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

ECE 417 Multimedia Signal Processing
Fall 2023

# EXAM 2

Tuesday, October 31, 2023

- This is a CLOSED BOOK exam.
- You are permitted one sheet of handwritten notes, 8.5x11.
- Calculators and computers are not permitted.
- Don't simplify explicit numerical expressions.
- If you're taking the exam online, you will need to have your webcam turned on. Your exam will appear on Gradescope at exactly 9:30am; you will need to photograph and upload your answers by exactly 11:00am.
- There are a total of 100 points in the exam. Each problem specifies its point total. Plan your work accordingly.
- You must SHOW YOUR WORK to get full credit.

Name: ________________________________

NetID: ________________________________

**Neural Nets**

$$a_{i,k}^{(\ell)} = g(\xi_{i,k}^{(\ell)}), \qquad \xi_{i,k}^{(\ell)} = b_k^{(\ell)} + \sum_{j=1}^{p} w_{k,j}^{(\ell)} a_{i,j}^{(\ell-1)}$$

$$\frac{d\mathcal{L}}{d\xi_{i,k}^{(\ell)}} = \frac{d\mathcal{L}}{da_{i,k}^{(\ell)}} \dot{g}(\xi_{i,k}^{(\ell)}), \qquad \dot{\sigma}(x) = \sigma(x)(1-\sigma(x))$$

$$\frac{d\mathcal{L}}{da_{i,j}^{(\ell-1)}} = \sum_k \frac{d\mathcal{L}}{d\xi_{i,k}^{(\ell)}} w_{k,j}^{(\ell)}, \qquad \frac{d\mathcal{L}}{dw_{k,j}^{(\ell)}} = \sum_i \frac{d\mathcal{L}}{d\xi_{i,k}^{(\ell)}} a_{i,j}^{(\ell-1)}$$

$$w_{k,j}^{(\ell)} \leftarrow w_{k,j}^{(\ell)} - \eta \frac{d\mathcal{L}}{dw_{k,j}^{(\ell)}}$$

**Viola-Jones**

$$II[m,n] = \sum_{m'=1}^{m} \sum_{n'=1^n} I[m',n'], \quad 1 \le m \le M, 1 \le n \le N$$

$$\epsilon_t = \sum_i w_t(x_i)|y_i - h_t(x_i)|, \quad w_{t+1}(x_i) = \beta_t w_t(x_i), \quad \beta_t = \frac{\epsilon_t}{1-\epsilon_t}$$

$$h(x) = \begin{cases} 1 & \sum_t \alpha_t h_t(x) > \frac{1}{2}\sum_t \alpha_t \\ 0 & \text{otherwise} \end{cases}, \qquad \alpha_t = -\ln \beta_t$$

**Hidden Markov Model**

$$\alpha_t(j) = \sum_{i=1}^{N} \alpha_{t-1}(i) a_{i,j} b_j(\boldsymbol{x}_t), \;\; 1 \le j \le N, \; 2 \le t \le T, \quad \hat{\alpha}_t(j) = \frac{\sum_{i=1}^{N} \hat{\alpha}_{t-1}(i) a_{i,j} b_j(\boldsymbol{x}_t)}{\sum_{j'=1}^{N}\sum_{i=1}^{N} \hat{\alpha}_{t-1}(i) a_{i,j'} b_{j'}(\boldsymbol{x}_t)}$$

$$\beta_t(i) = \sum_{j=1}^{N} a_{i,j} b_j(\boldsymbol{x}_{t+1}) \beta_{t+1}(j), \;\; 1 \le i \le N, \; 1 \le t \le T-1$$

$$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{k=1}^{N} \alpha_t(k)\beta_t(k)}, \quad \xi_t(i,j) = \frac{\alpha_t(i) a_{i,j} b_j(\boldsymbol{x}_{t+1}) \beta_{t+1}(j)}{\sum_{k=1}^{N}\sum_{\ell=1}^{N} \alpha_t(k) a_{k,\ell} b_\ell(\boldsymbol{x}_{t+1}) \beta_{t+1}(\ell)}$$

$$a'_{i,j} = \frac{\sum_{t=1}^{T-1} \xi_t(i,j)}{\sum_{j=1}^{N}\sum_{t=1}^{T-1} \xi_t(i,j)}, \quad b'_j(k) = \frac{\sum_{t:x_t=k} \gamma_t(j)}{\sum_{t=1}^{T} \gamma_t(j)}$$

