Xi'an Jiaotong-Liverpool University
西交利物浦大学

XJTLU Entrepreneur College (Taicang) Cover Sheet

Module code and Title	**DTS201TC Pattern Recognition**	
School Title	**School of AI and Advanced Computing**	
Assignment Title	**Final project**	
Submission Deadline	**23:59, 31st Dec.**	
Final Word Count		
If you agree to let the university use your work anonymously for teaching and learning purposes, please type **"yes"** here.		**Yes**

I certify that I have read and understood the University's Policy for dealing with Plagiarism, Collusion and the Fabrication of Data (available on Learning Mall Online). With reference to this policy I certify that:

- My work does not contain any instances of plagiarism and/or collusion.
  My work does not contain any fabricated data.

By uploading my assignment onto Learning Mall Online, I formally declare that all of the above information is true to the best of my knowledge and belief.

Scoring – For Tutor Use	
Student ID	**1930080**

Stage of Marking	**Marker Code**	**Learning Outcomes Achieved (F/P/M/D) (please modify as appropriate)**			**Final Score**
		A	**B**	**C**	
1st Marker – red pen					
Moderation – green pen	**IM Initials**	The original mark has been accepted by the moderator (please circle as appropriate):			Y / N
		Data entry and score calculation have been checked by another tutor (please circle):			Y
2nd Marker if needed – green pen					

For Academic Office Use			**Possible Academic Infringement (please tick as appropriate)**	
Date Received	**Days late**	**Late Penalty**	☐ **Category A**	Total Academic Infringement Penalty (A,B, C, D, E, Please modify where necessary) ______________________
			☐ **Category B**	
			☐ **Category C**	
			☐ **Category D**	
			☐ **Category E**	

DTS201TC Classification Demonstration

Project (Individual)

Yaqi Yu

1 Mathematical problems [40 marks]

Derive the Maximum Likelihood Estimate.

1.1 [20 marks]

Let $x_1, x_2, ..., x_N$ be vectors stemmed from a normal distribution with known covariance matrix and unknown mean, that is

$$p(x_k;\mu) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} exp(-\frac{1}{2}(x_k-\mu)^T\Sigma^{-1}(x_k-\mu)) \tag{1}$$

where D is the dimension of vector x_k $(k = 1, ..., N)$.

***TASK 1:* Derive the ML estimate of the mean μ.**

Solution:

For N available samples, we have

$$\begin{aligned} L(\mu) &= \ln\prod_{k=1}^{N} p(x_k;\mu) \\ &= \sum_{k=1}^{N}\ln[\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} exp(-\frac{1}{2}(x_k-\mu)^T\Sigma^{-1}(x_k-\mu))] \\ &= -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln(|\Sigma|) - \frac{1}{2}\sum_{k=1}^{N}(x_k-\mu)^T\Sigma^{-1}(x_k-\mu) \end{aligned} \tag{2}$$

We've known the derivation rule below(if A is a symmetric matrix)

$$\frac{\partial x^T Ax}{\partial x} = (A + A^T)x = 2Ax$$

For the covariance matrix Σ is a symmetric matrix, the inverse matrix Σ^{-1} is also symmetric.

Thus, we obtain:

$$
\begin{aligned}
\frac{\partial L}{\partial \mu} &= \frac{\partial(-\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln(|\Sigma|) - \frac{1}{2}\sum_{k=1}^{N}(x_k-\mu)^T\Sigma^{-1}(x_k-\mu))}{\partial \mu} \\
&= \frac{\partial(-\frac{1}{2}\sum_{k=1}^{N}(x_k-\mu)^T\Sigma^{-1}(x_k-\mu))}{\partial u} \\
&= \frac{\partial(-\frac{1}{2}\sum_{k=1}^{N}(x_k-\mu)^T\Sigma^{-1}(x_k-\mu))}{\partial(x_k-\mu)}\frac{\partial(x_k-\mu)}{\partial \mu} \\
&= \sum_{k=1}^{N}\Sigma^{-1}(x_k-\mu)
\end{aligned} \tag{3}
$$

We suppose that

$$
\frac{\partial L}{\partial \mu} = 0 = \sum_{k=1}^{N}\Sigma^{-1}(x_k-\mu)
$$

And we multiply Σ on both sides to get

$$
0 = \sum_{k=1}^{N}(x_k-\mu) \Rightarrow \mu = \frac{1}{N}\sum_{k=1}^{N}x_k \tag{4}
$$

So the MLE of the μ is : $\hat{\mu}_{ML} = \frac{1}{N}\sum_{k=1}^{N}x_k$

1.2 [20 marks]

Let $x_1, x_2, ..., x_N$ be vectors stemmed from a normal distribution with unknown mean μ and unknown convariance matrix Σ, that is

$$p(x_k; \mu, \Sigma) = \frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} exp(-\frac{1}{2}(x_k - \mu)^T \Sigma^{-1}(x_k - \mu)) \tag{5}$$

where D is the dimension of vector x_k $(k = 1, ..., N)$.

