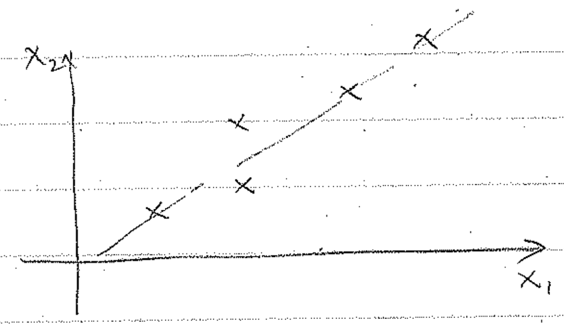
11/19/13

DIMENSIONALITY REDUCTION : PCA

① Goal: $\vec{x} \in \mathbb{R}^d \longrightarrow \vec{z} \in \mathbb{R}^k \quad k<d$



Often $\vec{x}$ may live in a high-dimensional space $\mathbb{R}^d$ but really occupies a small "lower dimension" subspace or manifold. The above figure show we really measured 1D information (with some noise)

② KL - Transform [No-dim-red yet]

$\vec{x} \in \mathbb{R}^d$ input features

W.L.O.G assume $E[X] = 0$

otherwise $X \leftarrow X - E[X]$ "center the data"

Now, Cov matrix $\Sigma_X = E[(X - E[X])(X - E[X])^T]$

$$= E[XX^T] = \begin{bmatrix} & & E(x_i x_j) \\ & E[x_i^2] & \\ & & \ddots \end{bmatrix}_{d \times d}$$

Goal: We want to transform features (not project yet)

$Z = U^T X$ where U is an orthonormal matrix

$$U = \begin{bmatrix} \vdots & & u_i & & \vdots \end{bmatrix}_{d \times d} \qquad U_i^T U_j = \begin{cases} 0 & \text{if } i \neq j \quad \text{(orthogonal)} \\ 1 & \text{if } i = j \quad \text{(unit norm)} \end{cases}$$

Find U s.t $\vec{z} \in \mathbb{R}^d$ are un-correlated
ie

$$\Sigma_z = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{bmatrix}_{d\times d}$$

How?

$$\Sigma_z = E[(z - E[z])(z - E[z])^T]$$
$$= E[zz^T] \quad \left[\text{Recall: } E[z] = E[U^T x] = U^T E[x] = 0\right]$$
$$= E[(U^T x)(U^T x)^T]$$

$$= E[U^T x x^T U]$$
$$= U^T E[xx^T] U = U^T \Sigma_x U$$

So

$$\Sigma_z = U^T \Sigma_x U$$

$$\Rightarrow \underbrace{UU^T}_{I} \Sigma_x U = U \Sigma_z$$

$$\Rightarrow \Sigma_x U = U \Sigma_z = U \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_d \end{bmatrix}$$

$$\Rightarrow \Sigma_x U_i = \lambda U_i \quad \boxed{\text{Eigen decomposition!}}$$

So Algorithm

$\rightarrow$ $[\underset{\text{evec}}{U}, \underset{\text{eval}}{\Lambda}] = \text{eig}(\hat{\Sigma}_x)$ $\quad \hat{\Sigma}_x$ = estimated cov matrix from data

$$= \frac{1}{N} \sum \vec{x}_i \vec{x}_i^T$$

after centering

$\rightarrow$ $\vec{z} = U^T \vec{x}$

$\rightarrow$ Now $\vec{z}$'s are uncorrelated $\quad E[z_i z_j] = 0$

$$E[z_i^2] = \lambda_i$$

2.1 Whitening of data:

Let $\vec{z} = \Lambda^{-1/2} U^T \vec{x}$

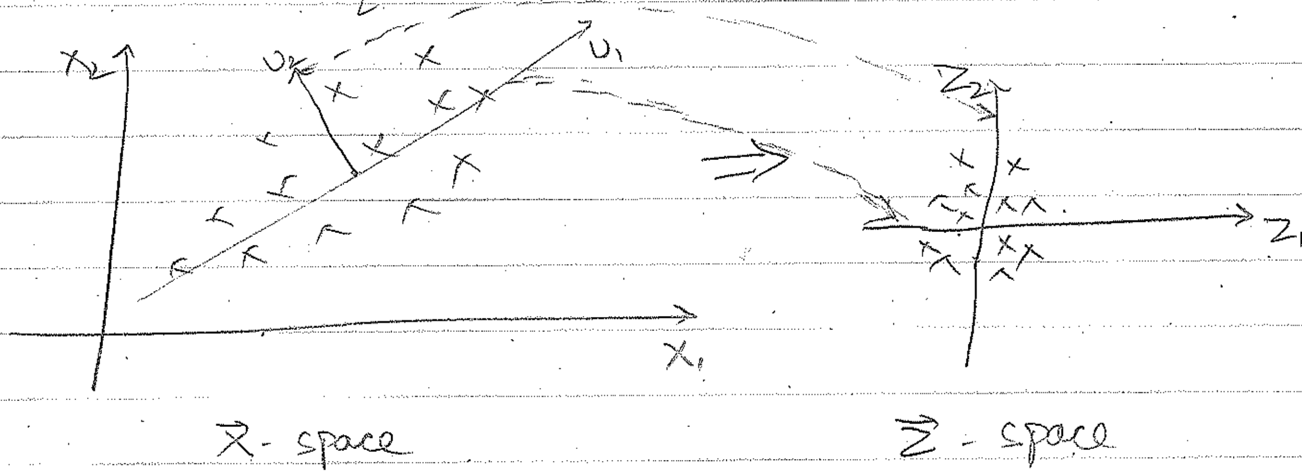
$$= \underbrace{\begin{bmatrix} \frac{1}{\sqrt{\lambda_1}} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\sqrt{\lambda_d}} \end{bmatrix} U^T x}_{}$$

