KTH Electrical Engineering

Exam in Pattern Recognition EQ2340

Date: Friday Oct 30, 2015, 14:00 – 19:00

Place: V01, V11.

Allowed: Beta, calculator with empty memory, one page handwritten note.

Grades: A: 31p; B: 27p; C: 23p; D: 20p; E: 17; of max 25p + 10p project bonus.

Language: English.

Results: Monday, Nov 23, 2015.

Review: Via scanned version

Contact: Saikat: 0738913581; Arun: 0727895538

Good Luck!

1 A signal source randomly selects a state $S = 1$ or $S = 2$ with equal probability. In both states the signal source generates stationary zero mean two-dimensional Gaussian mixture source $X(n) = (X_1(n),\ X_2(n))^\top$ with independent elements over time index n, that is,

$$P(X(n_1), X(n_2)) = P(X(n_1))\, P(X(n_2)),\ n_1 \neq n_2.$$

For state 1 and state 2, the Gaussian mixture distributions of states are

$$f_{X|S}(x|1) = 0.25\,\mathcal{N}(\mathbf{0}, C_{11}) + 0.75\,\mathcal{N}(\mathbf{0}, C_{12})$$
$$f_{X|S}(x|2) = 0.25\,\mathcal{N}(\mathbf{0}, C_{21}) + 0.75\,\mathcal{N}(\mathbf{0}, C_{22}),$$

where

$$C_{11} = \begin{bmatrix} 8 & 0 \\ 0 & 8 \end{bmatrix}, C_{12} = \begin{bmatrix} \frac{2}{0.75} & \frac{1}{0.75} \\ \frac{1}{0.75} & \frac{2}{0.75} \end{bmatrix}, C_{21} = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix}, C_{22} = \begin{bmatrix} \frac{1}{0.75} & 1 \\ 1 & \frac{1}{0.75} \end{bmatrix}.$$

(a) Approximate each state using a Gaussian distribution. Derive the expression. (2p)
Hint: Let us model $f_{X|S}(x|i) = \mathcal{N}(m_i, C_i)$. Find m_i, C_i for $i = 1, 2$.
Solution: See $m_i = E[X|S = i] = 0.25 \times \mathbf{0} + 0.75 \times \mathbf{0} = \mathbf{0}$. Next

$$C_1 = 0.25 \times C_{11} + 0.75 \times C_{12} = \begin{bmatrix} 4 & 1 \\ 1 & 4 \end{bmatrix},$$
$$C_2 = 0.25 \times C_{21} + 0.75 \times C_{22} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

(b) A two-category classifier receives L consecutive source samples $(X(0), \cdots, X(L-1))$ and then guesses the state of the signal source. Design this classifier for minimum error probability using computed m_i, C_i. (3p)
Hint: Design a discriminant function to produce a single design variable, such that optimal classification can be obtained by a simple threshold mechanism with this feature variable as input.
Solution: As both source alternatives are equally probable, we use the Maximum Likelihood decision rule. We can define a single discriminant function simply as

$$\begin{aligned}
g(\underline{\mathrm{x}}) &= \ln f_{X|S}(x|2) - \ln f_{X|S}(x|1) \\
&= \frac{1}{2}\sum_{n=0}^{L-1} \mathbf{x(n)} C_1^{-1} \mathbf{x(n)} - \frac{1}{2}\sum_{n=0}^{L-1} \mathbf{x(n)} C_2^{-1} \mathbf{x(n)} + \frac{1}{2} L \ln\left(\frac{\det(C_1)}{\det(C_2}\right) \\
&= \frac{1}{2}\sum_{n=0}^{L-1}\left(\frac{4}{15}x_1(n)^2 + \frac{4}{15}x_2(n)^2 - \frac{2}{15}x_1(n)x_2(n)\right) \\
&- \frac{1}{2}\sum_{n=0}^{L-1}\left(\frac{2}{3}x_1(n)^2 + \frac{2}{3}x_2(n)^2 - \frac{2}{3}x_1(n)x_2(n)\right) + \frac{1}{2} L \ln\left(\frac{15}{3}\right) \\
&= -\sum_{n=0}^{L-1}\left(\frac{1}{5}x_1(n)^2 + \frac{1}{5}x_2(n)^2 - \frac{4}{15}x_1(n)x_2(n)\right) + \frac{1}{2} L \ln 5
\end{aligned} \tag{1}$$

2 Determine for each of the following statements whether it is *true* or *false*, and give a brief argument for your choice: (1p each) (5p)

(a) A decision function $d(x)$ maps any observed feature vector x into a discrete output value representing the result of the classification.

Solution: TRUE

(b) An HMM with only one state may be represented by a GMM.

Solution: TRUE

(c) Bayes minimum-risk decision rule and ML decision rule always have two different solutions.

Solution: FALSE

(d) Expectation-maximization algorithm always converges to a globally optimum solution.

Solution: FALSE

(e) Evaluation of backward algorithm in HMM requires evaluation of forward algorithm, and vice-versa.

