안녕하세요, HW3 입니다. (추후에 문제가 추가될 수 있습니다.)

기한은 5/24 수업전까지 입니다. HW1과 동일한 포맷으로 제출해주시기 바랍니다.

15.1 Show that any second-order Markov process can be rewritten as a first-order Markov process with an augmented set of state variables. Can this always be done *parsimoniously*, i.e., without increasing the number of parameters needed to specify the transition model?

15.5 Equation (15.12) describes the filtering process for the matrix formulation of HMMs. Give a similar equation for the calculation of likelihoods, which was described generically in Equation (15.7).

$$\mathbf{f}_{1:t+1} = \alpha\, \mathbf{O}_{t+1}\mathbf{T}^{\top}\mathbf{f}_{1:t} \tag{15.12}$$

$$L_{1:t} = P(\mathbf{e}_{1:t}) = \sum_{\mathbf{x}_t} \ell_{1:t}(\mathbf{x}_t)\ . \tag{15.7}$$

15.13 A professor wants to know if students are getting enough sleep. Each day, the professor observes whether the students sleep in class, and whether they have red eyes. The professor has the following domain theory:

- The prior probability of getting enough sleep, with no observations, is 0.7.
- The probability of getting enough sleep on night t is 0.8 given that the student got enough sleep the previous night, and 0.3 if not.
- The probability of having red eyes is 0.2 if the student got enough sleep, and 0.7 if not.
- The probability of sleeping in class is 0.1 if the student got enough sleep, and 0.3 if not.

Formulate this information as a dynamic Bayesian network that the professor could use to filter or predict from a sequence of observations. Then reformulate it as a hidden Markov model that has only a single observation variable. Give the complete probability tables for the model.

15.14 For the DBN specified in Exercise 15.13 and for the evidence values

$\mathbf{e}_1$ = not red eyes, not sleeping in class
$\mathbf{e}_2$ = red eyes, not sleeping in class
$\mathbf{e}_3$ = red eyes, sleeping in class

perform the following computations:

a. State estimation: Compute $P(EnoughSleep_t|\mathbf{e}_{1:t})$ for each of $t = 1, 2, 3$.

b. Smoothing: Compute $P(EnoughSleep_t|\mathbf{e}_{1:3})$ for each of $t = 1, 2, 3$.

c. Compare the filtered and smoothed probabilities for $t = 1$ and $t = 2$.

20.1 The data used for Figure 20.1 on page 804 can be viewed as being generated by h_5. For each of the other four hypotheses, generate a data set of length 100 and plot the corresponding graphs for $P(h_i \mid d_1, \ldots, d_N)$ and $P(D_{N+1} = lime \mid d_1, \ldots, d_N)$. Comment on your results.

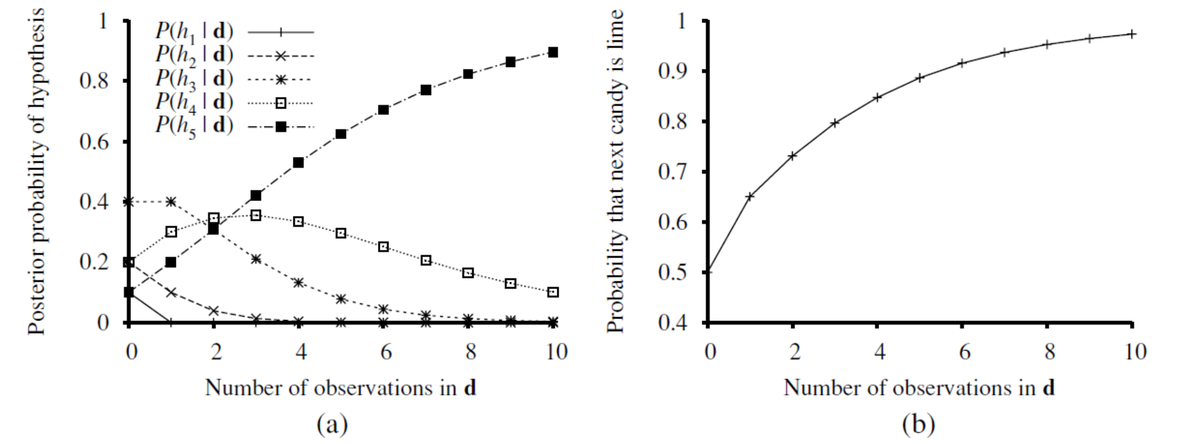

Figure 20.1 (a) Posterior probabilities $P(h_i \mid d_1, \ldots, d_N)$ from Equation (20.1). The number of observations N ranges from 1 to 10, and each observation is of a lime candy. (b) Bayesian prediction $P(d_{N+1} = lime \mid d_1, \ldots, d_N)$ from Equation (20.2).

18.19 Construct by hand a neural network that computes the XOR function of two inputs. Make sure to specify what sort of units you are using.

18.20 Recall from Chapter 18 that there are 2^{2^n} distinct Boolean functions of n inputs. How many of these are representable by a threshold perceptron?

