UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Department of Electrical and Computer Engineering

ECE 544NA Pattern Recognition
Fall 2016

SAMPLE EXAM 3

- This is a CLOSED BOOK exam. You may use one page, both sides, of handwritten notes.
- There are a total of 100 points in the exam. Plan your work accordingly.
- You must SHOW YOUR WORK to get full credit.

Problem	Score
1	
2	
3	
4	
5	
6	
Total	

Name: ____

Problem 1 (20 points)

An RNN has output $\hat{y}(t)$ defined for $0 \leq t < \infty$, input $x(t)$, error E_t, network weights a_t and b_t, and scalar nonlinearity $g(\cdot)$ related by

$$\begin{aligned}
\hat{y}(t) &= g\left(a_t \hat{y}(t-1) + b_t x(t)\right) &(1)\\
E_t &= \frac{1}{2}\left(y(t) - \hat{y}(t)\right)^2 &(2)\\
a_0 = b_0 &= 0 &(3)\\
a_{t+1} &= a_t - 0.02 \sum_{k=0}^{t} \frac{\partial E_t}{\partial a_k} &(4)
\end{aligned}$$

(a) Prove that, even if E_t is bounded, $\sum_{k=0}^{t} \frac{\partial E_t}{\partial a_k}$ might grow without bound.

(b) Modify the update equation to

$$a_{t+1} = \max\left(-a_{MAX}, \min\left(a_{MAX}, a_t - 0.02 \sum_{k=0}^{t} \frac{\partial E_t}{\partial a_k}\right)\right)$$

Find sufficient conditions on a_{MAX} and $g(\cdot)$ such that $\sum_{k=0}^{t} \frac{\partial E_t}{\partial a_k}$ is guaranteed to remain bounded as $t \to \infty$.

Problem 2 (10 points)

Consider a two-layer RNN, one node per layer, with input $x(t)$, hidden layer $h(t)$, output $\hat{y}(t)$, error E_t, scalar nonlinearity $g(\cdot)$, and coefficients a, b, α, β related by

$$\begin{aligned}
h(t) &= g\left(a h(t-1) + b x(t)\right) &(5)\\
\hat{y}(t) &= \left(\alpha \hat{y}(t-1) + \beta h(t)\right) &(6)\\
E_t &= y(t)\log(\hat{y}(t)) + (1 - y(t))\log(1 - \hat{y}(t)) &(7)
\end{aligned}$$

Find $\frac{\partial E_t}{\partial \beta}$ and $\frac{\partial E_t}{\partial b}$.

Problem 3 (20 points)

A simplified LSTM with inputs $x(t)$ is defined by

$$\begin{aligned}
c(t) &= i(t)x(t) + m(t)c(t-1) &(8)\\
i(t) &= u\left(w^i x(t) + b^i\right) &(9)\\
m(t) &= u\left(w^m x(t) + b^m\right) &(10)
\end{aligned}$$

where $u(\cdot)$ is the unit step function, defined as $u(z) = \frac{1}{2}\left(\text{sign}(z) + 1\right)$.

(a) Choose w^i, b^i, w^m and b^m so that

$$c(t) = \begin{cases} x(t) & x(t) > 2 \\ c(t-1) & x(t) < 2 \end{cases} \qquad (11)$$

(b) Suppose $x(t) = s(t) + v(t)$, where $s(t)$ is the desired signal,

$$s(t) = \begin{cases} 10 & t = 0 \\ 0 & \text{otherwise} \end{cases}$$

and $v(t)$ is a random noise process with distribution given by

$$\Phi(z) = \Pr\{v(t) \le z\} = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du$$

Define the "memory" of the network you designed in part (a) to be the expected number of time steps for which $c(t) = x(0)$. Find the memory of your network, as a function of $\Phi(z)$.

Problem 4 (25 points)

Define $c_t = i$ to be the event that the i^{th} coin is flipped at time t, where $1 \le i \le 3$. The possible observations are $x_t \in \{H, T\}$. Two of the coins are unfair; the heads probabilities of the three coins are given by $p(x_t = H|c_t = i) = i/4$. Coin $c_1 = 1$ is always the first one flipped. After each coin flip, the coin is changed with probability $\frac{1}{2}$; if the coin is changed, both of the other coins are equally likely.

(a) What is $p(c_1 = 1, c_2 = 1, c_3 = 2)$?

(b) What is the probability of getting three heads in a row?

(c) Suppose you observe two heads in a row. You know that the first flip was coin $c_1 = 1$. Which coin was most probably the second one flipped?

Problem 5 (15 points)

Suppose that an HMM has 6 states with transition probabilities arranged into a matrix A, whose $(i,j)^{\text{th}}$ element is $a_{ij} = \Pr\{s_t = j|s_{t-1} = i\}$. The a_{ij} are initialized randomly, but because of a programming error, a_{34} is accidentally set to $a_{34} = 0$. New transition probabilities $\hat{a}_{ij}$ are estimated as

$$\hat{a}_{ij} = \frac{\sum_{t=1}^{T} \xi_t(i,j)}{\sum_{t=1}^{T} \sum_{j=1}^{6} \xi_t(i,j)} \tag{12}$$

$$\xi_t(i,j) = p(s_{t-1} = i, s_t = j|A, X) \tag{13}$$

where $X = [x_1, \ldots, x_T]$ is a sequence of observations whose details you do not need to know. Prove that, under these circumstances, $\hat{a}_{34} = 0$.

Problem 6 (10 points)

A NN-HMM hybrid has known observations $X = [x_1, \ldots, x_T]$, and unknown state sequence $S = [s_1, \ldots, s_T]$. Its initialization probabilities are $\pi_i = \Pr\{s_1 = i\}$, and its transition probabilities are $a_{ij} = \Pr\{s_t = j|s_{t-1} = i\}$. Its observation probabilities are computed by the softmax layer of a neural network; they are $b_j(x_t) = \Pr\{s_t = j|x_t\} / \Pr\{s_t = j\}$. Consider the following algorithm:

$$\alpha_1(i) = \pi_i b_i(x_1), \quad 1 \le i \le N \tag{14}$$

$$\alpha_{t+1}(j) = \sum_{i=1}^{N} \alpha_t(i) a_{ij} b_j(x_t), \quad 1 \le i, j \le N, \quad 1 \le t \le T-1 \tag{15}$$

$$P_{final} = \sum_{j=1}^{N} \alpha_T(j) \tag{16}$$

Express $P_{final} = \ldots$, where the right-hand side of the equation contains some sort of arithmetic combination of joint, conditional, and marginal probability mass functions of the random variables x_t. No other variables should appear on the right-hand side.