

Lecture 18 - scribbles

Friday, November 4, 2016
13:40

today: factor analysis / PCA / CCA

Factor analysis (dimensionality reduction)

$$x_i \in \mathbb{R}^d \rightarrow z_i \in \mathbb{R}^k \quad k \ll d$$

(dimensionality reduction)

"Synthesis view" for PCA (principal component analysis)

get a orthonormal set of k vectors in \mathbb{R}^d

$$W = \begin{bmatrix} w_1 & \dots & w_k \\ | & & | \\ \mathbb{R}^{d \times k} & & \mathbb{R}^k \end{bmatrix} \quad \hookrightarrow \langle w_i, w_j \rangle = \delta_{ij}$$

$$W^T W = I_k$$

$k \times k$ identity

Orthogonal projection on $\text{span}\{w_1, \dots, w_k\} = \text{col}(W)$

$$P_W = W W^T \quad (P_W^2 = W \underbrace{W^T W}_{I_k} W^T = P_W)$$

$$P_W x = \sum_k w_k \underbrace{\langle w_k, x \rangle}_{(z)_k}$$

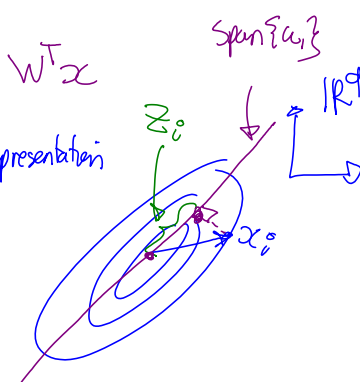
$$z = W^T x$$

new \mathbb{R}^k representation

PCA $\min_{W \in \mathbb{R}^{d \times k}} \sum_i \|x_i - W W^T x_i\|^2$

$W^T W = I_k$

$\text{col}(W) \triangleq$ "principal subspace"



$$X = \begin{pmatrix} -x_1^T \\ \vdots \end{pmatrix} \quad \|X^T - \overbrace{W W^T} X^T\|_F^2$$

$$= \|(I - P_W) X^T\|_F^2$$

$$= \text{tr}(X (I - P_W)^T (I - P_W) X^T)$$

$$= \text{tr}(X (I - P_W) X^T) \quad A_{..} = W W^T$$

$$= \text{tr}(X^T X (I - A_k))$$

scaling of empirical covariance matrix if data is centered

min reconstruction error

$$\Leftrightarrow \text{maximize } \text{tr}(X^T X W W^T)$$

"analyst view" of PCA

$$= \sum_k w_k^T X^T X w_k \rightarrow \text{maximize variance in new representation}$$

solution are top k -e-vectors of $X^T X$ (proportional to emp. cov. matrix)

SVD (singular value decomposition)

matrix $A \in \mathbb{R}^{m \times n}$ can always write as:

$$A = U \Sigma V^T$$

ordered singular values $\sigma_i > 0$
 $\sigma_1 > \sigma_2 > \dots > 0$

where: $U U^T = I_m = U^T U$ $\Sigma = \begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_{\min\{m,n\}} & \\ & & & 0 \end{pmatrix}$

$V V^T = I_n = V^T V$

$$r = \text{rank}(A) = \# \{ \sigma_i > 0 \} = \dim(\text{col}(A)) = \dim(\text{row}(A))$$

u_1, \dots, u_r ← orthonormal basis of $\text{col}(A)$
 v_1, \dots, v_r ← " " " $\text{row}(A)$

(identify U orthogonal $\left\langle \begin{array}{l} |U| = 1 \text{ (} U \text{ is just rotation matrix)} \\ |U| = -1 \text{ (" " " " + odd \# of flips)} \end{array} \right.$)

columns of U are e-vectors of AA^T top r -e-values of $AA^T \rightarrow \sigma_1^2, \dots, \sigma_r^2$
 V " " " $A^T A$ " " " $A^T A$

e-vectors of $X^T X$ if write $X = U \Sigma V^T$

for PCA, want first k columns of V

then reconstruction error $\sum_{i=k+1}^r \sigma_i^2$

example: face classification; get are "eigen-faces"

factor analysis \rightarrow generative model (latent variable)