Solution: FALSE

3 (Conceptual questions) Answer the following questions. Each answer should be with in few sentences. (1p each) (5p)

(a) A process p depends on some model parameters θ. The process p has an input x and output y. We observe x and y, and would like to estimate the process p. What is an optimal cost function, and how to estimate the process in a machine learning framework?

Solution: EM.

(b) A source is generating a sequence of scalars $\underline{X} = [X_1\, X_2\, X_3, \ldots\, X_n]$, where n is very large, and the scalars are correlated. The source is modeled using an Gaussian model $\lambda = \{\mathbf{m}_{\underline{X}}, \mathbf{C}_{\underline{X}}\}$. Our task is evaluation of $P(\underline{X}|\lambda)$ where such a large n is found to be computationally demanding. Suggest an approximate method to compute $P(\underline{X}|\lambda)$.

Solution: Divide the $\underline{X}$ into some blocks and then assume independence.

(c) In pattern recognition and machine learning, explain the motivations of using exponential distributions.

Solution: (1) Then we can use logarithm to make the analytical expressions much tractable.

(d) HMM is used for several pattern recognition tasks, such as speech recognition, genetic sequence identification, etc. Why HMM is suitable for such kind of applications?

Solution: Speech sounds may have a same pattern, but of different lengths. HMM is the only tool that can handle sequences of unequal lengths.

(e) Why and when ML decision rule is preferred over MAP rule?

Solution: (1) MAP rule comes from Bayes Minimum decision rule for a specific loss matrix. ML is a specific case of MAP. (2) The parameters for MAP rule is unknown and random, which is more general than ML. So, if there is a chance of modeling the randomness, and still remains tractable, then we should use MAP. Otherwise, if data is limited or we do not have much a-priori knowledge, then we should use ML.

4 (Expectation Maximization) To check speed of cars, a series of speed measurements are made with speed checkers placed at various locations on the main road in front of KTH. The speed measurements are listed as a sequence $x = \{x_1, ..., x_k, ..., x_L\}$. Each measurement or sample x_k in this list is assumed to be an outcome of a random variable X_k which has a normal (Gaussian) distribution $\mathcal{N}(\mu_k, \sigma^2)$ with different means and same variance σ^2. Each sample is statistically independent of the other samples in the sequence.

The natural mean speed $\{\mu_k\}$ is known from previous similar measurements. The purpose of the new measurements x is to estimate the unknown variance σ^2. The measurements are sometimes disturbed by spurious transients due to engineering malfunction of speed checkers when huge vehicles cause high pressure on the speed checkers. These disturbances can occur with a known probability d for any measurement. If a disturbance happened to occur at measurement number k, the mean μ_k is not changed, but the variance of x_k is $\alpha^2\sigma^2$ instead of just σ^2, with a known value of α. Apply the EM algorithm for an estimate of σ. (5p)
Hint: You need to write EM help function in terms of all necessary parameters and then maximize the help function via a standard approach. We represent the possibility of disturbances by a hidden random state variable sequence $S = (S_1, ..., S_L)$, with $S_k = 1$ if a disturbance occurred at location k, and $S_k = 0$ otherwise.

Solution: Conditional probablity that a disturbance occured at location k, given the measurement result, is
$\gamma_1(x_k) = P(S_k = 1 | X_k = x_k, \sigma) = \frac{d \frac{1}{\alpha\sigma\sqrt{2\pi}} e^{-(x_k-\mu_k)^2/2(\alpha^2\sigma^2)}}{d \frac{1}{\alpha\sigma\sqrt{2\pi}} e^{-(x_k-\mu_k)^2/2(\alpha^2\sigma^2)} + (1-d) \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_k-\mu_k)^2/2(\sigma^2)}}$ and $\gamma_0(x_k) = 1 - \gamma_1(x_k)$. The EM functon is

$$\begin{aligned}
Q(\sigma', \sigma) &= E_S(\ln P(x, S|\sigma')|x, \sigma) \\
&= \sum_{k=1}^{L} \sum_{i_1=0}^{1} \sum_{i_2=0}^{1} \cdots \sum_{i_L=0}^{1} P(S|x, \sigma') \ln P(X_k = x_k \cap S_k = i_k|\sigma) \\
&= \sum_{k=1}^{L} \sum_{i_k=0}^{1} P(S_k|x_k, \sigma') \ln P(X_k = x_k \cap S_k = i_k|\sigma) \\
&= \sum_{k=1}^{L} \gamma_0(x_k) \ln \frac{1-d}{\sqrt{2\pi}\sigma} - \gamma_0(x_k) \frac{(x_k - \mu_k)^2}{2\sigma^2} \\
&+ \gamma_1(x_k) \ln \frac{d}{\sqrt{2\pi}\sigma\alpha} - \gamma_1(x_k) \frac{(x_k - \mu_k)^2}{2(\alpha\sigma)^2}
\end{aligned} \tag{2}$$

