Homework 2

June 13, 2014

1. If we have $f : \mathbb{R}^{n\times p} \to \mathbb{R}$, then $\frac{\mathrm{d}f}{\mathrm{d}\mathbf{X}} = \frac{\partial f}{\partial x_{ij}}$. Assume $f = \mathrm{tr}(\mathbf{AX})$. prove that $\frac{\mathrm{d}f}{\mathrm{d}\mathbf{X}} = \mathbf{A}^T$.

   Solution

$$
\begin{aligned}
&\mathrm{tr}(\mathbf{AX}) \\
&= \mathrm{tr}\begin{bmatrix} \mathbf{a}_{(1)}^T \\ \mathbf{a}_{(2)}^T \\ \vdots \\ \mathbf{a}_{(p)}^T \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_p \end{bmatrix} \\
&= \mathrm{tr}\begin{bmatrix} \mathbf{a}_{(1)}^T\mathbf{x}_1 & \mathbf{a}_{(1)}^T\mathbf{x}_2 & \cdots & \mathbf{a}_{(1)}^T\mathbf{x}_p \\ \mathbf{a}_{(2)}^T\mathbf{x}_1 & \mathbf{a}_{(2)}^T\mathbf{x}_2 & \cdots & \mathbf{a}_{(2)}^T\mathbf{x}_p \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{a}_{(p)}^T\mathbf{x}_1 & \mathbf{a}_{(p)}^T\mathbf{x}_2 & \cdots & \mathbf{a}_{(p)}^T\mathbf{x}_p \end{bmatrix} \\
&= \mathbf{a}_{(1)}^T\mathbf{x}_1 + \mathbf{a}_{(2)}^T\mathbf{x}_2 + \cdots + \mathbf{a}_{(p)}^T\mathbf{x}_p \\
&= \sum_{j=1}^{p}\sum_{i=1}^{n} \mathbf{a}_{ji}\mathbf{x}_{ij}
\end{aligned}
$$

   Thus we can get:

$$
\frac{\mathrm{d}f}{\mathrm{d}\mathbf{X}} = \frac{\mathrm{tr}(\mathbf{AX})}{\mathrm{d}\mathbf{X}} = [\mathbf{a}_{ji}] = \mathbf{A}^T
$$

2. Let $\mathbf{X}$ be $n \times p$ matrices. Solve the following optimization problem: $\min \phi(\mathbf{Z}, \mathbf{V}|\mathbf{X}) = \|\mathbf{X} - \mathbf{ZV}\|_F^2$, s.t. $\mathbf{V}^T\mathbf{V} = \mathbf{I}_q$ and $\mathbf{Z}^T\mathbf{1}_n = 0$, where $\mathbf{V}$ is a $q \times p$ matrix and $\mathbf{Z}$ is a $n \times q$ matrix.

   Solution Construct the Lagrangian function

$$
L(\mathbf{X}, \mathbf{V}, \mathbf{C}, \mathbf{d}) = \mathrm{tr}[(\mathbf{X} - \mathbf{ZV}^T)(\mathbf{X} - \mathbf{ZV}^T)^T] - \mathrm{tr}[\mathbf{C}(\mathbf{V}^T\mathbf{V} - \mathbf{I}_q)] - \mathrm{tr}(\mathbf{Z}^T\mathbf{1}_n\mathbf{d}^T),
$$

   where $\mathbf{C} \in \mathbb{R}^{q\times q}$ and $\mathbf{d} \in \mathbb{R}^q$. Then we have

$$
\frac{\partial L}{\partial \mathbf{Z}} = \mathbf{0}, \quad \frac{\partial L}{\partial \mathbf{V}} = \mathbf{0}, \quad \frac{\partial L}{\partial \mathbf{d}} = \mathbf{0}, \quad \frac{\partial L}{\partial \mathbf{C}} = \mathbf{0},
$$

which implies

$$
\begin{aligned}
\mathbf{Z} - \mathbf{X}\mathbf{V} &= \mathbf{1}_n \mathbf{d}^T, & (1)\\
\mathbf{X}^T\mathbf{Z} - \mathbf{V}\mathbf{Z}^T\mathbf{Z} + \mathbf{V}\mathbf{C} &= \mathbf{0}, & (2)\\
\mathbf{V}^T\mathbf{V} &= \mathbf{I}_q, & (3)\\
\mathbf{Z}^T\mathbf{1}_n &= \mathbf{0}. & (4)
\end{aligned}
$$

