

# CSE 250A Quiz 6

Tuesday November 13, 2012

*Instructions.* You should do this quiz in partnership with exactly one other student. Write both your names at the top of this page. Discuss the answer to the question with each other, and then write your joint answer below the question. It is ok if you overhear other students' discussions, because you still need to decide if they are right or wrong. You have seven minutes.

Let  $\alpha = p(X_i = x | \text{pa}(X_i) = \pi)$  be one parameter of a Bayesian network. When using expectation-maximization to learn  $\alpha$ , in general the M step is

$$\alpha := \frac{\sum_{t=1}^T p(X_i = x, \text{pa}(X_i) = \pi | v_t)}{\sum_{x'} \sum_{t=1}^T p(X_i = x', \text{pa}(X_i) = \pi | v_t)}.$$

Obtain a simple upper bound for the denominator. Under what circumstances is the upper bound achieved?

*Answer.* Switching the order of the summations, the denominator is

$$\sum_{t=1}^T \sum_{x'} p(X_i = x', \text{pa}(X_i) = \pi | v_t) \leq \sum_{t=1}^T \sum_{x'} p(X_i = x' | v_t) = \sum_{t=1}^T 1 = T.$$

The upper bound  $T$  is achieved when the node  $X_i$  has no parents.

*Additional comments.* The answer relies on the basic fact that  $p(A, B) \leq p(A)$  for all events  $A$  and  $B$ .

The denominator is essentially the total weight of all training examples that are relevant for estimating  $\alpha$ . A training example is relevant only if it has the specified combination of values  $\pi$  for the parent nodes. The weight of the example is 1 if this exact combination of values is stated in  $v_t$ , 0 if any value included in  $\pi$  conflicts with a value stated in  $v_t$ , and otherwise between 0 and 1 if some variables in  $\text{pa}(X_i)$  are hidden.