## Homework 3 CS4100/5100, Fall 2014

Your Name:

1. In a support vector machine, the kernel function is a dot product over feature transforms:  $K(x, z) = F(x)^T F(z)$ . Suppose x, z are two dimensional with  $x = (x_1, x_2)^T$  and  $z = (z_1, z_2)^T$ .

a) Suppose  $F(x) = (x_1^2, x_1 x_2, x_2^2)^T$ . What is K(x, z)?

b) Suppose that  $K(x, z) = x^T z + x_1 x_2 z_1 z_2$ . What is F(x)?

2. Consider the Bayes net below.



a) Calculate P(C|D) in terms of the following: P(A), P(B|A), P(C|B), P(D|B).

b) Suppose we are using variable elimination to calculate P(C|D). Which variables get eliminated? How many different ways are there to order these variables? What is the size of the largest factor prior to marginalizing (size of the factor after executing the JOIN operation but before executing the SUM operation)?

3. Consider a binary perceptron with the weight vector, (1, 1, 0, -1).

a) How would this perceptron classify the following two test cases: (2, 0, 1, 1) and (-1, 0, 1, 1)?

b) Suppose the following additional training example is provided: ((0, 0, 0, -1), -1). How does the perceptron update rule change the weight vector is the learning rate  $(\alpha)$  is 0.1?

4. (AIMA 18.23) Suppose that a training set contains only a single example, repeated 100 times. In 80 of the 100 cases, the single output value is 1; in the other 20, it is 0. What will logistic regression predict for this example, assuming that it has been trained and reaches a global optimum? (Hint: to find the global optimum, differentiate the error function and set it to zero.)

5. Consider training the Naive Bayes model and the training data shown below.



$F_1$	1	0	0	0	0	1	1	1
$F_2$	0	1	1	1	0	0	1	1
$F_3$	1	0	1	0	1	1	1	0
Y	0	0	1	0	1	0	0	1

- a) Calculate $P(F_1=1 \mid Y=0)$  using maximum likelihood estimates.
- b) Calculate  $P(F_1 = 1 | Y = 0)$  using laplace smoothing.