Machine Learning	**Spring 2017**

Solution : Homework 4

Lecturer: Yang Yang — *Homework taker: Li Xu*

Due Time: June 9

Problem 1.

Solution:

Answer for problem (1):

We have

$$S = \frac{XX^T}{N}$$

As S is symmetric, it can be diagonalized:

$$S = ULU^T$$

where U is a matrix of eigenvectors($U = \{u_i\}$).

If

$$X = US'V^T$$

Then

$$S = U\frac{S'^2}{N}U^T$$

i.e. The principal components $\{u_i\}$ are columns of U.

Answer for problem (2):

SVD is better as XX^T can be very large if D >> N

■

Problem 2.

Solution:

Answer for problem (a):

i. roads salted, school cancellation

ii. none

iii. none

iv. temperature

Answer for problem (b):

$$\begin{aligned}
&p(\textbf{temperature}, \textbf{snow}, \textbf{roads salted}, \textbf{school cancellation}) = \\
&\qquad p(\textbf{temperature})p(\textbf{snow}|\textbf{temperature})p(\textbf{roads salted}|\textbf{snow}) \\
&\qquad\qquad p(\textbf{school cancellation}|\textbf{roads salted}, \textbf{snow})
\end{aligned}$$

Answer for problem (c):

$$\begin{aligned} p(\textbf{snow} = light) &= \sum_{\textbf{temperature}} p(\textbf{temperature})p(\textbf{snow} = light|\textbf{temperature}) \\ &= 0.208 \end{aligned}$$

$$\begin{aligned} p(\textbf{roads salted} = T, \textbf{snow} = light) &= p(\textbf{roads salted} = T|\textbf{snow} = light) * p(\textbf{snow} = light) \\ &= 0.1872 \end{aligned}$$

$$\begin{aligned} p(\textbf{roads salted} = F, \textbf{snow} = light) &= p(\textbf{roads salted} = F|\textbf{snow} = light) * p(\textbf{snow} = light) \\ &= 0.0208 \end{aligned}$$

$$\begin{aligned} p(\textbf{school cancellation} = T|\textbf{snow} = light) &= \frac{p(\textbf{school cancellation} = T, \textbf{snow} = light)}{p(\textbf{snow} = light)} \\ &= \frac{\sum_{\textbf{roads salted}} p(\textbf{school cancellation} = T, \textbf{snow} = light, \textbf{roads salted})}{p(\textbf{snow} = light)} \\ &= 0.22 \end{aligned}$$

$$p(\textbf{school cancellation} = F|\textbf{snow} = light) = 1 - p(\textbf{school cancellation} = T|\textbf{snow} = light) = 0.78$$

■

Problem 3.

Solution:

Answer for problem (a):

$$\begin{aligned} l(\theta^{(t+1)}) &>= \sum_i \sum_{z^{(i)}} Q_i^{(t)}(z^{(i)}) log \frac{p(x^{(i)}, z^{(i)}; \theta^{(t+1)})}{Q_i^{(t)}(z^{(i)})} \\ &>= \sum_i \sum_{z^{(i)}} Q_i^{(t)}(z^{(i)}) log \frac{p(x^{(i)}, z^{(i)}; \theta^{(t)})}{Q_i^{(t)}(z^{(i)})} \\ &= l(\theta^{(t)}) \end{aligned}$$

Answer for problem (b):

Differentiating the log likelihood directly we get

$$\begin{aligned} \frac{\partial}{\partial \theta_j} \sum_i \log \sum_{z^{(i)}} p(x^{(i)}, z^{(i)}; \theta) &= \sum_i \frac{1}{\sum_{z^{(i)}} p(x^{(i)}, z^{(i)}; \theta)} \sum_{z^{(i)}} \frac{\partial}{\partial \theta_j} p(x^{(i)}, z^{(i)}; \theta) \\ &= \sum_i \sum_{z^{(i)}} \frac{1}{p(x^{(i)}; \theta)} \cdot \frac{\partial}{\partial \theta_j} p(x^{(i)}, z^{(i)}; \theta) \end{aligned}$$

For the GEM algorithm,

$$\frac{\partial}{\partial \theta_j} \sum_i \sum_{z^{(i)}} Q_i(z^{(i)}) \log \frac{p(x^{(i)}, z^{(i)}; \theta)}{Q_i(z^{(i)})} = \sum_i \sum_{z^{(i)}} \frac{Q_i(z^{(i)})}{p(x^{(i)}, z^{(i)}; \theta)} \cdot \frac{\partial}{\partial \theta_j} p(x^{(i)}, z^{(i)}; \theta)$$

But for E-step of the GEM algorithm chooses

$$Q_i(z^{(i)}) = p(x^{(i)}|z^{(i)};\theta) = \frac{p(x^{(i)}, z^{(i)};\theta)}{p(x^{(i)};\theta)}$$

So

$$\sum_i \sum_{z^{(i)}} \frac{Q_i(z^{(i)})}{p(x^{(i)}, z^{(i)};\theta)} \cdot \frac{\partial}{\partial \theta_j} p(x^{(i)}, z^{(i)};\theta) = \sum_i \sum_{z^{(i)}} \frac{1}{p(x^{(i)};\theta)} \cdot \frac{\partial}{\partial \theta_j} p(x^{(i)}, z^{(i)};\theta)$$

which is the same as the derivative of the log likelihood.3 ■