**Gaussians**

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^{D/2}|\boldsymbol{\Sigma}|^{1/2}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})}$$

$$\boldsymbol{\mu}'_i = \frac{\sum_{t=1}^{T} \gamma_t(i) \boldsymbol{x}_t}{\sum_{t=1}^{T} \gamma_t(i)}, \quad \boldsymbol{\Sigma}'_i = \frac{\sum_{t=1}^{T} \gamma_t(i)(\boldsymbol{x}_t - \boldsymbol{\mu}_i)(\boldsymbol{x}_t - \boldsymbol{\mu}_i)^T}{\sum_{t=1}^{T} \gamma_t(i)}$$

1. (20 points) Consider a new type of neural network called a Fourier network, invented for this problem. Given an input vector $\boldsymbol{x} \in \Re^{n_x}$, a Fourier network computes $\boldsymbol{y} \in \Re^{n_y}$ as

$$\boldsymbol{y} = \boldsymbol{R}\cos(\boldsymbol{P}\boldsymbol{x}) + \boldsymbol{S}\sin(\boldsymbol{Q}\boldsymbol{x}),$$

where the functions $\cos(\cdot)$ and $\sin(\cdot)$ are element-wise scalar cosine and sine, respectively, and where $\boldsymbol{P}, \boldsymbol{Q} \in \Re^{n_h \times n_x}$ and $\boldsymbol{R}, \boldsymbol{S} \in \Re^{n_y \times n_h}$ are weight matrices that must be learned from training data. Suppose the loss is

$$\mathcal{L} = \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{t}\|_2^2$$

for some target vector $\boldsymbol{t}$. Express either $\frac{\partial \mathcal{L}}{\partial q_{\ell,m}}$, where $q_{\ell,m}$ is the $(\ell, m)^{\text{th}}$ element of $\boldsymbol{Q}$, in terms of any of the elements of the vectors $\boldsymbol{t}, \boldsymbol{x}, \boldsymbol{y}$ and/or any of the matrices $\boldsymbol{P}, \boldsymbol{Q}, \boldsymbol{R}, \boldsymbol{S}$.

**Solution:**

$$y_i = \sum_{j=1}^{n_h} s_{i,j} \sin\left(\sum_{k=1}^{n_x} q_{j,k} x_k\right) + \text{other terms}$$

$$\mathcal{L} = \frac{1}{2}\sum_{i=1}^{n_y}(y_i - t_i)^2$$

$$\frac{\partial \mathcal{L}}{\partial q_{\ell,m}} = \sum_{i=1}^{n_y} \frac{\partial \mathcal{L}}{\partial y_i}\frac{\partial y_i}{\partial q_{\ell,m}}$$

$$= \sum_{i=1}^{n_y}(y_i - t_i) s_{i,\ell} \cos(\sum_{k=1}^{n_x} q_{\ell,k} x_k) x_m$$

2. (20 points) Consider a 1D convnet with input $x[n]$ and output $y[n]$ related according to

$$z[k] = \sum_{m=0}^{M-1} h[m]x[k-m],$$
$$y[n] = \max\left(z[2n], z[2n+1]\right),$$

where $h[n]$ is a set of learned filter weights. Suppose the loss is

$$\mathcal{L} = \frac{1}{2}\sum_{n=0}^{N-1}(y[n]-s[n])^2,$$

where $s[n]$ is a target output. In terms of $h[n], s[n], x[n], y[n], z[n]$, $M$, $N$, and/or $\ell$, what is $\frac{\partial \mathcal{L}}{\partial h[\ell]}$?

**Solution:**

$$\frac{\partial \mathcal{L}}{\partial h[\ell]} = \sum_{n=0}^{N-1}(y[n]-s[n])x[k^*(n)-\ell],$$

where $k^*(n) = \text{argmax}(z[2n], z[2n+1])$.

3. (20 points) A particular Adaboost classifier is being trained using 6 training examples, $x_1$ through $x_6$. In the first round of training, all 6 have equal weight, so that $w_1(x_1) = \ldots = w_1(x_6) = \frac{1}{6}$. The first weak classifier is able to correctly classify tokens $x_1$ through $x_4$, but it incorrectly classifies tokens $x_5$ and $x_6$, so its weighted error is $\epsilon_1 = \frac{1}{3}$. After computing the weights $w_2(x_i)$ for $1 \leq i \leq 6$, a second weak classifier is chosen. The second weak classifier correctly classifies tokens $x_3$ through $x_6$, but incorrectly classifies $x_1$ and $x_2$. What is $\epsilon_2$?

