KTH Electrical Engineering

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

Date:	Friday Oct 27, 2017, 14:00 – 19:00
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 marks.
Language:	English.
Results:	Wednesday, Nov 23, 2016.
Review:	Via scanned version
Contact:	Saikat: 0738913581; Arun: 0707755587; Sina: 0739946726; Dong: 0767103249
	Good Luck!

1 (Causality in stock market) Suppose a company is observing two shares in a stock market for a period of time. You were asked to investigate the effect of them on each other. Assume two time series of length N, $\{X_{11}, \ldots, X_{1N}\}$ and $\{X_{21}, \ldots, X_{2N}\}$ to be the value of shares in time. In a linear regime, we are expressing the data with an auto-regressive model as bellow:

$$\vec{X}_i = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \vec{X}_{i-1} + \vec{W}_i$$

where $\vec{X}_i = \begin{bmatrix} X_{1i} \\ X_{2i} \end{bmatrix}$ and $\vec{W}_i = \begin{bmatrix} W_{1i} \\ W_{2i} \end{bmatrix}$.

Two find the effect of one share to another, consider a scenarios S_1 where second share is not influencing on the first one, i.e. $b_1 = 0$. In a different scenario S_2, $b_1 \neq 0$ indicating the influence of second share on the first one.

Assume $\vec{W}_i$ to be i.i.d. Gaussian random vector with zero mean and covariance matrix C, find discriminant function for your decision making task in parts (a) and (b).

(a) $C = \sigma^2 I$ and $P(S_1) = P(S_2) = 0.5$. (3p)

Solution: Since there is no prior on systems, we use ML estimate. We have:

$$\begin{aligned} P(\vec{X}_1, \ldots, \vec{X}_N | S) &= \prod_{i=1}^{N} P(\vec{X}_i | \vec{X}_{i-1}, S) \\ &= \frac{1}{2\pi\sqrt{\det(C)}} \exp\left(- \frac{\sum_{i=1}^{N} (\vec{X}_i - A\vec{X}_{i-1})^T \, C^{-1} \, (\vec{X}_i - A\vec{X}_{i-1})}{2} \right), \end{aligned}$$

and decision rule is:

$$d = \begin{cases} 1 & P(\vec{X}_1, \ldots, \vec{X}_N | S_1) > P(\vec{X}_1, \ldots, \vec{X}_N | S_2) \\ 2 & otherwise \end{cases}$$

$$P(\vec{X}_1, \ldots, \vec{X}_N | S_1) = \frac{1}{2\pi\sigma^2} \exp\left(- \frac{\sum_{i=1}^{N} (\vec{X}_i - A_1\vec{X}_{i-1})^T \, (\vec{X}_i - A_1\vec{X}_{i-1})}{2\sigma^2} \right)$$

$$\frac{1}{2\pi\sigma^2} \exp\left(- \frac{\sum_{i=1}^{N} (X_{1i} - a_1 X_{1i-1})^2 + (X_{2i} - a_2 X_{1i-1} - b_2 X_{2i-1})^2}{2\sigma^2} \right).$$

Similarly,

$$P(\vec{X}_1, \ldots, \vec{X}_N | S_2) =$$

$$\frac{1}{2\pi\sigma^2} \exp\left(- \frac{\sum_{i=1}^{N} (X_{1i} - a_1 X_{1i-1} - b_1 X_{2i-1})^2 + (X_{2i} - a_2 X_{1i-1} - b_2 X_{2i-1})^2}{2\sigma^2} \right),$$

so decision rule becomes:

$$d = \begin{cases} 1 & \sum_{i=1}^{N} (X_{1i} - a_1 X_{1i-1} - b_1 X_{2i-1})^2 > \sum_{i=1}^{N} (X_{1i} - a_1 X_{1i-1})^2 \\ 2 & otherwise \end{cases}$$

(b) $C = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ and $3P(S_1) = 2P(S_2) = 1.2$. Also assume $2a_1 = -a_2 = b_2 = 2$ and in system model S_2, $b_1 = 1$. (2p)

Solution: Now since the systems are biased you should take $\log \frac{P(S_1)}{P(S_2)}$ into account. Also note that the variances are different.

$$d = \begin{cases} 1 & \sum_{i=1}^{N}(X_{1i} - a_1 X_{1i-1} - b_1 X_{2i-1})^2 - \sum_{i=1}^{N}(X_{1i} - a_1 X_{1i-1})^2 > \log \frac{P(S_2)}{P(S_1)} \\ 2 & otherwise \end{cases}$$

(c) Suppose you have observed shares as:

$$Share\,1 : \{1, 1, 2, 1, 1\} \quad Share\,2 : \{0, 2, 2, 1, 0\},$$

what is the probability of $P(S_1|\text{Observed data})$ using the parameters in (b). (2p)

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

(a) For a random vector $\boldsymbol{x}$, there exists a linear transform matrix A, which can transform $\boldsymbol{x}$ into $\boldsymbol{y}$ ($\boldsymbol{y} = A\boldsymbol{x}$) so that the covariance matrix of $\boldsymbol{y}$ is diagonal.
Solution: True.

