STA 414/2104 — Answers to Third Test — 2012-04-05

Question 1: [30 marks total] We have two i.i.d. observations of seven variables, as follows:

$$5 \ 7 \ 8 \ 2 \ 3 \ 5 \ 2$$
$$3 \ 3 \ 6 \ 6 \ 1 \ 1 \ 0$$

a) [20 marks] Find a 7-dimensional vector of length one that points in the direction of the first principal component of this data. Explain how you obtained it.

First, we subtract the sample means from the two observed vectors, giving the following centred data:

$$\begin{matrix} 1 & 2 & 1 & -2 & 1 & 1 & 1 \\ -1 & -2 & -1 & 2 & -1 & -2 & -1 \end{matrix}$$

With only two training cases, each of these vectors must point in the direction of the first principal component. Taking the first, its length is 4, so one vector of length 1 in the direction of the first principal component is

$$\left[\frac{1}{4} \ \frac{1}{2} \ \frac{1}{4} \ -\frac{1}{2} \ \frac{1}{4} \ \frac{1}{2} \ \frac{1}{4} \right]^T$$

The other possible answer is the negation of the above.

It's also possible to answer this question by computing

$$XX^T \ = \ \begin{bmatrix} 16 & -16 \\ -16 & 16 \end{bmatrix}$$

and then finding its eigenvectors, $[1 \ -1]^T$ and $[1 \ 1]^T$, which have eigenvalues 32 and 0. PC1 is in the direction $X^T[1 \ -1]^T$. After scaling to unit length, this gives the same answer as above.

b) [10 marks] Find the projection on this principal component of the new observation shown below:

$$4 \ 1 \ 9 \ 3 \ 2 \ 2 \ 1$$

Subtracting the sample means from the training data gives $[0 \ -4 \ 2 \ -1 \ 0 \ -1 \ 0]^T$. The dot product of this with the PC1 vector from (a) is $-3/2$.

Question 2: [30 marks total] Recall that in a factor analysis model an observed data point, x, is modeled using M latent factors as

$$x = \mu + Wz + \epsilon$$

where μ is a vector of means for the p components of x, W is a $p \times M$ matrix, z is a vector of M latent factors, assumed to have independent $N(0, 1)$ distributions, and ϵ is a vector of p residuals, assumed to be independent, and to come from normal distributions with mean zero. The variance of ϵ_j is σ_j^2.

Suppose that $p = 5$ and $M = 2$, and that the parameters of the model are mean $\mu = [0\ 0\ 0\ 0\ 0]^T$, residual standard deviations $\sigma_1 = 1$, $\sigma_2 = 1$, $\sigma_3 = 2$, $\sigma_4 = 2$, $\sigma_5 = 2$, and

$$W = \begin{bmatrix} 1 & 2 \\ -1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

a) [20 marks] Find the covariance matrix for x. Show your work.

$$\text{Cov}(x) = E[(Wz + \epsilon)(Wz + \epsilon)^T] = WW^T + \text{diag}(\sigma_1^2, \ldots, \sigma_5^2)$$

$$= \begin{bmatrix} 6 & 1 & 1 & 1 & 2 \\ 1 & 3 & -1 & -1 & 1 \\ 1 & -1 & 5 & 1 & 0 \\ 1 & -1 & 1 & 5 & 0 \\ 2 & 1 & 0 & 0 & 5 \end{bmatrix}$$

b) [10 marks] Suppose that we don't observe vectors x of dimension five, but rather we observe vectors y of dimension four, where $y_1 = x_1$, $y_2 = 3x_2$, $y_3 = -x_3$, and $y_4 = 2x_4 + x_5$. Assuming that the distribution of x is given by the factor analysis model with parameters above, write down a factor analysis model (including values of its parameters) for the distribution of y.

