0.1 P-PCA

In this section, we'll learn PCA from perspective of probability which is called P-PCA.

0.1.1 Given

$$x \in R^p,\ z \in R^q,\ q < p$$

$$\begin{cases} z \backsim N(0_q, I_q) \\ x = wz + \mu + \epsilon \\ \epsilon \backsim N(0_p, \sigma^2 I_p) \end{cases}$$

0.1.2 To solve

$$x|z \quad x \quad z|x$$

0.1.3 Derivation

derive x

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

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

$$\therefore x \backsim N(\mu, ww^T + \sigma^2 I_p)$$

derive $x|z$

construct dist t

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

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

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

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

construct $x.z$

$$
\begin{aligned}
set\ x.z &= x - \Sigma_{xz}\Sigma_{zz}^{-1} z \\
\therefore x.z &= x - wz \\
&= \begin{pmatrix} I_p & -w \end{pmatrix} \begin{pmatrix} x \\ z \end{pmatrix} \\
&= \begin{pmatrix} I_p & -w \end{pmatrix} t
\end{aligned} \tag{6}
$$

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

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

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

derive $x|z$

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

Here z is known, so we can treat wz as constant.

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

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

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

derive $z|x$

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

let's slightly deform the formula above:

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

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

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

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

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

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

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

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

0.1.4 Conclusion

the parameters we are to seek are:

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

We can obtain them by maximizing $p(x|z)$ via EM algorithm which means:

- we calculate $p(x|z)$ in the process of training.
- we calculate $p(z|x)$ in the process of inference.