Abstract

本期我们学习概率视角下的主成分分析。

Given

$$
\begin{aligned}
&z \in R^p \quad x \in R^q \quad q < p \\
&\begin{cases} z \sim N(0, I) \\ x = wz + \mu + \epsilon \\ \epsilon \sim N(0, \sigma^2 I) \end{cases}
\end{aligned} \tag{1}
$$

我们很容易可以得出其中的隐含条件: $z \perp \epsilon$

solve $x|z$

因为是对z进行降维，因此我们在推理的时候需要计算$p(x|z)$

solve x

$$
\begin{aligned}
E[x] &= E[wz + \mu + \epsilon] = \mu \\
Var[x] &= Var[wz + \mu + \epsilon] \\
&= Var[wz] + Var[\mu] + Var[\epsilon] \\
&= wVar[z]w^T + 0 + \sigma^2 I \\
&= ww^T + \sigma^2 I \\
&\therefore x \sim N(\mu, ww^T + \sigma^2 I)
\end{aligned} \tag{2}
$$

construct t

下面我们构造一个联合分布:

$$
t = \begin{pmatrix} x \\ z \end{pmatrix} \tag{3}
$$

$$
\begin{aligned}
E[t] &= \begin{pmatrix} E[x] \\ E[z] \end{pmatrix} = \begin{pmatrix} \mu \\ 0 \end{pmatrix} \\
Var[t] &= \begin{pmatrix} ww^T + \sigma^2 I & \Delta \\ \Delta^T & I \end{pmatrix} \\
\Delta &= Cov(x, z) \\
&= E[(x - E[x])(z - E[z])^T] \\
&= E[(x - \mu)z^T] \\
&= E[(wz + \epsilon)z^T] \\
&= wE[zz^T] + E[\epsilon z^T] \\
&\because z \perp \epsilon \\
&= wVar[z] + 0 \\
&= w
\end{aligned} \tag{4}
$$

$$\therefore t \sim N(\begin{bmatrix}\mu\\0\end{bmatrix}, \begin{bmatrix}ww^T+\sigma^2 I & w\\ w^T & I\end{bmatrix}) \tag{5}$$

construct $x.z$

下面为了求出 $x|z$，我们构造一个新的概率分布:

$$\begin{aligned} x.z &= x - \Sigma_{xz}\Sigma_{zz}^{-1}z \\ &= x - wI^{-1}z \\ &= \begin{pmatrix}I & -w\end{pmatrix}\begin{pmatrix}x\\z\end{pmatrix} \\ &= \begin{pmatrix}I & -w\end{pmatrix}\cdot t \end{aligned} \tag{6}$$

$$\begin{aligned} E[x.z] &= \begin{pmatrix}I & -w\end{pmatrix}E[t] \\ &= \begin{pmatrix}I & -w\end{pmatrix}\begin{pmatrix}\mu\\0\end{pmatrix} \\ &= \mu \end{aligned} \tag{7}$$

$$\begin{aligned} Var[x.z] &= \begin{pmatrix}I & -w\end{pmatrix}Var[t]\begin{pmatrix}I\\-w^T\end{pmatrix} \\ &= \begin{pmatrix}I & -w\end{pmatrix}\begin{pmatrix}ww^T+\sigma^2 I & w\\ w^T & I\end{pmatrix}\begin{pmatrix}I\\-w^T\end{pmatrix} \\ &= \begin{pmatrix}\sigma^2 I & 0\end{pmatrix}\begin{pmatrix}I\\-w^T\end{pmatrix} \\ &= \sigma^2 I \end{aligned} \tag{8}$$

$$\therefore x.z \sim N(\mu, \sigma^2 I) \tag{9}$$

derive $x|z$

将(6)变形，我们得到:

$$x|z = x.z + wz \tag{10}$$

$$\begin{aligned} \therefore E[x|z] &= E[x.z] + wz \\ &= wz + \mu \\ Var[x.z] &= Var[x.z] \\ &= \sigma^2 I \end{aligned} \tag{11}$$

$$\therefore x|z \sim N(wz+\mu, \sigma^2 I) \tag{12}$$

conclusion

因此我们在训练时只需用EM求出这三个参数即可:

$$parameters = \{w, \mu, \sigma\} \tag{13}$$

而在推理的时候求出$p(x|z)$即可