After projecting on $U_1 \ldots U_d$ also "normalize" the co-ordinates

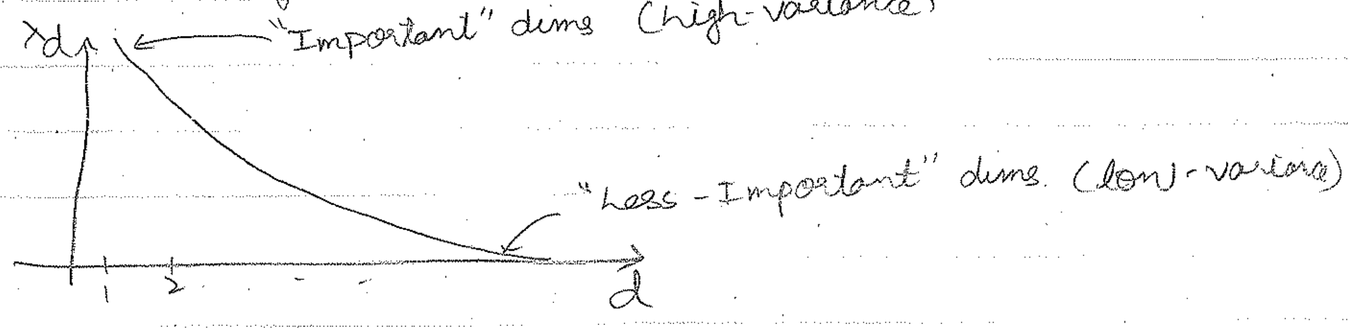
Now $\Sigma_z = E[\vec{z}\vec{z}^T]$

$$= \Lambda^{-1/2} \underbrace{U^T \Sigma_x U}_{\begin{bmatrix} \lambda_1 & \\ & \ddots \\ & & \lambda_d \end{bmatrix}} \Lambda^{-1/2} = \boxed{I_{d \times d}}$$

Ah! data looks like white noise



Plot of Sorted Eigen-values



③ MAX-VARIANCE VIEW of PCA

Notice in KLT $\quad z = U^T x$

$$= \underbrace{\begin{bmatrix} — U_1^T — \\ — \; — \\ — U_d^T — \end{bmatrix}}_{\text{project onto eig-vectors}} x$$

Also notice this means $\quad x = Uz = \underbrace{\begin{bmatrix} | & & | \\ U_1 & -- & U_d \\ | & & | \end{bmatrix}}_{x \text{ as a linear combination of } z} z$

How about we reduce dim by ignoring less-variance dims

$$\vec{z} = \underbrace{\begin{bmatrix} — U_1^T — \\ — U_k^T — \end{bmatrix}}_{\text{top-}k \text{ only}} x$$

Formally, $\quad \max_{\|U_1\|=1} \frac{1}{N} \sum_{i=1}^{N} (U_1^T \vec{x}_i)^2 \qquad$ [Maximize variance after projection]

Solution is eigen vectors again

$\max_{\substack{\|U_2\|=1 \\ U_2^T U_1 = 0}} \frac{1}{N} \sum_{i=1}^{N} (U_2^T \vec{x}_i) \qquad$ Find another direction orthogonal to previous one

$\max_{\substack{\|U_k\|=1 \\ U_k^T U_i = 0 \\ i=1,\dots,k-1}} \frac{1}{N} \sum_{i=1}^{N} (U_k^T x_i)^2$

④ MIN-Reconstruction Error View

Note in KLT $x = Uz$. If we use d-dims, this give perfect reconstruction of x. How about if we use fewer dims?

Formally, find directions $U_1 \ldots U_k$ & new co-ordinates $\vec{z}_1 \in \mathbb{R}^k \ldots \vec{z}_N \in \mathbb{R}^k$

$$\min_{\substack{U_1 \ldots U_k \\ \vec{z}_1, \ldots, \vec{z}_N}} \text{error} = \sum_{i=1}^{N} \|\vec{x}_i - \underbrace{\tilde{x}_i}_{\text{reconstruction}}\|^2$$

$$= \sum_{i=1}^{N} \|\vec{x}_i - \sum_{j=1}^{k} U_j \vec{z}_j\|_2^2$$

≡ Leads to same solution

$$U = \text{top eig-vecs}$$
$$\vec{z}_i = U^T \vec{x}_i$$

⑤ PCA via SVD

Consider an image (100×100) $\quad \vec{x}_i \in \mathbb{R}^{10{,}000}$

So data-matrix $X = \begin{bmatrix} \text{---} \; \vec{x}_1^T \; \text{---} \\ \text{---} \; \vec{x}_N^T \; \text{---} \end{bmatrix}_{N \times d} \quad d = 10{,}000$

PCA requires $\text{eig}\left(\frac{1}{N} X^T X\right) \quad O(d^3)$

SVD of X $\quad X = \underbrace{U}_{\text{orthonormal}} \underbrace{S}_{\text{diagonal}} \underbrace{V^T}_{\text{orthonormal}} = \underbrace{\sum s_{ii} u_i v_i^T}_{\text{Generalization of eigen-values of non-square matrices}}$

Notice $X^T X = (USV^T)^T USV^T$

$= V S^T \underbrace{U^T U}_{I} S V^T$

$= V S^T S V^T$

$= V \underbrace{S^2}_{\substack{\text{diagonal} \\ \Rightarrow \text{symetric}}} V^T$

$\Rightarrow$ S_{ii} is eigen-value of $X^T X$

& V^T = eigen-vectors of $X^T X$

Economy SVD can be done in $O(Nd \min(N,d))$ time

$= O(N^2 d)$ time

Much Better!