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

ECE 417 Multimedia Signal Processing
Fall 2017

EXAM 2 SOLUTIONS

Thursday, October 26, 2017

Problem 1 (10 points)

A speech signal can be modeled as an excitation passed through a filter, $s[n] = h[n] * e[n]$. A reasonable (very) simplified model of voiced speech might use

$$e[n] = \delta[n] + \sum_{m=-\infty}^{\infty} \delta[n - mT_0], \quad h[n] = \left(A_1 p_1^n + A_1^*(p_1^*)^n\right) u[n]$$

where p_1 is the complex first formant frequency, and A_1 is some appropriate constant. For purposes of this problem, define the cepstrum to be the inverse DTFT of the log DTFT:

$$\hat{s}[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln S(\omega) e^{j\omega n} d\omega$$

You are given the cepstrum of $h[n]$:

$$\hat{h}[n] = \left(\frac{p_1^n}{n} + \frac{(p_1^*)^n}{n}\right) u[n]$$

and you may assume that

$$\ln\left(1 + \frac{2\pi}{T_0} \sum_{k=0}^{T_0-1} \delta\left(\omega - \frac{2\pi k}{T_0}\right)\right) \approx \ln\left(\frac{2\pi}{T_0}\right) + \sum_{k=0}^{T_0-1} \delta\left(\omega - \frac{2\pi k}{T_0}\right)$$

Under these assumptions, find the cepstrum $\hat{s}[n]$ of the speech signal, in terms of the parameters T_0 and p_1.

Solution

$$\hat{s}[n] = \left(\frac{p_1^n}{n} + \frac{(p_1^*)^n}{n}\right) u[n] + \ln\left(\frac{2\pi}{T_0}\right) \delta[n] + \frac{T_0}{2\pi} \sum_{m=\infty}^{\infty} \delta[n - mT_0]$$

Problem 2 (10 points)

A speech signal, $x[n]$, has been sampled at F_s samples/second. Suppose the signal is only N samples long, so we can compute its N-sample DFT, $X[k]$. The mel spectrum is then defined to be

$$\tilde{X}[m] = \begin{cases} \sum_{k=0}^{N/2} W_m[k] \cdot |X[k]| & 1 \le m \le M \\ 0 & \text{otherwise} \end{cases}$$

where the filters $W_m[k]$ are uniformly spaced in a mel-frequency scale. Suppose that the speech signal is known to be $x[n] = h[n] * e[n]$, where $e[n]$ has mel spectrum $\tilde{E}[m]$, and $h[n]$ has mel spectrum $\tilde{H}[m]$. Define the mel-frequency cepstral coefficients to be

$$\tilde{x}[n] = \text{DFT}^{-1}\left\{\ln\left(\tilde{X}[m] + \tilde{X}[2M+2-m]\right)\right\}$$

where the DFT has a length of $N = 2M + 2$. Prove that $\tilde{x}[n]$ is a symmetric real-valued sequence.

Solution

The mel-frequency samples at $\tilde{X}[0]$ and $\tilde{X}[M+1]$ are unspecified. If we assume that both of those samples are zero, then the IDFT formula

$$\tilde{x}[n] = \frac{1}{N}\sum_{k=0}^{N-1} \ln\left(\tilde{X}[m] + \tilde{X}[2M+2-m]\right) e^{j2\pi mn/N}$$

simplifies to

$$\tilde{x}[n] = \frac{2}{N}\sum_{k=1}^{M} \ln \tilde{X}[m] \cos\left(\frac{\pi mn}{M+1}\right)$$

This is real and symmetric, as long as $\ln \tilde{X}[m]$ is real, which is true, for example, if $W_m[k]$ and $|X[k]|$ are both real non-negative quantities.

Problem 3 (10 points)

Three scalar random variables A, V, and Y are jointly distributed as

$$p_{Y|V}(y|v) = \frac{1}{3}C\left(y_k = y\right), \quad p_{A|Y}(a|y) = \mathcal{N}(a; \mu_y, 1) \tag{1}$$

where $\mathcal{N}(a; \mu_y, 1)$ is a scalar Gaussian with unit variance, and with class-dependent means

$$\mu_0 = -1, \quad \mu_1 = 1 \tag{2}$$

The operator $C(y_k = y)$ is a nearest-neighbor count operator: it finds three training samples v_k with the three smallest values of $(v_k - v)^2$, then counts how many of those samples have the label y. The training samples are

$$\{v_1, \ldots, v_9\} = [-2, -1, 0, 1, 2, 3], \quad \{y_1, \ldots, y_9\} = [0, 0, 0, 1, 1, 1] \tag{3}$$

(a) Sketch $p_{Y|V}(1|v)$ as a function of v.

Solution

$$p_{Y|V}(y|v) = \begin{cases} 0 & v < -\frac{1}{2} \\ \frac{1}{3} & -\frac{1}{2} < v < \frac{1}{2} \\ \frac{2}{3} & \frac{1}{2} < v < \frac{3}{2} \\ 1 & \frac{3}{2} < v \end{cases}$$

(b) Based on Eqs. 1 through 3 on the previous page, design a classifier $\hat{y}(a,v)$ such that

$$\hat{y}(a,v) = \arg\max_y \ln\left(p_{A|Y}(a|y)p_{Y|V}(y|v)\right)$$

Draw a two-dimensional space whose axes are v and a; show the region $-5 \le v, a \le 5$. In the two-dimensional space, draw the decision boundary between the class $\hat{y}(a,v) = 0$ and the class $\hat{y}(a,v) = 1$.

Solution

The plot should show a boundary composed of vertical and horizontal line segments at $v = -\frac{1}{2}$,$a = -\frac{1}{2}\ln(2)$, $v = \frac{1}{2}$, $a = \frac{1}{2}\ln(2)$, and $v = \frac{3}{2}$.

Problem 4 (10 points)

Suppose you have a scalar random variable X, with training examples $[x_1, x_2, x_3, x_4] = [-1, 0, 1, 2]$. You want to try to cluster these into two clusters. You have initial estimates of the cluster centroids, as $\mu_0 = 0$, $\mu_1 = 1$.

(a) Perform one iteration of K-means clustering: assign each token to a cluster, then recompute the new cluster centroids. What are the new cluster centroids?

Solution

$$\mu_0 = -\frac{1}{2}, \quad \mu_1 = \frac{3}{2}$$

(b) Assume that part (a) never happened. Instead, perform one iteration of EM training. You have initial parameter estimates $\mu_0 = 0$, $\mu_1 = 1$, $\sigma_0^2 = \sigma_1^2 = 1$, and $c_0 = c_1 = 0.5$. Perform one iteration of EM training. What are the new cluster centroids? Give numerical values for the new values of μ_0 and μ_1; in order to do so, assume that the unit normal Gaussian pdf has the following values: $\mathcal{N}(0;0,1) \approx \frac{2}{5}$, $\mathcal{N}(1;0,1) \approx \frac{1}{4}$, $\mathcal{N}(2;0,1) \approx \frac{1}{20}$, and $\mathcal{N}(x;0,1) \approx 0$ for $x \ge 3$. Note: since you didn't bring a calculator to this exam, feel free to leave your answer in the form of a sum of fractions.

Solution:

$$\mu_0 = -\frac{3}{52}, \quad \mu_1 = \frac{55}{52}$$