Lecture 16 PCA & ICA

1. PCA

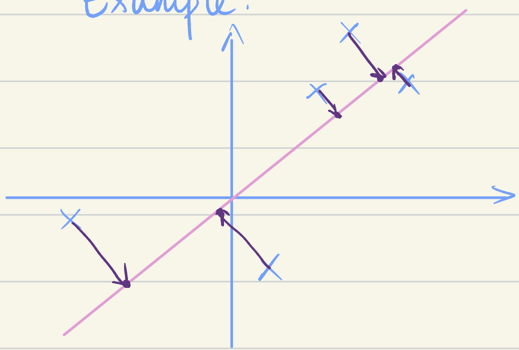
Assume $\{x^{(i)}\}_{i=1}^{n}$, $x^{(i)} \in \mathbb{R}^d \approx \mathbb{R}^k$?

① Standardize

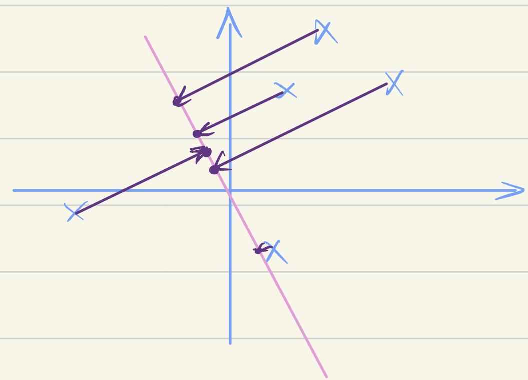
$$x_j^{(i)} := \frac{x_j^{(i)} - \mu_j}{\sigma_j} \quad \left(\mu_j = \frac{1}{n}\sum_{i=1}^{n} x_j^{(i)},\ \sigma_j^2 = \frac{1}{n}\sum_{i=1}^{n}(x_j^{(i)} - \mu_j)^2\right)$$

② Find the subspace which the variance of the projected data is maximized.

Example:



larger variance.



Parameter: the low dimensional subspace $u \in \mathbb{R}^d$, unit length

$\text{Proj}(u)\vec{x}: \quad \frac{uu^T}{u^Tu}\vec{x} = (\vec{x}^{(i)}u)\vec{u}$

Find u such that:

$$\max_u \frac{1}{n}\sum_{i=1}^{n} \|\text{Proj}(u)\vec{x}\|^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} \|(x^{(i)T}u)u\|^2$$

$$= \frac{1}{n}\sum_{i=1}^{n} (x^{(i)T}u)^2 = \frac{1}{n}\sum_{i=1}^{n} x^{(i)T}u\, x^{(i)T}u$$

$$= \frac{1}{n}\sum_{i=1}^{n} x^{(i)T}uu^Tx^{(i)} = \frac{1}{n}\sum_{i=1}^{n} u^Tx^{(i)}x^{(i)T}u = u^T\Big[\overbrace{\frac{1}{n}\sum_{i=1}^{n} x^{(i)}x^{(i)T}}^{\text{sample covariance matrix}}\Big]u$$

$$\Rightarrow u = \underset{u}{\text{argmax}}\; u^T\Big[\frac{1}{n}\sum_{i=1}^{n} x^{(i)}x^{(i)T}\Big]u$$

$$* \; \underset{u}{\text{argmax}}\; u^TAu \Rightarrow \text{Eigenvector of the largest eigenvalue of } A$$

$$\Rightarrow y^{(i)} = \begin{bmatrix} u_1^Tx^{(i)} \\ u_2^Tx^{(i)} \\ \vdots \\ u_k^Tx^{(i)} \end{bmatrix} \in \mathbb{R}^k$$

2. Summary of the 4 unsupervised learning algorithms

	Non Probabilistic	Probabilistic (use EM)	
Clustering	K-Means	GMM	Classification
Subspace	PCA (subspace: XU^T)	Factor Analysis (subspace: z, $n \ll d$)	Regression

3. ICA

Example: Speaker 1 (S_1) Speaker 2 (S_2)

Microphone 1 (X_1) Microphone 2 (X_2)

time i: $S^{(i)} = (S_1^{(i)}, S_2^{(i)}) \in \mathbb{R}^d$, $X^{(i)} = (x_1^{(i)}, x_2^{(i)}) \in \mathbb{R}^d$, $d=2$

W is an unmixing matrix $\Rightarrow s^{(i)} = Wx^{(i)}$

Assumptions: $\#s = \#x$, $s = Wx$

S_j is independent of S_k, $j \neq k$.

S_j is non Gaussian.

$$P(x) = \prod_{s=1}^{d} P_s(w_j^T x) \cdot |W|$$

$W = \begin{bmatrix} - w_j \\ \cdots \end{bmatrix}$ is an unmixing matrix. $\quad WS = X$

$P_s \sim$ logistic distribution

(CDF of logistic: $F(x) = \frac{1}{1+e^{-x}} = \sigma(x)$;

PDF of logistic: $f(x) = \sigma(x)(1-\sigma(x))$)

$$\Rightarrow l(W) = \sum_{i=1}^{n} \left[\left(\sum_{j=1}^{d} \log\left[\sigma(x^{(i)})(1-\sigma(x^{(i)}))\right] \right) + \log|W| \right]$$

using gradient descent: $\quad W := W + \alpha \left(\begin{bmatrix} 1-2\sigma(w_1^T x^{(i)}) \\ 1-2\sigma(w_2^T x^{(i)}) \\ \vdots \end{bmatrix} x^{(i)T} + (W^T)^{-1} \right)$