16/9/19

EE801: Representation Learning

- Review
- PCA
  - Proof
  - Demo
- ML Estimation

- Recall: $C_{XX} \equiv \frac{1}{N} X \cdot X^T$ where X is a $d \times N$ matrix of data points
  - C_{XX} is symmetric $\Rightarrow C_{XX} = E \cdot D \cdot E^T$ — ①

For any symmetric $n \times n$ matrix A with unique evals,

$$A = E \cdot D \cdot E^T \text{ where } E = \begin{bmatrix} | & | & & | \\ \underline{e}_1 & \underline{e}_2 & \cdots & \underline{e}_n \\ | & | & & | \end{bmatrix}, \; D = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

i.e. $A \cdot \underline{e}_1 = \lambda_1 \cdot \underline{e}_1$. ⊛

Further $E \cdot E^T = I$: — ②

Let show this for a pair of evecs $\underline{e}_1, \underline{e}_2$. and correp λ_1, λ_2.

$$\lambda_1 \langle \underline{e}_1, \underline{e}_2 \rangle = (\lambda_1 \underline{e}_1)^T \underline{e}_2$$

$$= (A \cdot \underline{e}_1)^T \underline{e}_2 \quad \text{from ⊛}$$

$$= \underline{e}_1^T \cdot A^T \cdot \underline{e}_2$$

$$= \underline{e}_1^T \cdot A \cdot \underline{e}_2 \quad (\because A = A^T)$$

$$= \underline{e}_1^T \cdot \lambda_2 \cdot \underline{e}_2$$

$$= \lambda_2 \cdot \langle \underline{e}_1, \underline{e}_2 \rangle$$

$$\Rightarrow (\lambda_1 - \lambda_2) \cdot \langle \underline{e}_1, \underline{e}_2 \rangle = 0.$$

$$\because \lambda_1 \neq \lambda_2, \quad \langle \underline{e}_1, \underline{e}_2 \rangle = 0$$

$$\therefore \; C_{XX} = E \cdot D \cdot E^T.$$

Goal: find a linear transformation $Y = PX$ where P is an $d \times d$ matrix such that

$C_{YY} \equiv \frac{1}{N} Y \cdot Y^T$ is a diagonal matrix

Soln: Make use of ① and ②

$$C_{YY} = \frac{1}{N} \cdot Y \cdot Y^T$$

$$= \frac{1}{N} (P \cdot X) \cdot (P \cdot X)^T$$

$$= \frac{1}{N} P \cdot X \cdot X^T \cdot P^T$$

$$= P \cdot C_{XX} \cdot P^T \qquad \left(\because C_{XX} = \frac{1}{N} \cdot X \cdot X^T\right)$$

$$= P \cdot E \cdot D \cdot E^T \cdot P^T \qquad (\because C_{XX} = E \cdot D \cdot E^T)$$

If $P = E^T$, $P \cdot E = E^T \cdot P^T = I$ (from ②)

$$\therefore \; \boxed{P = E^T}$$

- Maximum Likelihood Estimation:

given a set of N i.i.d observations $X = \{\underline{x}_1 \cdots \underline{x}_N\}$ drawn from an underlying distribution $f_{\underline{x}}(\underline{x};\underline{\theta})$ with parameters $\underline{\theta}$.

find $\underline{\theta}_{ML}$ that maximizes the likelihood that X came from $f_{\underline{x}}(\underline{x};\underline{\theta})$

$$\underline{\theta}_{ML} = \underset{\underline{\theta}}{\arg\max}\ L(X;\theta)$$

$$\text{where } L(X;\theta) = \prod_{i=1}^{N} f_x(\underline{x}_i;\underline{\theta})$$