Quiz 2
November 30, 2015

Time: 50 minutes; Total Marks: 43
One double-sided 8.5" x 11" cheat sheet allowed

This test contains 4 questions and 6 pages

NAME: _____

STUDENT NUMBER: _____

Question	Marks	Time budget
1	/16	15 min
2	/8	10 min
3	/11	10 min
4	/8	10 min

1. (16 marks) True or False questions. **No explanation required.**

   (a) True or False. Back propagation can compute the partial derivatives of the error on one example with respect to all W weights in a network in $O(W)$ time.

   (b) True or False. Gradient descent for training a neural network with 2 hidden layers is guaranteed to find the network parameters with lowest training error.

   (c) True or False. Stochastic gradient descent for training a neural network with 2 hidden layers is guaranteed to find the network parameters with lowest training error.

   (d) True or False. The 2 Bayesian networks below have the same conditional independence assumptions.

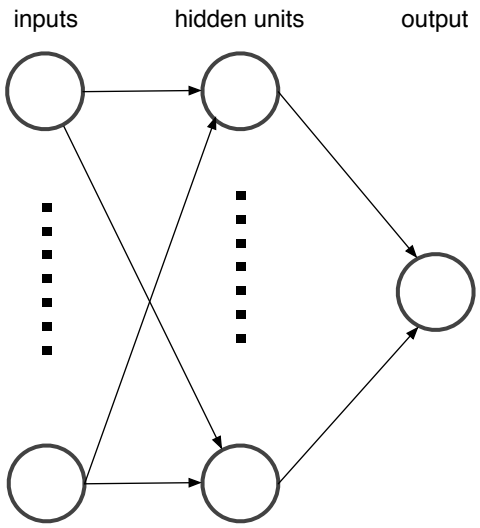
   (e) True or False. If A is a discrete random variable, then $P(B|C) = \sum_A P(A, B|C)$.

   (f) True or False. Any potential function defined over 3 variables can be written as a product of potential functions over pairs of variables.
   I.e. $\psi_{ABC}(A, B, C)$ can be written as $\psi_{AB}(A, B) \cdot \psi_{BC}(B, C) \cdot \psi_{AC}(A, C)$.

   (g) True or False. In a Bayesian network, when learning the parameters for $P(X_i|parents(X_i))$, it is sufficient to only consider the values of X_i and the parents of X_i in the training data.

   (h) True or False. The K-means algorithm is guaranteed to converge.

2. (8 marks) Consider the neural network below.

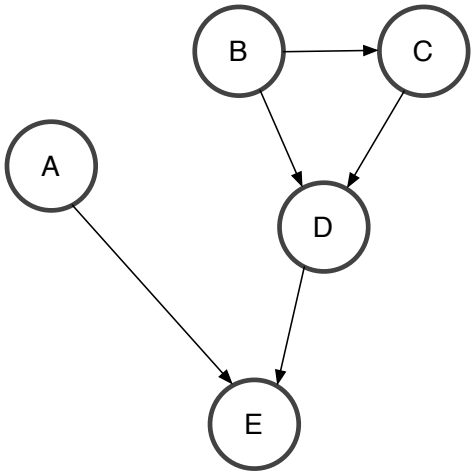


Suppose the hidden units use **linear** activation functions, $h(a) = a$, and the output unit uses a sigmoid activation function, $h(a) = 1/(1 + exp(-a))$.

- (3 marks) Write down a mathematical formula for what the output of this neural network is, given an input $\boldsymbol{x} = (x_1, \ldots, x_D)$.

- (5 marks) Is this model equivalent to logistic regression? What if regularization is used when learning model parameters?

3. (11 marks) Consider the Bayesian network below. Assume each random variable is boolean (takes values *true* or *false*).



- (3 marks) How many independent parameters are needed to specify each of the conditional distributions in this network?

- (2 marks) Draw the Markov random field that is obtained by moralizing this Bayesian network.

- (3 marks) Write down the factorization of the joint probability distribution that is given by this Markov random field.

- (3 marks) How many independent parameters are needed to specify each of the potential functions in this Markov random field?

4. (8 marks) Consider the problem of determining K, the number of Gaussians in a mixture of Gaussians model.

$$p(x) = \sum_{k=1}^{K} \pi_k \mathcal{N}(x|\mu_k, \Sigma_k)$$

- (4 marks) What happens to the likelihood obtained using expectation-maximization as we increase K?

- (4 marks) Suggest how we might decide on a good value for K when doing unsupervised learning.