**Solution:**

$$\beta_1 = \frac{\epsilon_1}{1-\epsilon_1} = \frac{1}{2}$$

$$\boldsymbol{w}_2 = \left[\frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{8}, \frac{1}{4}, \frac{1}{4}\right]$$

$$\epsilon_2 = w_2(x_1) + w_2(x_2) = \frac{1}{4}$$

4. (20 points) In a given length-$T$ observation sequence, the sixth observation is missing, because at the moment $x_6$ should have been recorded, a cyberattack caused the sensor to briefly go offline. Therefore, you have valid observations of $x_1, \ldots, x_5$ and of $x_7, \ldots, x_T$, but the observation $x_6$ is unknown. The observations are drawn from a discrete finite set $x_t \in \{1, \ldots, K\}$, and are known to have been generated by an N-state HMM with known model parameters $\Lambda = \{\pi_i, a_{i,j}, b_j(k) \forall 1 \leq i, j \leq N, 1 \leq k \leq K\}$.

   Note that, because of the limited observations, you can only find the forward probabilities $\alpha_t(i)$ for $1 \leq t \leq 5$. Similarly, you can only find the backward probabilities $\beta_t(i)$ for $6 \leq t \leq T$.

   Suppose you want to find the following quantity:

$$\zeta_6(\ell) = \Pr\{x_6 = \ell | x_1, \ldots, x_5, x_7, \ldots, x_T, \Lambda\}, \quad 1 \leq \ell \leq K$$

   Find an expression for $\zeta_6(\ell)$ in terms of the model parameters, and in terms of the forward and backward probabilities at times for which they are known. Make sure that your answer is a function of no random variable other than $\ell$.

**Solution:**

$$\begin{aligned}
\zeta_6(\ell) &= \frac{\sum_{j=1}^N \sum_{i=1}^N \Pr\{x_1, \ldots, x_5, q_5 = i\} a_{i,j} b_j(\ell) \Pr\{x_7, \ldots, x_T | q_6 = j\}}{\sum_{k=1}^K \sum_{j=1}^N \sum_{i=1}^N \Pr\{x_1, \ldots, x_5, q_5 = i\} a_{i,j} b_j(k) \Pr\{x_7, \ldots, x_T | q_6 = j\}} \\
&= \frac{\sum_{i=1}^N \sum_{j=1}^N \alpha_5(i) a_{i,j} b_j(\ell) \beta_6(j)}{\sum_{k=1}^K \sum_{i=1}^N \sum_{j=1}^N \alpha_5(i) a_{i,j} b_j(k) \beta_6(j)}
\end{aligned}$$

5. (20 points) Consider an $N$-state Gaussian HMM with $D$-dimensional observations, $\boldsymbol{x} \in \Re^D$, in which all states have the same already-known covariance matrix, $\Sigma_1 = \ldots = \Sigma_N = \Sigma$, but their mean vectors are distributed as $\boldsymbol{\mu}_i = \boldsymbol{\mu}_0 + \boldsymbol{\delta}_i$, where $\boldsymbol{\mu}_0 \in \Re^D$ is a known global mean, and $\boldsymbol{\delta}_i \in \Re^D, 1 \le i \le N$, is an unknown state-dependent offset vector. Suppose you are given a training sequence $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T]$, for which it is known that $q_1 = q_2 = \ldots = q_T = 5$, i.e., all of these observations were generated by the fifth HMM state. Find the maximum-likelihood estimate of the parameter vector $\boldsymbol{\delta}_5$ in terms of the known parameters $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}$ and the observation vectors $\boldsymbol{x}_1$ through $\boldsymbol{x}_T$.

**Solution:**

$$\mathcal{L} = -\frac{1}{2}\sum_{t=1}^{T}\Big((\boldsymbol{x}_t - \boldsymbol{\mu}_0 - \boldsymbol{\delta}_5)^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}_t - \boldsymbol{\mu} - \boldsymbol{\delta}_5) + \ln|\boldsymbol{\Sigma}| + 2\pi D\Big) \quad (1)$$

$$\frac{\partial \mathcal{L}}{\partial \boldsymbol{\delta}_5} = \sum_{t=1}^{T}(\boldsymbol{x}_t - \boldsymbol{\mu}_0 - \boldsymbol{\delta}_5)^T\boldsymbol{\Sigma}^{-1} \quad (2)$$

Since $\mathcal{L}$ is quadratic with negative sign, its global maximum is achieved by setting its derivative equal to zero. Re-arranging terms, we get:

$$\boldsymbol{\delta}_5 = \frac{1}{T}\sum_{t=1}^{T}\boldsymbol{x}_t - \boldsymbol{\mu}_0$$

THIS PAGE INTENTIONALLY BLANK.