Homework 5 Liu Yan

(a) Explanation

Step 1: 4 base time series are generated by drawing independently from a uniform distribution within $[-5,5]$.

Let's call each time series ts_1, ts_2, ts_3, ts_4.

Step 2: the first 5 time stamps of ts_1 is set to 0. ;

the second 5 time stamps of ts_2 is set to 0 ;

the third 5 time stamps of ts_3 is set to 0 ;

the fourth 5 time stamps of ts_4 is set to 0 .

Step 3: an assign probability vector is generated as

$\left(\frac{1}{1+2+3+4}, \frac{2}{1+2+3+4}, \frac{3}{1+2+3+4}, \frac{4}{1+2+3+4}\right)$

or $\left(\frac{1}{10}, \frac{2}{10}, \frac{3}{10}, \frac{4}{10}\right)$, which indicates the probability that a sample is assigned to a time series.

Step 4: 1000 samples with value 1,2,3,4 are generated with the assignment probability.

A sample matrix with 1000 rows, 20 columns are created, where each row's value is equal to the base time series it is drawn from.

Step 5: Random noise with the distribution $N(0,2)$ is added to each entry of the sample matrix.

(b) $Z = (Z_1, Z_2, \cdots, Z_n)$ follows multivariate normal with mean μ and covariance Σ.

$\hat{z}^{(i)} = (\hat{z}^{(1)}, \hat{z}^{(2)}, \cdots, \hat{z}^{(N)})$ is iid samples from Z, $i = 1, 2, \cdots, N$.

(i) Show $\hat{\mu}_{MLE} = \frac{1}{N}\sum_{i=1}^{N} \hat{z}^{(i)}$

Proof: Since the samples are iid, the joint density of the data given $\{\hat{z}^{(i)}, i = 1, 2, \cdots, N\}$ is the product of the individual density which is $\prod_{i=1}^{N} f_{Z^{(i)}}(\hat{z}^{(i)} \mid \mu, \Sigma)$. Take the logarithm, we have the log-likelihood function ℓ

$$\ell(\mu, \Sigma \mid \hat{z}^{(i)}) = \log \prod_{i=1}^{N} f_{z^{(i)}}(\hat{z}^{(i)} \mid \mu, \Sigma)$$

$$= \log \prod_{i=1}^{N} \frac{1}{\sqrt{(2\pi)^n \det \Sigma}} \exp\left(-\frac{1}{2}(\hat{z}^{(i)}-\mu)^T \Sigma^{-1} (\hat{z}^{(i)}-\mu)\right)$$

$$= \sum_{i=1}^{N} \left(-\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(\det\Sigma) - \frac{1}{2}(\hat{z}^{(i)}-\mu)^T \Sigma^{-1} (\hat{z}^{(i)}-\mu)\right)$$

$$= -\frac{nN}{2}\log(2\pi) - \frac{N}{2}\log(\det\Sigma) - \frac{1}{2}\sum_{i=1}^{N}(\hat{z}^{(i)}-\mu)^T \Sigma^{-1} (\hat{z}^{(i)}-\mu)$$

Then take the partial derivative of ℓ with respect to μ

Base on the matrix calculus identity below:

$$\frac{\partial w^T A w}{\partial w} = 2Aw \text{ if } w \text{ is } \cancel{\text{not}} \text{ dependent on } A \text{ and } A \text{ is symmetric.}$$

We know μ does not depend on Σ and Σ is symmetric. thus

$$\frac{\partial \ell(\mu,\Sigma)}{\partial \mu} = \sum_{i=1}^{N} \Sigma^{-1}(\mu - \hat{z}^{(i)}) = 0$$

Since Σ is positive definite, we have

$$\sum_{i=1}^{N}(\mu - \hat{z}^{(i)}) = 0 \Rightarrow N\mu - \sum_{i=1}^{N}\hat{z}^{(i)} = 0$$

$$\Rightarrow \hat{\mu}_{MLE} = \frac{1}{N}\sum_{i=1}^{N}\hat{z}^{(i)}$$

(ii) Show that $\hat{\Sigma}_{kj} = \frac{1}{N}\sum_{i=1}^{N}(z_k^{(i)} - \hat{\mu}_k)(z_j^{(i)} - \hat{\mu}_j)$

The proof below is based on the answer by Xavier on stackexchange and the ML book.

To derive the MLE estimator for Σ, we use the following linear algebra property:

1. The trace is invariant under cyclic permutations of matrix products

$$tr[ACB] = tr[CAB] = tr[BCA]$$

2. Since $x^T A x$ is scalar, its trace equals the value of itself, which is

$$x^T A x = tr(x^T A x) = tr(x^T x A)$$

3. $\frac{\partial}{\partial A} tr[AB] = B^T$

4. $\frac{\partial}{\partial A} \log|A| = A^{-T}$

Based on 1-4, we have

$$\frac{\partial}{\partial A} x^T A x = \frac{\partial}{\partial A} tr[x^T x A] = (x x^T)^T = x^{TT} x^T = x x^T$$

Let $C = -\frac{nN}{2}\log(2\pi)$ as a constant.

rewrite $\ell(\mu,\Sigma)$ and compute the derivative with respect to Σ

$$\ell(\mu,\Sigma \mid \hat{z}^{(i)}) = C - \frac{N}{2}\log(\det\Sigma) - \frac{1}{2}\sum_{i=1}^{N}(\hat{z}^{(i)}-\mu)^T\Sigma^{-1}(\hat{z}^{(i)}-\mu))$$

$$= C - \frac{N}{2}\log(\det\Sigma^{-1}) - \frac{1}{2}\sum_{i=1}^{N}\mathrm{tr}[(\hat{z}^{(i)}-\mu)(\hat{z}^{(i)}-\mu)^T\Sigma^{-1}]$$

$$\frac{\partial\ell}{\partial\Sigma} = \frac{N}{2}(\Sigma^{-1})^{-T} - \frac{1}{2}\sum_{i=1}^{N}(\hat{z}^{(i)}-\mu)(\hat{z}^{(i)}-\mu)^T)$$

Since Σ is symmetric, we have $(\Sigma^{-1})^{-T} = \Sigma$. Thus,

$$\frac{\partial\ell}{\partial\Sigma} = \frac{N}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{N}(\hat{z}^{(i)}-\mu)(\hat{z}^{(i)}-\mu)^T$$

Let $\frac{\partial\ell}{\partial\Sigma} = 0$, we have

$$\frac{N}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{N}(\hat{z}^{(i)}-\mu)(\hat{z}^{(i)}-\mu)^T = 0$$

$$\Rightarrow \hat{\Sigma}_{MLE} = \frac{1}{N}\sum_{i=1}^{N}(\hat{z}^{(i)}-\hat{\mu}_{MLE})(\hat{z}^{(i)}-\hat{\mu}_{MLE})^T.$$

which is equivalent to $\hat{\Sigma}_{kj} = \frac{1}{N}\sum_{i=1}^{N}(\hat{z}_k^{(i)}-\hat{\mu}_k)(\hat{z}_j^{(i)}-\hat{\mu}_j)$.