Multiplying (1) by $\mathbf{H}$ on both sides, we have

$$
\begin{aligned}
& \mathbf{HZ} = \mathbf{HXV} \\
\Longrightarrow\ & (\mathbf{1}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T)\mathbf{Z} = \mathbf{HXV} \quad // \text{ using (4)} \\
\Longrightarrow\ & \mathbf{Z} = \mathbf{HXV}. \qquad (5)
\end{aligned}
$$

Substitute (5) into (2), we have

$$
\begin{aligned}
& \mathbf{X}^T\mathbf{HXV} - \mathbf{V}\mathbf{V}^T\mathbf{X}^T\mathbf{HHXV} + \mathbf{VC} = \mathbf{0} \\
\Longrightarrow\ & \mathbf{V}^T\mathbf{X}^T\mathbf{HXV} - \mathbf{V}^T\mathbf{X}^T\mathbf{HHXV} + \mathbf{C} = \mathbf{0} \quad // \text{ multiplying } \mathbf{V}^T \\
\Longrightarrow\ & \mathbf{C} = \mathbf{0}
\end{aligned}
$$

Therefore $\mathbf{X}^T\mathbf{HXV} = \mathbf{V}\mathbf{V}^T\mathbf{X}^T\mathbf{HXV}$, and $\mathbf{V}$ consists of the eigenvectors associate with the q largest eigenvalues of $\mathbf{X}^T\mathbf{HX}$ and $\mathbf{Z} = \mathbf{HXV}$.

3. Assume a mapping from the latent space into the data space is $\mathbf{x} = \mathbf{W}\mathbf{z} + \boldsymbol{\mu} + \sigma\boldsymbol{\epsilon}$, where $\mathbf{x} \in \mathbb{R}^p$, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_p)$, $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_q)$,$\mathbf{z} \perp\!\!\!\perp \boldsymbol{\epsilon}$, and $\mathbf{W}^T\mathbf{W} = \mathbf{I}_q$. Please use EM algorithm to solve parameters: $\mathbf{W}, \boldsymbol{\mu}, \sigma$.
Solution
E-step:

$$
\begin{aligned}
\mathbf{Q}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t)}) = &-\frac{np}{2}\log\tau - \frac{n}{2\tau}\mathrm{tr}(\mathbf{S}) \\
&- \sum_{i=1}^{n}\left\{\frac{1}{2\tau}\mathrm{tr}(\mathbf{W}^T\mathbf{W}\langle\mathbf{z}_i, \mathbf{z}_i^T\rangle) - \frac{1}{\tau}(\mathbf{x}_i - \boldsymbol{\mu})^T\mathbf{W}\langle\mathbf{z}_i\rangle + \frac{1}{2}\mathrm{tr}(\langle\mathbf{z}_i, \mathbf{z}_i^T\rangle)\right\}
\end{aligned}
$$

where $\tau = \sigma^2$, $\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i - \hat{\boldsymbol{\mu}})(\mathbf{x}_i - \hat{\boldsymbol{\mu}})^T$, $\langle\mathbf{z}_i\rangle = \mathbf{M}_{(t)}^{-1}\mathbf{W}_{(t)}^T(\mathbf{x}_i - \boldsymbol{\mu})$, $\langle\mathbf{z}_i, \mathbf{z}_i^T\rangle = \tau_{(t)}\mathbf{M}_{(t)}^{-1} + \langle\mathbf{z}_i\rangle\langle\mathbf{z}_i\rangle^T$, $\mathbf{M} = \sigma^2\mathbf{I}_q + \mathbf{W}^T\mathbf{W}$
M-step:
We need to maximize $\mathbf{Q}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t)})$, subject to $\mathbf{W}^T\mathbf{W} = \mathbf{I}_q$.