(b) Bayes Minimum-Risk Decision Rule is equivalent to MAP rule if the loss matrix has form as

$$L(d = i|s = j) = \begin{cases} 0, & i \neq j \\ r, & \text{otherwise,} \end{cases}$$

where r is positive constant.
Solution: True, r is not necessary to be 1 stated in textbook, as long as it is positive constant.

(c) MAP is equivalent to ML if source categories are chosen with equal probabilities (uniform priori probability).
Solution: True.

(d) In a reducible Markov model, all states are possible to be periodic.
Solution: False. In an reducible Markov model, some states could not possibly be entered once the model enters some states and gets locked there.

(e) When using Hidden Markov model for sequence classification, either Forward algorithm or Backward algorithm can only calculate the probability of state at time t conditioned on partial observed data sequence before or after time t, instead of full data sequence.
Solution: True. That is why both Forward and Backward algorithms are needed to calculated to probability of state conditioned on full data sequence.

(f) For a Markov model with transition matrix $A = \begin{pmatrix} 0.9 & 0.075 & 0.025 \\ 0.15 & 0.8 & 0.05 \\ 0.25 & 0.25 & 0.5 \end{pmatrix}$, it is irreducible and has the stationary state distribution $\boldsymbol{p} = (0.625, 0.3125, 0.0625)$.
Solution: True, $\boldsymbol{p} = \boldsymbol{p}A$

(g) Viterbi decoding is not a globally optimum solution.
Solution: False.

(h) It is possible to design classifier discriminant functions, that normally use all K elements of the feature vector, such that the classifier can allow one feature element to be missing but still make optimal use of the remaining features (although possibly with reduced performance)(copy of previous exam, ???)
Solution: TRUE. The discriminant functions only need to include a pre-designed variant that uses only $K - 1$ features.

(i) Expectation-maximization algorithm always has one solution, which is global optimum.
Solution: False.

(j) Using too complex model structure with too little training data may bring overfit, which is possible to be avoided by Bayesian Learning.
Solution: True.

3 (Life-time of lights) Assume that you are part of the testing and analysis team of a startup which manufactures light bulbs for Christmas. You are required to analyze a set of N bulbs and you are given the length of time for which each of these bulbs lasted. The life-times are listed as a sequence $\underline{\mathrm{x}} = (x_1, x_2, \cdots, x_N)$. The lifetime of all bulbs assumed to be the output of a random variable X_n which has an exponential distribution given by

$$f_X(x|\mu) = \frac{1}{\mu}\exp(-\frac{x}{\mu}).$$

Each sample is statistically independent of other samples in the set. The purpose of the analysis is to estimate the unknown average life-time parameter μ. Your startup has one major unit (S1) and one new manufacturing unit (S2). Hence, any given bulb comes from S1 with probability d, from S2 with probability $1-d$. S1 being an old unit is known to produce bulbs with lifetime of μ. S2 being new produces bulbs with lesser lifetime $\alpha\mu$, where $0 < \alpha < 1$ is known.

(a) (3p) Use the Expectation maximization rule to derive an update rule for the lifetime parameter μ.

Solution: Let $\gamma_1(x_n) = P(S_n = 1|X_n = x_n, \hat{\mu}) = \frac{\frac{d}{\mu}\exp(-x_n/\mu)}{\frac{d}{\mu}\exp(-x_n/\mu) + \frac{(1-d)}{\alpha\mu}\exp(-x_n/(\alpha\mu))}$ and $\gamma_2(x_n) = 1-\gamma_1(x_n)$

Then, the EM function is given by

$$\begin{aligned} Q(\hat{\mu}', \hat{\mu}) &= E_{\underline{\mathrm{S}}}[\ln P(\underline{\mathrm{x}}, \underline{\mathrm{S}}|\hat{\mu'})|\underline{\mathrm{x}}, \hat{\mu}] \\ &= \sum_{n=1}^{N}\sum_{i_n=1}^{2} P(S_n = i_n|x_n\mu)\ln P(X_n = \hat{x}_n, S_n = i_n|\hat{\mu}') \\ &= \sum_{n=1}^{N}\gamma_1(x_n)[\ln(d) - \ln\mu - \frac{x_n}{\mu}] + \gamma_2(x_n)[\ln((1-d)/\alpha) - \ln\mu - \frac{x_n}{\alpha\mu}] \end{aligned}$$

Setting derivative with respect to $\hat{\mu'}$:

$$-\frac{\sum_{n=1}^{N}\gamma_1(x_n)}{\hat{\mu}'} - \frac{\sum_{n=1}^{N}\gamma_2(x_n)}{\hat{\mu}'} + \frac{\sum_{n=1}^{N}\gamma_1(x_n)x_n}{\hat{\mu}'^2} + \frac{\sum_{n=1}^{N}\gamma_2(x_n)x_n}{\alpha\hat{\mu}'^2} = 0.$$

Therefore,

$$\hat{\mu'} = \frac{\sum_{n=1}^{N}\gamma_1(x_n)x_n + \frac{\sum_{n=1}^{N}\gamma_2(x_n)x_n}{\alpha}}{\sum_{n=1}^{N}\gamma_1(x_n) + \sum_{n=1}^{N}\gamma_2(x_n)}.$$

(b) (1p) What would be your estimate for the life-time if both units produce bulbs with the same life-time μ? That is when $\alpha = 1$.

