KTH VETENSKAP OCH KONST

KTH Electrical Engineering

Exam for Pattern Recognition EQ2340 and EN2202 and for partial fulfillment of EO3274

Date:	Thursday Oct 27, 2016, 8:00 – 13:00
Place:	Q1.
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:	Wednesday, Nov 23, 2016.
Review:	Via scanned version
Contact:	Saikat: 0738913581; Arun: 0707755587
	Good Luck!

1 A signal source randomly selects a state $S = 1$ or $S = 2$ with equal probability. State 1 consists of signal source generating stationary zero mean two-dimensional Gaussian noise $X(n) = (X_1(n),\, X_2(n))^\top$ with covariance matrix $C = \mathcal{E}(X(n)X(n)^\top|1) = \begin{pmatrix} 4 & 1 \\ 1 & 4 \end{pmatrix}$. The signal is assumed to be uncorrelated over time instant n, that is, $\mathcal{E}(X(n_1)X(n_2)^\top) = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$, $n_1 \neq n_2$. In state 2, the source generates two dimensional vectors, such that each component $X_i(n),\ i \in \{1, 2\}$ takes values as follows:

$$\begin{aligned} X_i(n) &= 1, \text{ with probability } p_i \\ X_i(n) &= 0, \text{ with probability } 1 - p_i. \end{aligned} \tag{1}$$

A two-category classifier receives and stores L consecutive noise samples $(X(0), \cdots, X(L-1))$ and then guesses the state of the signal source.

(a) Design this classifier for minimum error probability with $p_1 = p_2 = 0.25$. (3p)
Hint: Design an appropriate discriminant function.
Solution: As both source alternatives are equally probable, we use the Maximum Likelihood decision rule. We can define a single discriminant function simply as

$$\begin{aligned} g(\underline{\mathrm{x}}) &= \ln f_{\underline{\mathrm{X}}|S}(\underline{\mathrm{x}}|2) - \ln f_{\underline{\mathrm{X}}|S}(\underline{\mathrm{x}}|1) \\ &= \frac{1}{2}\sum_{n=0}^{L-1} \mathbf{x(n)} C_1^{-1} \mathbf{x(n)} - \frac{1}{2}\sum_{n=0}^{L-1}\sum_{i=1}^{2} x_i(n)\log(p_i) + (1 - x_i(n))\log(1 - p_i) + \frac{1}{2} L \ln\left(2\pi \det(C)\right) \end{aligned} \tag{2}$$

(b) What is the conditional probability that source $S = 1$ was active given observations $X(0) = (0, 0)^\top$, $X(1) = (1, 0)^\top$, and $X(2) = (0, 1)^\top$, that is, $P(S = 1|\underline{\mathrm{X}} = ((0, 0)^\top, (1, 0)^\top, (0, 1)^\top)$? (2p)

Solution: We compute the log likelihood values:

$$L_1 = \ln f_{\underline{\mathrm{X}}|S}(\underline{\mathrm{x}}|i) = -\frac{1}{2}\sum_{n=0}^{L-1} \mathbf{x(n)} C^{-1} \mathbf{x(n)} - \frac{L}{2}\ln(2\pi \det C) \tag{3}$$

$$L_2 = \frac{1}{2}\sum_{n=0}^{L-1}\sum_{i=1}^{2} x_i(n)\log(p_i) + (1 - x_i(n))\log(1 - p_i) \tag{4}$$

Omitting common factors and substituting values, we have

$$L_1 = ? \tag{5}$$

and

$$L_2 = ? \tag{6}$$

then the conditional probability for $S = 1$ is:
$P(S = 1|\underline{\mathrm{X}} = ((0, 0)^\top, (1, -1)^\top, (-1, 1)^\top) = \dfrac{e^{L1}}{e^{L_1} + e^{L_2}} \approx 0?$

2 Determine for each of the following statements whether it is *true* or *false*, and give a brief argument for your choice: (1p each) (10p)

(a) Viterbi decoding is a globally optimum solution.
Solution: TRUE

(b) Baum-Welch algorithm is more intuitively driven, but still gives same solution of EM applied to HMM learning.
Solution: TRUE

(c) ML and MAP rules are specific cases of Bayes minimum-risk decision rule.
Solution: TRUE

(d) A first order Markov modeling means only having 1-step time wise correlation between feature vectors.
Solution: FALSE

(e) Uncorrelated implies independence between random variables.
Solution: FALSE

(f) A GMM can be generated from HMM.
Solution: TRUE

(g) If $\mathbf{x}$ and $\mathbf{n}$ are uncorrelated multivariate Gaussian distributions, then $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}$ is also uncorrelated distribution.

