include $c$ and $x_1, \ldots, x_M$.

(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. The bayesian network should include $c$, $s$ and $x_1, \ldots, x_M$.

Problem 2

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 $M$ of the HMM has the following special structure: $M(i, i) = a$ and $M(i, j) = b$ when $j \neq i$ for two parameters $a > b$. Show how we can modify the Viterbi algorithm to run in $O(nk)$ time in this case.
Problem 3

In this problem we will consider a 1-nearest neighbor classifier. Suppose we have two classes $C_1$ and $C_2$ with equal probabilities $p(C_1) = p(C_2) = 0.5$. Suppose the conditional probability density on an observable feature $x$ is given by

\[
p(x|C_1) = \begin{cases} 
2x & 0 \leq x \leq 1 \\
0 & \text{otherwise} 
\end{cases} \tag{1}
\]

\[
p(x|C_2) = \begin{cases} 
2(1-x) & 0 \leq x \leq 1 \\
0 & \text{otherwise} 
\end{cases} \tag{2}
\]

(a) Derive the Bayes optimal classifier for an example $x$ and the probability this classifier makes a mistake on a random example from $p(x)$. You should use a 0/1 loss when deriving the Bayes optimal classifier (loss of 0 if the prediction is correct, and loss of 1 if the prediction is incorrect).

(b) Suppose we have a single training sample from each class. That is, we have a sample $x_1$ from $p(x|C_1)$ and a sample $x_2$ from $p(x|C_2)$. Now suppose we classify a random test sample from $p(x)$ using a 1-nearest neighbor classifier. What is the probability that this classifier will classify the test sample incorrectly.

Problem 4

Suppose you have an instrument that measures the speed of a running animal. You install the instrument in a habitat that contains two kinds of animals: tigers and cheetahs. The speed of a tiger is a random variable with probability density $p_1(x)$. The speed of a cheetah is a random variable with probability density $p_2(x)$. Suppose $p_1$ and $p_2$ are known. These densities are different but they have significant overlap, so it is hard to decide if an animal is a tiger or a cheetah based on a single observation of its speed.

(a) Let $x$ be the speed of a random animal in the habitat. Let $w_1$ be the fraction of animals in the habitat that are tigers, and $w_2$ be the fraction of animals in the habitat that are cheetahs (since there are no other animals in this habitat we have $w_1 + w_2 = 1$). What is the probability density of $x$?

(b) Suppose we record the speed of $n$ animals $x_1, \ldots, x_n$. Assume each measurement comes from a random animal in the habitat. How can we estimate $w_1$ and $w_2$ from these measurements? Derive an algorithm for estimating $w_1$ and $w_2$. Justify your answer.