$z \sim N(0, I_k)$ noise
 $x = Wz + \mu + \epsilon$



~

$$\varepsilon \perp z \quad \varepsilon \sim N(0, D)$$

d x d diagonal matrix

$$x|z \sim N(Wz + \mu, D)$$



what is $p(x)$ $E[x] = E[E[x|z]] = 0 + \mu = \mu$

$$\begin{aligned} \text{cov}(x, x) &= \text{cov}(Wz + \mu + \varepsilon, \sim) \\ &= \text{cov}(Wz, Wz) + \underbrace{\text{cov}(\varepsilon, \varepsilon)}_{\text{indep.}} \\ &= W \text{cov}(z, z) W^T + D \end{aligned}$$

$$\text{cov}(x, x) = WW^T + \underbrace{I_K}_D$$

equivalent model for x : $x \sim N(\mu, \underbrace{WW^T}_{d \times d} + \underbrace{D}_{\text{diagonal}}) \Rightarrow d \text{ degrees of freedom}$

\rightarrow low rank covariance model for x

to get $p(z|x)$ $\begin{pmatrix} x \\ z \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_x \\ \mu_z \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xz} \\ \Sigma_{zx} & \Sigma_{zz} \end{pmatrix}\right)$

$$\mu_x = \mu \quad \mu_z = 0 \quad \begin{aligned} \Sigma_{xx} &= WW^T + D \\ \Sigma_{zz} &= I_K \end{aligned}$$

$$\begin{aligned} \text{cov}(x, z) &= \text{cov}(Wz + \underbrace{\mu + \varepsilon}_0, z) = \text{cov}(Wz, z) \\ &= W \underbrace{\text{cov}(z, z)}_{I_K} = W \end{aligned}$$

$$\begin{aligned} E[z|x] &= \mu_z + \Sigma_{zx} \Sigma_{xx}^{-1} (x - \mu_x) \\ &= 0 + W^T (WW^T + D)^{-1} (x - \mu_x) \end{aligned}$$

[from conditioning block in Gaussian formula]

probabilistic PCA: suppose $D = \sigma^2 I_d$

$$\lim_{\sigma \rightarrow 0} W^T (WW^T + \sigma^2 I)^{-1} = W^T \text{ pseudo-inverse} = W^T \text{ (here if } W^T W = I_K)$$

(prove this using matrix inversion lemma)

$$E[z|x] = W^T (x - \mu_x) \quad [\text{PCA representation}]$$

PCA as limit $\sigma \rightarrow 0$ of PPCA

* factor analysis, estimate $\hat{W}, \hat{D}, \hat{\mu}$ using ML

* factor analysis, estimate $\hat{W}, \hat{D}, \hat{\mu}$ using ML

(use EM) \rightarrow need $p(z|x)$

model is not identifiable; let $W' = WR$ where R $k \times k$ rotation
 $R^T = R^T = I_k$

$$W'W'^T = WR R^T W^T = WW^T$$

because $\Sigma_{xx} = WW^T \in \mathcal{D}$

W is unique up to rotation & sign flip

* why factor analysis?

- richer noise model \triangleright
- easy to plug into bigger model

Canonical correlation analysis

setup is two aligned signals (x_i, y_i)

$\uparrow \searrow$
two views

$$W_x \quad d_x \times k$$

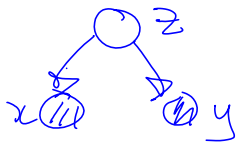
$$W_y \quad d_y \times k$$

the goal is find reduced representation $z_x = W_x^T x$
 $z_y = W_y^T y$

find W_x, W_y s.t. z_x & z_y are maximally correlated

(get generalized e-vector problem)

generative model:



$$z \sim N(0, I_k)$$

$$x|z \sim N(W_x z + \mu_x, \Psi_x)$$

$$y|z \sim N(W_y z + \mu_y, \Psi_y)$$

full covariance

it turns out that MLE in this model related to CCA subspace solution
 [Bach & Jordan 2005]

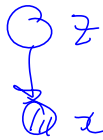
$$\text{cov} \begin{pmatrix} x \\ y \end{pmatrix}, \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} W_x W_x^T + \Psi_x & W_x W_y^T \\ W_y W_x^T & W_y W_y^T + \Psi_y \end{pmatrix}$$

Example: x is image of somebody speaking
 y is sound " " "

State space model & Kalman filter

Go for ... $\rightarrow ?$

factor analysis



SSM unroll in time



$$z_t | z_{t-1} \sim N(Az_{t-1}, B)$$

goal of filtering is to compute $z_t | x_{1:t}$

can still use sum-product algorithm / α - β recursion

where sum is replaced by integral \rightarrow KF
 \rightarrow analytical because have Gaussian variables

* switching KF: $z_t | z_{t-1}$ is mixture of 2 Gaussians

$\Rightarrow z_t | x_{1:t}$ is mixture of 2^t || (intractable)
 \Rightarrow need approximation? (next class)