18.22 Suppose you had a neural network with linear activation functions. That is, for each unit the output is some constant c times the weighted sum of the inputs.

a. Assume that the network has one hidden layer. For a given assignment to the weights **w**, write down equations for the value of the units in the output layer as a function of **w** and the input layer **x**, without any explicit mention of the output of the hidden layer. Show that there is a network with no hidden units that computes the same function.

b. Repeat the calculation in part (a), but this time do it for a network with any number of hidden layers.

c. Suppose a network with one hidden layer and linear activation functions has n input and output nodes and h hidden nodes. What effect does the transformation in part (a) to a network with no hidden layers have on the total number of weights? Discuss in particular the case $h \ll n$.

Neural Networks 강의자료의 Backpropagation 유도식을 스스로 유도해보시오.

23.2 An *HMM grammar* is essentially a standard HMM whose state variable is N (nonterminal, with values such as Det, $Adjective$, $Noun$ and so on) and whose evidence variable is W (word, with values such as is, $duck$, and so on). The HMM model includes a prior $\mathbf{P}(N_0)$, a transition model $\mathbf{P}(N_{t+1}|N_t)$, and a sensor model $\mathbf{P}(W_t|N_t)$. Show that every HMM grammar can be written as a PCFG. [Hint: start by thinking about how the HMM prior can be represented by PCFG rules for the sentence symbol. You may find it helpful to illustrate for the particular HMM with values A, B for N and values x, y for W.]

23.3 Consider the following PCFG for simple verb phrases:

$0.1 : VP \rightarrow Verb$
$0.2 : VP \rightarrow Copula\ Adjective$
$0.5 : VP \rightarrow Verb\ the\ Noun$
$0.2 : VP \rightarrow VP\ Adverb$
$0.5 : Verb \rightarrow is$
$0.5 : Verb \rightarrow shoots$
$0.8 : Copula \rightarrow is$
$0.2 : Copula \rightarrow seems$
$0.5 : Adjective \rightarrow$ **unwell**
$0.5 : Adjective \rightarrow$ **well**
$0.5 : Adverb \rightarrow$ **well**
$0.5 : Adverb \rightarrow$ **badly**
$0.6 : Noun \rightarrow$ **duck**
$0.4 : Noun \rightarrow$ **well**

a. Which of the following have a nonzero probability as a VP? (i) shoots the duck well well well (ii) seems the well well (iii) shoots the unwell well badly

b. What is the probability of generating "is well well"?

c. What types of ambiguity are exhibited by the phrase in (b)?

d. Given any PCFG, is it possible to calculate the probability that the PCFG generates a string of exactly 10 words?

23.6 Consider the sentence "Someone walked slowly to the supermarket" and a lexicon consisting of the following words:

$Pronoun \rightarrow$ **someone**	$Verb \rightarrow$ **walked**
$Adv \rightarrow$ **slowly**	$Prep \rightarrow$ **to**
$Article \rightarrow$ **the**	$Noun \rightarrow$ **supermarket**

Which of the following three grammars, combined with the lexicon, generates the given sentence? Show the corresponding parse tree(s).

(A):	(B):	(C):
$S \rightarrow NP\ VP$	$S \rightarrow NP\ VP$	$S \rightarrow NP\ VP$
$NP \rightarrow Pronoun$	$NP \rightarrow Pronoun$	$NP \rightarrow Pronoun$
$NP \rightarrow Article\ Noun$	$NP \rightarrow Noun$	$NP \rightarrow Article\ NP$
$VP \rightarrow VP\ PP$	$NP \rightarrow Article\ NP$	$VP \rightarrow Verb\ Adv$
$VP \rightarrow VP\ Adv\ Adv$	$VP \rightarrow Verb\ Vmod$	$Adv \rightarrow Adv\ Adv$
$VP \rightarrow Verb$	$Vmod \rightarrow Adv\ Vmod$	$Adv \rightarrow PP$
$PP \rightarrow Prep\ NP$	$Vmod \rightarrow Adv$	$PP \rightarrow Prep\ NP$
$NP \rightarrow Noun$	$Adv \rightarrow PP$	$NP \rightarrow Noun$
	$PP \rightarrow Prep\ NP$	

For each of the preceding three grammars, write down three sentences of English and three sentences of non-English generated by the grammar. Each sentence should be significantly different, should be at least six words long, and should include some new lexical entries (which you should define). Suggest ways to improve each grammar to avoid generating the non-English sentences.

23.9 Consider the following toy grammar:

$S \rightarrow NP\ VP$
$NP \rightarrow Noun$
$NP \rightarrow NP$ **and** NP
$NP \rightarrow NP\ PP$
$VP \rightarrow Verb$
$VP \rightarrow VP$ **and** VP
$VP \rightarrow VP\ PP$
$PP \rightarrow Prep\ NP$

$Noun \rightarrow$ **Sally** | **pools** | **streams** | **swims**
$Prep \rightarrow$ **in**
$Verb \rightarrow$ **pools** | **streams** | **swims**

a. Show all the parse trees in this grammar for the sentence "Sally swims in streams and pools."

b. Show all the table entries that would be made by a (non-probabalistic) CYK parser on this sentence.