Problem 1: PPCA

Wednesday, 23 January 2019 15:27

Problem 1: Consider the latent space distribution

$$p(\boldsymbol{z}) = \mathcal{N}(\boldsymbol{z}|\boldsymbol{0}, \boldsymbol{I})$$

and a conditional distribution for the observed variable $\boldsymbol{x} \in \mathbb{R}^d$,

$$p(\boldsymbol{x}|\boldsymbol{z}) = \mathcal{N}(\boldsymbol{x}|\cancel{\boldsymbol{W}\boldsymbol{z} + \boldsymbol{\mu}}\ A\overset{\mu_z}{\hat{z}} + b, \boldsymbol{\Phi})$$

where $\boldsymbol{\Phi}$ is an arbitrary symmetric, positive-definite noise covariance variable. Now suppose that we make a nonsingular linear transformation of the data variables $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$ where $\boldsymbol{A}$ is a non-singular $d \times d$ matrix. If $\boldsymbol{\mu}_{ML}$, $\boldsymbol{W}_{ML}$, and $\boldsymbol{\Phi}_{ML}$ represent the maximum likelihood solution corresponding to the original untransformed data, show that $\boldsymbol{A}\boldsymbol{\mu}_{ML}$, $\boldsymbol{A}\boldsymbol{W}_{ML}$, and $\boldsymbol{A}\boldsymbol{\Phi}_{ML}\boldsymbol{A}^T$ will represent the corresponding maximum likelihood solution for the transformed data set. Finally, show that the form of the model is preserved if $\boldsymbol{A}$ is orthogonal and $\boldsymbol{\Phi}$ is proportional to the unit matrix so $\boldsymbol{\Phi} = \sigma^2\boldsymbol{I}$ (i.e. probabilistic PCA). The transformed $\boldsymbol{\Phi}$ matrix remains proportional to the unit matrix, and hence probabilistic PCA is covariant under a rotation of the axes of data space, as is the case for conventional PCA.

Gaussian Identities

Bishop 2.3

1)
$$p(x) = \mathcal{N}(\mu, \Sigma)$$
$$P(Ax) = \mathcal{N}(A\mu, A\Sigma A^T)$$

2)
$$p(z) = \mathcal{N}(\mu_z, \Sigma_z)$$
$$p(x|z) = \mathcal{N}(Az + b, \Sigma_{x|z})$$
$$p(x) = \mathcal{N}(A\mu_z + b, \Sigma_{x|z} + A\Sigma_z A^T)$$

3)
$$\mathcal{N}(x|\mu_1, \Sigma_1) \cdot \mathcal{N}(x|\mu_2, \Sigma_2)$$
$$= \mathcal{N}(x|\mu, \Sigma) \qquad \mu = \Sigma(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2)$$
$$\Sigma = (\Sigma_1^{-1} + \Sigma_2^{-1})$$

$$\mathcal{N}(x|\mu, \Sigma) \propto e^{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)}$$

$$\mathcal{N}(x|\mu,\Sigma) \propto e^{-\frac{1}{2}\underbrace{(x-\mu)^T\Sigma^{-1}(x-\mu)}}$$

$$-\frac{1}{2}\left[x^T\Sigma^{-1}x - \mu^T\Sigma^{-1}x - x^T\Sigma^{-1}\mu + \mu^T\Sigma^{-1}\mu\right]$$

$$A = \Sigma^{-1} \qquad r = \Sigma^{-1}\mu$$

$$\mathcal{N}(x|\mu,\Sigma) \propto e^{-\frac{1}{2}x^TAx + r^Tx}$$

$$e^{-\frac{1}{2}x^TA_1x + r_1^Tx} \cdot e^{-\frac{1}{2}x^TA_2x + r_2^Tx} = e^{\left[-\frac{1}{2}x^T(A_1+A_2)x + (r_1+r_2)^Tx\right]}$$

$$\tilde{A} = A_1 + A_2 = \Sigma_1^{-1} + \Sigma_2^{-1}$$

$$\tilde{r} = r_1 + r_2 = \Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2$$

$$\Sigma = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}$$

$$\mu = \Sigma\tilde{r} = \Sigma(r_1 + r_2) = \Sigma(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2)$$

$$p(z) = \mathcal{N}(z|0,I)$$

$$p(x|z) = \mathcal{N}(x|Wz + \mu, \Phi)$$

$$p(y) \qquad y = Ax$$

$$p(x) = \mathcal{N}(x \,|\, \underset{A\mu_z + b}{W\mu_z + \mu}, \Phi + WIW^T)$$