Solution: Then it reduces to standard ML estimate of sample average $\mu' = (1/N)\sum_n x_n$.

4 (An AR(2) problem) Suppose the data X is generated with auto-regressive model bellow:

$$\begin{cases} X_0 = X_1 = 0 \\ X_t = AX_{t-1} + BX_{t-2} + cW_t \qquad t \geq 2 \end{cases}$$

where A and B are unknown and only a prior knowledge about their distribution is available. Assume W_t to have zero mean and variance 1, then answer the following questions:

(a) Compute the posterior distribution for A and B, if in the prior distribution A and B were independent Gaussian random variables of zero mean with variance σ_A^2 and σ_B^2 respectively. (2p)

Solution:

$$\begin{aligned} f(\vec{X} = \vec{x}, A = a, B = b) &= f(A = a, B = b) f(\vec{X} = \vec{x} | A = a, B = b) \\ &\propto f(A = a, B = b | \vec{X} = \vec{x}) \end{aligned}$$

where

$$\begin{aligned} f(\vec{X} = \vec{x} | A = a, B = b) &= \prod_{i=1}^{N} f(x_i | x_{i-1}, x_{i-2}, a, b) \\ &\propto \prod_{i=1}^{N} \exp(-\frac{(x_i - ax_{i-1} - bx_{i-2})^2}{2c^2}) \\ &= \exp(-\frac{\sum_{i=1}^{N} (x_i - ax_{i-1} - bx_{i-2})^2}{2c^2}). \end{aligned}$$

On the other hand,

$$f(A = a, B = b) \propto \exp(-0.5[a\,, b]\, \Sigma^{-1}\, [a\,, b]^T)$$

a) In this case we can simplify the joint distribution of A and B:

$$f(A = a, B = b) \propto \exp(-\frac{a^2}{2\sigma_A^2} - \frac{b^2}{2\sigma_B^2}).$$

So

$$\begin{aligned} & f(A = a, B = b | \vec{x}) \\ & \propto \exp(-[a^2(\frac{1}{2\sigma_A^2} + \frac{\sum x_{i-1}^2}{2c^2}) - 2a(\frac{\sum x_i x_{i-1}}{2c^2}) + b^2(\frac{1}{2\sigma_B^2} + \frac{\sum x_{i-2}^2}{2c^2}) - 2b(\frac{\sum x_i x_{i-2}}{2c^2}) + 2ab(\frac{\sum x_{i-1} x_{i-2}}{2c^2})]). \end{aligned}$$

To compare this result with distribution of joint Guassian we have:

$$\begin{aligned} & \exp(-0.5[a - \mu'_A\,, b - \mu'_B] \begin{bmatrix} p & q \\ q & r \end{bmatrix} [a - \mu'_A\,, b - \mu'_B]^T) \\ & \propto \exp(-0.5[a^2 p - 2a(p\mu'_A + q\mu'_B) + b^2 r - 2b(r\mu'_B + q\mu'_A) + 2abq]) \end{aligned}$$

So by comparison we have:

$$p = \frac{1}{\sigma_A^2} + \frac{\sum x_{i-1}^2}{c^2}$$
$$r = \frac{1}{\sigma_B^2} + \frac{\sum x_{i-2}^2}{c^2}$$
$$q = \frac{\sum x_{i-1}x_{i-2}}{c^2}$$

and μ'_A and μ'_B are derived by solving the following system of equations

$$\begin{cases} p\mu'_A + q\mu'_B = \frac{\sum x_i x_{i-1}}{c^2} \\ r\mu'_B + q\mu'_A = \frac{\sum x_i x_{i-2}}{c^2} \end{cases}$$

Then,

$$\Sigma' = \begin{bmatrix} p & q \\ q & r \end{bmatrix}^{-1}$$

(b) Repeat part (a) if in the prior $[A, B]^T \sim \mathcal{N}(\mathbf{0}, \Sigma)$ where $\Sigma = \begin{bmatrix} 2/3 & 1/3 \\ 1/3 & 2/3 \end{bmatrix}$. (2p)

Solution: If A, B were correlated,

$$f(a, b) \propto \exp\left(-0.5(2a^2 - 2ab + 2b^2)\right)$$

In this case:

$$p = 2 + \frac{\sum x_{i-1}^2}{c^2}$$
$$r = 2 + \frac{\sum x_{i-2}^2}{c^2}$$
$$q = 2 + \frac{\sum x_{i-1}x_{i-2}}{c^2}$$