Using the relation of y to x and the model for x above, we can write $y = W'z + \epsilon'$, where

$$W' = \begin{bmatrix} 1 & 2 \\ -3 & 3 \\ -1 & 0 \\ 2 & 1 \end{bmatrix}$$

The standard deviations of the ϵ'_i will be $\sigma'_1 = \sigma_1 = 1$, $\sigma'_2 = 3\sigma_2 = 3$, $\sigma'_3 = \sigma_3 = 2$, and $\sigma'_4 = \sqrt{4\sigma_4^2 + \sigma_5^2} = \sqrt{20}$.

Question 3: [40 marks total] Consider a two-component Gaussian mixture model for univariate data, in which the probability density for an observation, x, is

$$(1/2)N(x|\mu, 1) \;+\; (1/2)N(x|\mu, 2^2)$$

Here, $N(x|\mu, \sigma^2)$ denotes the density for x under a univariate normal distribution with mean μ and variance σ^2. Notice that mixing proportions are equal for this mixture model, that the two components have the same mean, and that the standard deviations of the two components are fixed at 1 and 2. There is only one model parameter, μ.

Suppose we wish to estimate the μ parameter by maximum likelihood using the EM algorithm. Answer the following questions regarding how the E step and M step of this algorithm operate, if we have the three data points below:

$$4.0, \; 4.6, \; 2.0$$

Here is a table of standard normal probability densities that you may find useful:

x	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	2.0
$N(x|0,1)$	.40	.40	.39	.38	.37	.35	.33	.31	.29	.27	.24	.22	.19	.17	.15	.13	.11	.09	.08	.07	.05

a) [20 marks] Find the responsibilities that will be computed in the E step if the model parameter estimates from the previous M step are $\mu = 4$, $\sigma_1 = 1$, and $\sigma_2 = 2$. Since the responsibilities for the two components must add to one, it is enough to give $r_{i1} = P(\text{component } 1 \,|\, x_i)$ for $i = 1, 2, 3$. Show your work.

First, note that the normal density function with mean μ and variance σ^2 is $N(x|\mu, \sigma^2) = (1/\sigma)N((x - \mu)/\sigma|0, 1)$. Also $N(-x|0, 1) = N(x|0, 1)$.

Using Bayes' Rule, we get that

$$P(\text{component } 1|x) \;=\; \frac{(1/2)N(x|\mu, 1)}{(1/2)N(x|\mu, 1) + (1/2)N(x|\mu, 2^2)}$$

Applying this the three observations, we get

$$r_11 \;=\; \frac{(1/2)0.40}{(1/2)0.40 + (1/2)(1/2)0.40} \;=\; 2/3$$

$$r_21 \;=\; \frac{(1/2)0.33}{(1/2)0.33 + (1/2)(1/2)0.38} \;=\; 33/52$$

$$r_31 \;=\; \frac{(1/2)0.05}{(1/2)0.05 + (1/2)(1/2)0.24} \;=\; 5/17$$

b) [20 marks] Using the responsibilities that you computed in part (a), find the estimate for μ that will be found in the next M step. Recall that the M step maximizes the expected value of the log of the probability density for x_1, x_2, x_3 and the unknown component indicators, with the expectation taken with respect to the distribution for the component indicators found in the previous E step. Show your work. Your final answer may be an arithmetic expression (with no symbols) rather than an actual number.

The expected log likelihood is

$$\sum_{i=1}^{3} \left[r_{i1}(-(1/2)(x_i - \mu)^2) \; + \; (1 - r_{i1})(-(1/2)(x_i - \mu)/2^2) \right]$$

To find the maximum of this with respect to μ, we take the derivative with respect to μ, which is

$$\sum_{i=1}^{3} \left[r_{i1}(x_i - \mu) \; + \; (1 - r_{i1})(x_i - \mu)/4 \right]$$

Setting this to zero and solving for μ gives

$$\hat{\mu} \;\; = \;\; \frac{\sum_{i=1}^{3} (r_{i1} + (1 - r_{i1})/4)\, x_i}{\sum_{i=1}^{3} (r_{i1} + (1 - r_{i1})/4)} \;\; = \;\; \frac{(3/4)4.0 + (151/208)4.6 + (25/68)2.0}{(3/4) + (151/208) + (25/68)}$$