$$= \mathcal{N}(x|\mu, \Phi + WW^T)$$

$$p(y) = p(Ax) = \mathcal{N}(x|A\mu, A(\Phi + WW^T)A^T)$$

$$= \mathcal{N}(y|A\mu, \underbrace{A\Phi A^T + AWW^TA^T})$$

$$p(x) = \mathcal{N}(x|\mu, \Phi + WW^T)$$

$$p(x) = \mathcal{N}(x \mid \mu, \Phi + WW^T)$$

$\mu_{ML} = \mu$ $\qquad \Phi_{ML} = \Phi$ $\qquad W_{ML} = W$

$\mu_{ML(y)} = A\mu$ $\qquad \Phi_{ML(y)} = A\Phi A^T$ $\qquad W_{ML(y)} = AW$

1. Rule for $p(z)$ $p(x|z)$ $\rightarrow$ $p(x)$
2. Rule $p(Ax)$ $p(x)$
3. pattern matching

$\Phi = \sigma^2 I$ $\qquad AA^T = A^TA = I$

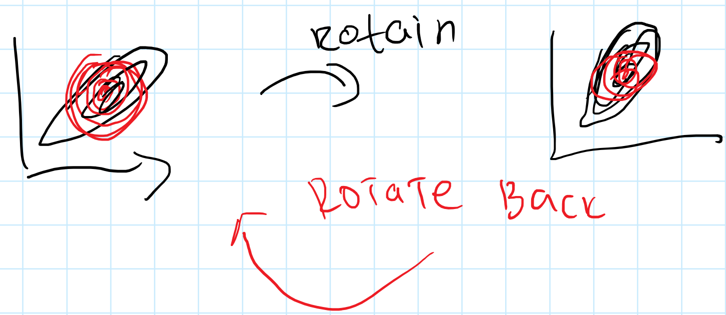
$$A\Phi A^T + AWW^TA^T =$$

$$A\sigma^2 I A^T + AWW^TA^T =$$

$$\sigma^2 I \underbrace{AA^T}_{I} + AWW^TA^T =$$

$$\Sigma_y = \sigma^2 I + AWW^TA^T$$

$$\Sigma_x = \sigma^2 I + WW^T$$



Problem 2: SVD

Wednesday, 23 January 2019 15:28

Problem 2: Use the SVD shown below. Suppose a new user Leslie assigns rating 3 to Alien and rating 4 to Titanic, giving us a representation of Leslie in the 'original space' of $[0, 3, 0, 0, 4]$. Find the representation of Leslie in concept space. What does that representation predict about how well Leslie would like the other movies appearing in our example data?

	Matrix	Alien	Star Wars	Casablanca	Titanic
Joe	1	1	1	0	0
Jim	3	3	3	0	0
John	4	4	4	0	0
Jack	5	5	5	0	0
Jill	0	0	0	4	4
Jenny	0	0	0	5	5
Jane	0	0	0	2	2

Figure 11.6: Ratings of movies by users

$$\underset{M}{\begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 3 & 3 & 3 & 0 & 0 \\ 4 & 4 & 4 & 0 & 0 \\ 5 & 5 & 5 & 0 & 0 \\ 0 & 0 & 0 & 4 & 4 \\ 0 & 0 & 0 & 5 & 5 \\ 0 & 0 & 0 & 2 & 2 \end{bmatrix}} = \underset{U}{\begin{bmatrix} .14 & 0 \\ .42 & 0 \\ .56 & 0 \\ .70 & 0 \\ 0 & .60 \\ 0 & .75 \\ 0 & .30 \end{bmatrix}} \underset{\Sigma}{\begin{bmatrix} 12.4 & 0 \\ 0 & 9.5 \end{bmatrix}} \underset{V^T}{\begin{bmatrix} .58 & .58 & .58 & 0 & 0 \\ 0 & 0 & 0 & .71 & .71 \end{bmatrix}}$$

clusteing of movies

clustering of the users

$M = U \Sigma V^T$

$U^{(k)} \Sigma^{(k)} V^{T(k)}$

$P = U \Sigma$

$P = M \cdot V = U \Sigma \underbrace{V^T V}_{I} = U \Sigma$

Cas Titanic

$[0, 3, 0, 0, 4] \cdot V =$

$= [1.74, 2.84]$

Problem 5: Autoencoders

Wednesday, 23 January 2019 15:28

Problem 5: We train a linear autoencoder to D-dimensional data. The autoencoder has a single K-dimensional hidden layer, there are no biases, and all activation functions are identity ($\sigma(x) = x$).

- Why is it usually impossible to get zero reconstruction error in this setting if $K < D$?
- Under which conditions is this possible?

$X \in \mathbb{R}^D$

$\sigma^{(i)}(X^{(i-1)} \cdot W^{(i)} + b^{(i)})$

D-dimensional

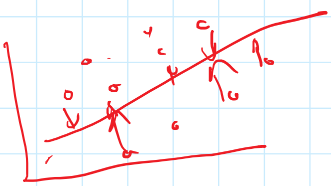
K-dimensional

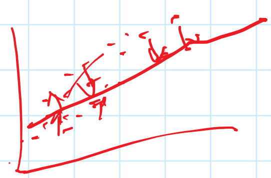
D-dimensional

$(X \cdot W_1)\, W_2$

$W_1 \in \mathbb{R}^{d \times K}$

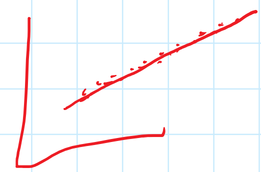
$W_2 \in \mathbb{R}^{K \times D}$





$D = 2$

$K = 1$



$\|dec(enc(x)) - x\|_2^2$

$W_1 = W_2^T$

$\downarrow$

tieD weights

$\sigma(x) = x$

$(X W_1 W_2^T - X)$

$W_1 = W_2^T$ $\quad W_1 W_1^T = I$ $\quad (X I - X) = (X - X)$