***TASK 2:* Derive the ML Estimate of μ and Σ.**

Solution:

For N available samples, we have

$$\begin{aligned} L(\mu, \Sigma) &= \ln \prod_{k=1}^{N} p(x_k; \mu) \\ &= \sum_{k=1}^{N} \ln[\frac{1}{(2\pi)^{D/2}|\Sigma|^{1/2}} exp(-\frac{1}{2}(x_k - \mu)^T \Sigma^{-1}(x_k - \mu))] \\ &= -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln(|\Sigma|) - \frac{1}{2}\sum_{k=1}^{N}(x_k - \mu)^T \Sigma^{-1}(x_k - \mu) \end{aligned} \tag{6}$$

We've known the derivation rule below(if A is a symmetric matrix)

$$\frac{\partial x^T A x}{\partial x} = (A + A^T)x = 2Ax$$

For the covariance matrix Σ is a symmetric matrix, the inverse matrix Σ^{-1} is also symmetric.

Thus, we obtain:

$$\begin{aligned} \frac{\partial L}{\partial \mu} &= \frac{\partial(-\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln(|\Sigma|) - \frac{1}{2}\sum_{k=1}^{N}(x_k - \mu)^T \Sigma^{-1}(x_k - \mu))}{\partial \mu} \\ &= \frac{\partial(-\frac{1}{2}\sum_{k=1}^{N}(x_k - \mu)^T \Sigma^{-1}(x_k - \mu))}{\partial u} \\ &= \frac{\partial(-\frac{1}{2}\sum_{k=1}^{N}(x_k - \mu)^T \Sigma^{-1}(x_k - \mu))}{\partial(x_k - \mu)} \frac{\partial(x_k - \mu)}{\partial \mu} \\ &= \sum_{k=1}^{N} \Sigma^{-1}(x_k - \mu) \end{aligned} \tag{7}$$

We suppose that

$$\frac{\partial L}{\partial \mu} = 0 = \sum_{k=1}^{N} \Sigma^{-1}(x_k - \mu)$$

And we multiply Σ on both sides to get

$$0 = \sum_{k=1}^{N}(x_k - \mu) \Rightarrow \mu = \frac{1}{N}\sum_{k=1}^{N} x_k \tag{8}$$

For Σ, we have a partial derivation rule that

$$\frac{\partial \ln(|X|)}{\partial X} = (X^{-1})^T$$

Derivation of the differential form of the determinant by partial derivation:

$$d\ln(|X|) = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\partial \ln(|X|)}{\partial X_{ij}}\mathrm{dX_{ij}} = \sum_{i=1}^{m}\sum_{j=1}^{n} X_{ji}^{-1}\mathrm{dX_{ij}} = tr(X^{-1}\mathrm{dX}) \tag{9}$$

So the differential rule that: $d\ln(|X|) = tr(X^{-1}dX)$

And we have another one derivation rule that:

$$0 = dI = d(XX^{-1}) = dXX^{-1} + XdX^{-1}$$

so:

$$dX^{-1} = -X^{-1}dXX^{-1}$$

$$L(\mu, \Sigma) = -\frac{ND}{2}\ln(2\pi) - \frac{N}{2}\ln(|\Sigma|) - \frac{1}{2}\sum_{k=1}^{N}(x_k - \mu)^T\Sigma^{-1}(x_k - \mu) \tag{10}$$

We apply the matrix differentiation

$$\begin{aligned} dL &= tr(dL) \\ &= -\frac{N}{2}tr(\Sigma^{-1}d\Sigma) - \frac{1}{2}\sum_{k=1}^{N} tr((x_k - \mu)^T d\Sigma^{-1}(x_k - \mu)) \\ &= -\frac{N}{2}tr(\Sigma^{-1}d\Sigma) - \frac{1}{2}\sum_{k=1}^{N} tr((x_k - \mu)^T(-\Sigma^{-1}d\Sigma\Sigma^{-1})(x_k - \mu)) \\ &= -\frac{N}{2}tr(\Sigma^{-1}d\Sigma) + \frac{1}{2}\sum_{k=1}^{N} tr(\Sigma^{-1}(x_k - \mu)(x_k - \mu)^T\Sigma^{-1}d\Sigma) \end{aligned} \tag{11}$$

for we have

$$df = tr(\frac{df}{\partial x}^T dx)$$

we can obtain

$$\frac{\partial L}{\partial \Sigma} = -\frac{N}{2}\Sigma^{-1} + \frac{1}{2}\sum_{k=1}^{N}\Sigma^{-1}(x_k - \mu)(x_k - \mu)^T\Sigma^{-1} \tag{12}$$

We suppose that $\frac{\partial L}{\partial \Sigma} = 0$, so

$$\Sigma = \frac{1}{N}\sum_{k=1}^{N}(x_k - \mu)(x_k - \mu)^T$$

So the MLE of the μ is : $\hat{\mu}_{ML} = \frac{1}{N}\sum_{k=1}^{N} x_k$;

The MLE of the Σ is : $\hat{\Sigma}_{ML} = \frac{1}{N}\sum_{k=1}^{N}(x_k - \mu)(x_k - \mu)^T$

2 Practical problems [60 marks]

I will present the code and comments of this part in another file.