Solution: FALSE

(h) A Gaussian distribution is conjugate to another Gaussian distribution.
Solution: TRUE

(i) HMM is a discriminative model.
Solution: FALSE

(j) A binary classification problem can be always optimally solved using a linear discriminant.
Solution: FALSE

3 (Expectation Maximization) To obtain a good estimate of a bridge condition, a series of load measurements are made at different locations on a bridge. The results are listed as a sequence $\mathbf{x} = (x_1, \cdots, x_T)$. Each sample in this list is assumed to be an outcome of a random variable X_t which has a normal (Gausssian) distribution $N(\mu, \sigma^2)$. Each sample is statistically independent of the other samples in the sequence. The natural mean load of the bridge is known to be μ from previous measurements. Under normal conditions, it is known that an unbiased estimate for the variance would simply have been the sample variance

$$\sigma^2 = \frac{1}{T} \sum_t (x_t - \mu)^2. \tag{7}$$

However, the measurements are sometimes disturbed by presence of strong winds. These disturbances can occur with an unknown small probability d at any time during the measurement. If a disturbance happened to occur at measurement number t, the variance is not changed, but the mean of X_t is $\alpha\mu$ instead of just μ, with a known value α. What should be a good estimate of the variance σ^2 and disturbance probability d? Apply the EM algorithm to derive a formal update rule to improve previous estimate $\{\hat{\sigma}^2, \hat{d}\}$ by searching for a new estimate $\{\hat{\sigma}^{2\prime}, \hat{d}'\}$ that maximizes relevant help function. (5p)
Hint: You may treat σ^2 itself as a single parameter in your calculations, instead of using σ directly. We represent the possibility of disturbances by a hidden random state variable sequence $S = (S_1, \cdots, S_T)$ with $S_t = 1$ if a disturbance occurred at time t, and $S_t = 0$ otherwise.

Solution: The help function is

$$Q(\hat{\sigma}^{2\prime}, \hat{\sigma}^2) = E_s[\ln P(x, S|\hat{\sigma}^{2\prime})|x, \hat{\sigma}^2]. \tag{8}$$

The conditional probability that a disturbance occurred at time t, given the measurement result, is

$$\begin{aligned}
\gamma_1(x_t) &= P(S_t = 1|X_t = x_t, \hat{\sigma}^2) = \frac{d\mathcal{N}(x_t|\alpha\mu, \hat{\sigma}^2)}{(1-d)\mathcal{N}(x_t|\hat{\mu}, \hat{\sigma}^2) + d\mathcal{N}(x_t|\alpha\mu, \hat{\sigma}^2)} \\
\gamma_0(x_t) &= 1 - \gamma_1(x_t)
\end{aligned}$$

The EM function is

$$\begin{aligned}
Q(\hat{\sigma}^{2\prime}, \hat{\sigma}^2) &= E_{\underline{\mathrm{S}}}[\ln P(\underline{\mathrm{x}}, \underline{\mathrm{S}}|\hat{\sigma}^{2\prime})|\underline{\mathrm{x}}, \hat{\sigma}^2] = \\
&= \sum_{i_1=0}^{1} \cdots \sum_{i_T=0}^{1} P(S_1 = i_1 \ldots S_T = i_T|x_1 \ldots x_T, \hat{\sigma}^2) \cdot \\
&\quad \cdot \ln P(x_1 \ldots x_T, \underline{\mathrm{S}} = (i_1 \ldots i_T)|\hat{\sigma}^{2\prime})
\end{aligned}$$

Because of the statistical independence between samples we have

$$\ln P(x_1 \dots x_T, \underline{\mathrm{S}} = (i_1 \dots i_T)|\hat{\sigma}^{2\prime}) = \sum_{t=1}^{T} \ln P(X_t = x_t \cap S_t = i_t|\hat{\sigma}^{2\prime})$$

