Abstract

In this issue, we study principal component analysis from the perspective of probability.

Given

$$
\begin{gathered}
z \in R^p \quad x \in R^q \quad q < p \\
\begin{cases}
z \sim N(0, I) \\
x = wz + \mu + \epsilon \\
\epsilon \sim N(0, \sigma^2 I)
\end{cases}
\end{gathered} \tag{14}
$$

It is easy for us to figure out the implicit conditions: $z \perp \epsilon$

solve $x|z$

Since we are reducing the dimension of z, we need to calculate $p(x|z)$ in our inference stage.

solve x

$$
\begin{aligned}
E[x] &= E[wz + \mu + \epsilon] = \mu \\
Var[x] &= Var[wz + \mu + \epsilon] \\
&= Var[wz] + Var[\mu] + Var[\epsilon] \\
&= wVar[z]w^T + 0 + \sigma^2 I \\
&= ww^T + \sigma^2 I \\
\therefore x &\sim N(\mu, ww^T + \sigma^2 I)
\end{aligned} \tag{15}
$$

construct t

Let's construct a joint distribution:

$$
t = \begin{pmatrix} x \\ z \end{pmatrix} \tag{16}
$$

$$
\begin{aligned}
E[t] &= \begin{pmatrix} E[x] \\ E[z] \end{pmatrix} = \begin{pmatrix} \mu \\ 0 \end{pmatrix} \\
Var[t] &= \begin{pmatrix} ww^T + \sigma^2 I & \Delta \\ \Delta^T & I \end{pmatrix} \\
\Delta &= Cov(x, z) \\
&= E[(x - E[x])(z - E[z])^T] \\
&= E[(x - \mu)z^T] \\
&= E[(wz + \epsilon)z^T] \\
&= wE[zz^T] + E[\epsilon z^T] \\
&\because z \perp \epsilon \\
&= wVar[z] + 0 \\
&= w
\end{aligned} \tag{17}
$$

$$\therefore t \sim N(\begin{bmatrix} \mu \\ 0 \end{bmatrix}, \begin{bmatrix} ww^T + \sigma^2 I & w \\ w^T & I \end{bmatrix}) \tag{18}$$

construct $x.z$

To solve $x|z$, we construct a specific distribution:

$$\begin{aligned} x.z &= x - \Sigma_{xz}\Sigma_{zz}^{-1} z \\ &= x - wI^{-1}z \\ &= (I \quad -w) \begin{pmatrix} x \\ z \end{pmatrix} \\ &= (I \quad -w) \cdot t \end{aligned} \tag{19}$$

$$\begin{aligned} E[x.z] &= (I \quad -w) E[t] \\ &= (I \quad -w) \begin{pmatrix} \mu \\ 0 \end{pmatrix} \\ &= \mu \end{aligned} \tag{20}$$

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

$$\therefore x.z \sim N(\mu, \sigma^2 I) \tag{22}$$

derive $x|z$

deform formula (6), we got:

$$x|z = x.z + wz \tag{23}$$

$$\begin{aligned} \therefore E[x|z] &= E[x.z] + wz \\ &= wz + \mu \\ Var[x.z] &= Var[x.z] \\ &= \sigma^2 I \end{aligned} \tag{24}$$

$$\therefore x|z \sim N(wz + \mu, \sigma^2 I) \tag{25}$$

conclusion

Therefore, we only need EM to calculate these three parameters during training process:

$$parameters = \{w, \mu, \sigma\} \tag{26}$$

And calculate the $p(x|z)$ during inference process.