$$
\begin{aligned}
L = & -\frac{np}{2}\log\tau - \frac{n}{2\tau}\mathrm{tr}(\mathbf{S}) \\
& -\sum_{i=1}^{n}\left\{\frac{1}{2\tau}\mathrm{tr}(\mathbf{W}^T\mathbf{W}\langle\mathbf{z}_i,\mathbf{z}_i^T\rangle)\right. \\
& \left.-\frac{1}{\tau}(\mathbf{x}_i-\boldsymbol{\mu})^T\mathbf{W}\langle\mathbf{z}_i\rangle + \frac{1}{2}\mathrm{tr}(\langle\mathbf{z}_i,\mathbf{z}_i^T\rangle)\right\} \\
& -\mathrm{tr}(\mathbf{D}(\mathbf{W}^T\mathbf{W}-\mathbf{I}_q))
\end{aligned} \tag{1}
$$

$$
\frac{\mathrm{d}L}{\mathrm{d}\mathbf{W}} = -\frac{1}{\tau}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})\langle\mathbf{z}_i^T\rangle + \frac{1}{\tau}\sum_{i=1}^{n}\mathbf{W}\langle\mathbf{z}_i,\mathbf{z}_i^T\rangle - \mathbf{W}\mathbf{D} = 0
$$

Multiply both sides by $\mathbf{W}^T$, we have

$$
\begin{aligned}
\sum_{i=1}^{n}\langle\mathbf{z}_i,\mathbf{z}_i^T\rangle) &= \sum_{i=1}^{n}\mathbf{W}^T(\mathbf{x}_i-\boldsymbol{\mu})\langle\mathbf{z}_i^T\rangle - \tau\mathbf{D} \\
\mathbf{D} &= \frac{1}{\tau}\sum_{i=1}^{n}[\mathbf{W}^T(\mathbf{x}_i-\boldsymbol{\mu})\langle\mathbf{z}_i^T\rangle - \langle\mathbf{z}_i,\mathbf{z}_i^T\rangle]
\end{aligned} \tag{2}
$$

Substituting (2) in to (1), we have

$$
\sum_{i=1}^{n}\mathbf{W}^T\langle\mathbf{z}_i,\mathbf{z}_i^T\rangle = \sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})\langle\mathbf{z}_i^T\rangle - \sum_{i=1}^{n}\mathbf{W}\mathbf{W}^T(\mathbf{x}_i-\boldsymbol{\mu})\langle\mathbf{z}_i^T\rangle + \sum_{i=1}^{n}\mathbf{W}^T\langle\mathbf{z}_i,\mathbf{z}_i^T\rangle
$$

$$
\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})\langle\mathbf{z}_i^T\rangle = \sum_{i=1}^{n}\mathbf{W}\mathbf{W}^T(\mathbf{x}_i-\boldsymbol{\mu})\langle\mathbf{z}_i^T\rangle
$$

$$
\mathbf{W}\mathbf{W}^T\mathbf{S}\mathbf{W}^{(t)} = \mathbf{S}\mathbf{W}^{(t)}
$$

We perform SVD on $\mathbf{S}\mathbf{W}^{(t)}$, which is $\mathbf{S}\mathbf{W}^{(t)} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T$, then we have $\mathbf{W}\mathbf{W}^T\mathbf{U}\boldsymbol{\Sigma} = \mathbf{U}\boldsymbol{\Sigma}$. Therefore, we can know that $\mathbf{W} = \mathbf{U}$, which indicates that $\mathbf{W}^{(t+1)}$ is the left singular matrix of $\mathbf{S}\mathbf{W}^{(t)}$.

$$
\begin{aligned}
\frac{\mathrm{d}L}{\mathrm{d}\tau} &= 0 \\
\tau^{(t+1)} &= \frac{1}{p}[\mathrm{tr}(\mathbf{S}) - \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}_i-\boldsymbol{\mu})^T\mathbf{W}^{(t+1)}\langle\mathbf{z}_i\rangle]
\end{aligned}
$$

4. Please give the formula of Probabilistic Kernel PCA, and solve it.
   Solution Please refer to this paper: "Probabilistic kernel principal component analysis".