Therefore,

$$\begin{aligned}
Q(\hat{\sigma}^{2\prime}, \hat{\sigma}^2) &= E_{\underline{\mathrm{S}}}[\ln P(\underline{\mathrm{x}}, \underline{\mathrm{S}}|\hat{\sigma}^{2\prime})|\underline{\mathrm{x}}, \hat{\sigma}^2] = \\
&= \sum_{t=1}^{T} \sum_{i_1=0}^{1} \dots \sum_{i_T=0}^{1} P(S_1 = i_1 \dots S_T = i_T|x_1 \dots x_T, \hat{\sigma}^2) \cdot \\
&\quad \cdot \ln P(X_t = x_t \cap S_t = i_t|\hat{\sigma}^{2\prime}) = \\
&= \sum_{t=1}^{T} \sum_{i_t=0}^{1} P(S_t = i_t|x_t, \hat{\sigma}^2) \ln P(X_t = x_t \cap S_t = i_t|\hat{\sigma}^{2\prime}) = \\
&= \sum_{t=1}^{T} \gamma_0(x_t) \ln(1-d)\mathcal{N}(x_t|\mu, \hat{\sigma}^{2\prime}) + \gamma_1(x_t) \ln d\mathcal{N}(x_t|\alpha\mu, (\hat{\sigma}^{2\prime})) = \\
&= \sum_{t=1}^{T} \gamma_0(x_t) \ln \frac{(1-d)}{\sqrt{2\pi\hat{\sigma}^{2\prime}}} - \gamma_0(x_t)\frac{(x_t-\mu)^2}{2\hat{\sigma}^{2\prime}} + \\
&\quad + \gamma_1(x_t) \ln \frac{d}{\sqrt{2\pi\hat{\sigma}^{2\prime}}} - \gamma_1(x_t)\frac{(x_t-\alpha\mu)^2}{2\hat{\sigma}^{2\prime}}
\end{aligned}$$

The differentiate Q wrt to $\hat{\sigma}^{2\prime}$ and d

4 (Bayesian Learning) Let us consider a scenario where a group of people is entering Stockholm city via Arlanda airport. In the immigration check counter, the airport authority wants to know whether they are affected by Zika virus infection. In case of Zika virus infection, the concerned people has 4 major symptoms like fever, rash, body pain and headache. Let symptom vector is a 4-dimensional vector. Typically hospitals say that if all people in a group have Zika infection then the symptom vector is distributed as a multivariate Gaussian $\mathcal{N}(\boldsymbol{\mu}_1, \sigma_1^2\mathbf{I})$. On the other hand if nobody in a group is infected then the symptom vector is distributed as $\mathcal{N}(\boldsymbol{\mu}_2, \sigma_2^2\mathbf{I})$. Prior probability of being infected or not is equal.

(a) Let a person is subjected to medical check via collecting symptom vector N times and obtain a symptom sequence $\mathbf{X} = [\mathbf{x}_1\, \mathbf{x}_2\, \dots, \mathbf{x}_N]$, where $\mathbf{x}_n$ is the symptom vector at n'th time. Assume there is no statistical correlation in time between symptom vectors to have a simplistic model. For collecting a symptom vector, medical devices have additive measurement error term that is distributed as $\mathcal{N}(\mathbf{0}, \sigma_e^2\mathbf{I})$. Formulate and describe a Bayesian learning method to estimate the true symptom vector. (4p)

Solution: Vector version of the following. Also products as the things are uncorrelated Gaussian.

$$f_v(v) \propto e^{-(v-\mu_1)^2/\sigma_1^2} + e^{-(v-\mu_2)^2/\sigma_2^2}$$

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

$$f_{X_1,\cdots,X_N,V}(x_1,\cdots,x_N,v) \propto e^{-\sum_{i=1}^{N}(x_i-v)^2/\sigma^2} e^{-(v-\mu_v)^2/\sigma_v^2} \tag{9}$$

Then, the posterior probabilty

$$f_{V|X_1,\cdots,X_N}(v|x) \propto e^{-(v-\mu_p 1)^2/\sigma_p 1^2} + e^{-(v-\mu_p 2)^2/\sigma_p 2^2}.$$

(b) Using predictive distribution, describe a method to decide whether a person is infected or not. (1p)

Solution: We need to find predictive distribution of state S

$$p(S|\mathbf{X}) = \int_{\mathbf{V}} p(S|\mathbf{V})p(\mathbf{V}|\mathbf{X})d\mathbf{V} \tag{10}$$

where $p(V|X)$ is found earlier. $p(S|V) = \frac{p(V|S)p(S)}{p(V)}$