$$\frac{\partial Q}{\partial \sigma} = \sum_{k=1}^{L} -\gamma_0(x_k)/\sigma + 2\gamma_0(x_k)\frac{(x_k-\mu_k)^2}{2\sigma^3} + -\gamma_1(x_k)/(\alpha\sigma) + 2\gamma_1(x_k)\frac{(x_k-\mu_k)^2}{2\alpha^2\sigma^3} = 0 \tag{3}$$

$$\sum_{k=1}^{L} \gamma_0(x_k)/\sigma + \gamma_1(x_k)/(\alpha\sigma) = \sum_{k=1}^{L} 2\gamma_0(x_k)\frac{(x_k-\mu_k)^2}{2\sigma^3} + 2\gamma_1(x_k)\frac{(x_k-\mu_k)^2}{2\alpha^2\sigma^3} \tag{4}$$

or,

$$\sigma^2 = \frac{\sum_{k=1}^{L} \gamma_0(x_k)(x_k-\mu_k)^2 + \gamma_1(x_k)\frac{(x_k-\mu_k)^2}{\alpha^2}}{\sum_{k=1}^{L} \gamma_0(x_k) + \gamma_1(x_k)/(\alpha)} \tag{5}$$

5 (Bayesian Learning) Based on extensive statistical measurements, your heating provider has guaranteed that the heating supply in winter has a mean temperature of μ_v, with a standard deviation σ_v, across all their customers in their homes in Stockholm city. You want to measure the actual temperature in your home.

(a) Your thermometer shows the true temperature $\pm$ a random error with standard deviation σ when used in the range required for this measurement. You measure the temperature N times and obtain a temperature measurement sequence $x = (x_1, ..., x_N)$. Assume that all random variables involved are Gaussian. Formulate and evaluate a Bayesian estimate of the actual temperature using the a-priori knowledge and the N samples. (3p)

Solution:

$$f_v(v) \propto e^{-(v-\mu_v)^2/\sigma_v^2}$$

$$f_{X_i|V}(x_i) \propto e^{-(x-v)^2/\sigma^2}$$

$$f_{X_1,\cdots,X_N,V}(x_1,\cdots,x_N,v) \propto e^{-\sum_{i=1}^{N}(x_i-v)^2/\sigma^2} e^{-(v-\mu_v)^2/\sigma_v^2} \tag{6}$$

Then, the posterior probabilty

$$f_{V|X_1,\cdots,X_N}(v|x) \propto e^{-(v-\mu_p)^2/\sigma_p^2}$$

where, by collecting terms dependent on v from (6), we get that

$$\frac{1}{\sigma_p^2} = \frac{1}{\sigma_v^2} + \sum_{i=1}^{N}\frac{1}{\sigma^2}, \text{and}$$

$$\frac{\mu_p}{\sigma_p^2} = \frac{\mu_v}{\sigma_v^2} + \sum_{i=1}^{N}\frac{x_i}{\sigma^2}$$

(b) Evaluate a Bayesian learning based estimate of the actual temperature using your a-priori knowledge (i.e., the given guarantee) and measurements from two thermometers as follows: Thermometer 1 measures samples $(x_1, ..., x_N)$ with variance σ_1^2, and Thermometer 2 measures samples $(x_{N+1}, ..., x_{2N})$ with variance σ_2^2. (2p)

Solution:

$$f_v(v) \propto e^{-(v-\mu_v)^2/\sigma_v^2} \tag{7}$$

For Thermometer 1, we have $f_{X_i|V}(x_i) \propto e^{-(x-v)^2/\sigma_1^2}$, and for Thermometer 2, we have $f_{X_i|V}(x_i) \propto e^{-(x-v)^2/\sigma_2^2}$.

$$f_{X_1,\cdots,X_{2N},V}(x_1,\cdots,x_{2N},v) \propto e^{-\sum_{i=1}^{N}(x_i-v)^2/\sigma_1^2} e^{-\sum_{i=N+1}^{2N}(x_i-v)^2/\sigma_2^2} e^{-(v-\mu_v)^2/\sigma_v^2} \tag{8}$$

Then, the posterior probabilty

$$f_{V|X_1,\cdots,X_{2N}}(v|x) \propto e^{-(v-\mu_p)^2/\sigma_p^2} \tag{9}$$

where, by collecting terms dependent on v from (8), we get that

$$\begin{aligned} \frac{1}{\sigma_p^2} &= \frac{1}{\sigma_v^2} + \sum_{i=1}^{N} \frac{1}{\sigma_1^2} + \sum_{i=N+1}^{2N} \frac{1}{\sigma_2^2} \text{ and} \\ \frac{\mu_p}{\sigma_p^2} &= \frac{\mu_v}{\sigma_v^2} + \sum_{i=1}^{N} \frac{x_i}{\sigma_1^2} + \sum_{i=N+1}^{2N} \frac{x_i}{\sigma_2^2} \end{aligned} \tag{10}$$