

0.1 P-PCA

在本节，我们将学习概率视角的 PCA，这被称为 P-PCA。

0.1.1 已知

$$\begin{aligned} x &\in R^p, z \in R^q, q < p \\ \begin{cases} z \sim N(0_q, I_q) \\ x = wz + \mu + \epsilon \\ \epsilon \sim N(0_p, \sigma^2 I_p) \end{cases} \end{aligned}$$

0.1.2 所求

$$x|z, x, z|x$$

0.1.3 推导

推导 x

$$\begin{aligned} E[x] &= wE[z] + E[\mu] + E[\epsilon] \\ &= \mu \end{aligned} \tag{1}$$

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

推导 $x|z$

构造分布 t

$$\begin{aligned} t &= \begin{pmatrix} x \\ z \end{pmatrix} \\ E[t] &= \begin{pmatrix} E[x] \\ E[z] \end{pmatrix} \\ &= \begin{pmatrix} \mu \\ 0_q \end{pmatrix} \end{aligned} \tag{3}$$

$$\begin{aligned}
cov(x, z) &= E[(x - E[x])(z - E[z])^T] \\
&= E[(x - \mu)z^T] \\
&= E[(wz + \epsilon)z^T] \\
&= E[wzz^T] + E[\epsilon z^T] \\
&= wE[(z - E[z])(z - E[z])^T] + E[\epsilon z^T] \\
&= wVar[z] + 0 \\
&= wI_q \\
&= w
\end{aligned} \tag{4}$$

$$\therefore Var[t] = \begin{pmatrix} ww^T + \sigma^2 I_p & w \\ w^T & I_q \end{pmatrix} \tag{5}$$

$$\therefore t \sim N\left(\begin{bmatrix} \mu l_p \\ 0_q \end{bmatrix}, \begin{bmatrix} ww^T + \sigma^2 I_p & w \\ w^T & I_q \end{bmatrix}\right)$$

构造分布 $x.z$
我们设

$$x.z = x - \Sigma_{xz} \Sigma_{zz}^{-1} z$$

$$\therefore x.z = x - wz$$

$$= \begin{pmatrix} I_p & -w \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} \tag{6}$$

$$= \begin{pmatrix} I_p & -w \end{pmatrix} t$$

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

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

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

推导 $x|z$

$$\begin{aligned}\therefore x.z &= x - \Sigma_{xz}\Sigma_{zz}^{-1}z \\ \therefore x|z &= x.z + \Sigma_{xz}\Sigma_{zz}^{-1}z \\ &= x.z + wz\end{aligned}\tag{10}$$

这里 z 是已知的, 因此我们可以将 wz 视为常数。

$$\begin{aligned}E[x|z] &= E[x.z] + wz \\ &= \mu l_p + wz\end{aligned}\tag{11}$$

$$\begin{aligned}Var[x|z] &= Var[x.z] + 0 \\ &= \sigma^2 I_p\end{aligned}\tag{12}$$

$$\therefore x|z \sim N(\mu l_p + wz, \sigma^2 I_p)\tag{13}$$

推导 $z|x$

我们设

$$\begin{aligned}z.x &= z - \Sigma_{zx}\Sigma_{xx}^{-1}x \\ &= z - w^T \Sigma_{xx}^{-1}x \\ \Sigma &= \Sigma_{xx} \\ \therefore z.x &= z - w^T \Sigma^{-1}x\end{aligned}\tag{14}$$

让我们对上式稍作变形:

$$\begin{aligned}z.x &= z - w^T \Sigma^{-1}x \\ &= \begin{pmatrix} -w^T \Sigma^{-1} & I \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} \\ &= \begin{pmatrix} -w^T \Sigma^{-1} & I \end{pmatrix} t\end{aligned}\tag{15}$$

$$\begin{aligned}E[z.x] &= \begin{pmatrix} -w^T \Sigma^{-1} & I \end{pmatrix} \begin{pmatrix} \mu l_p \\ 0_q \end{pmatrix} \\ &= -\mu w^T \Sigma^{-1}\end{aligned}\tag{16}$$

$$\begin{aligned}Var[z.x] &= \begin{pmatrix} -w^T \Sigma^{-1} & I \end{pmatrix} \begin{pmatrix} \Sigma & w \\ w^T & I \end{pmatrix} \begin{pmatrix} -(\Sigma^{-1})^T w \\ I \end{pmatrix} \\ &= \begin{pmatrix} 0 & -w^T \Sigma^{-1} w + I \end{pmatrix} \begin{pmatrix} -\Sigma^{-1} w \\ I \end{pmatrix} \\ &= -w^T \Sigma^{-1} w + I\end{aligned}\tag{17}$$

$$\therefore z.x \sim N(-\mu w^T \Sigma^{-1}, -w^T \Sigma^{-1} w + I)\tag{18}$$

$$\begin{aligned}\therefore z.x &= z - w^T \Sigma^{-1} x \\ \therefore z|x &= z.x + w^T \Sigma^{-1} x\end{aligned}\tag{19}$$

$$\begin{aligned}E[z|x] &= E[z.x] + w^T \Sigma^{-1} x \\ &= -\mu w^T \Sigma^{-1} + w^T \Sigma^{-1} x \\ &= w^T \Sigma^{-1} (x - \mu l_p) \\ &= w^T (w w^T + \sigma^2 I)^{-1} (x - \mu l_p)\end{aligned}\tag{20}$$

$$\begin{aligned}Var[z|x] &= Var[z.x] + 0 \\ &= -w^T \Sigma^{-1} w + I \\ &= -w^T (w w^T + \sigma^2 I)^{-1} w + I\end{aligned}\tag{21}$$

$$\therefore z|x \sim N(w^T (w w^T + \sigma^2 I)^{-1} (x - \mu l_p), -w^T (w w^T + \sigma^2 I)^{-1} w + I)\tag{22}$$

0.1.4 总结

我们需要搜索的参数是：

$$\{w, \sigma, \mu\}$$

我们可以用 EM 算法求得他们，通过最大化 $p(x|z)$ ，这意味着：

- 训练阶段我们计算 $p(x|z)$
- 推理阶段我们计算 $p(z|x)$