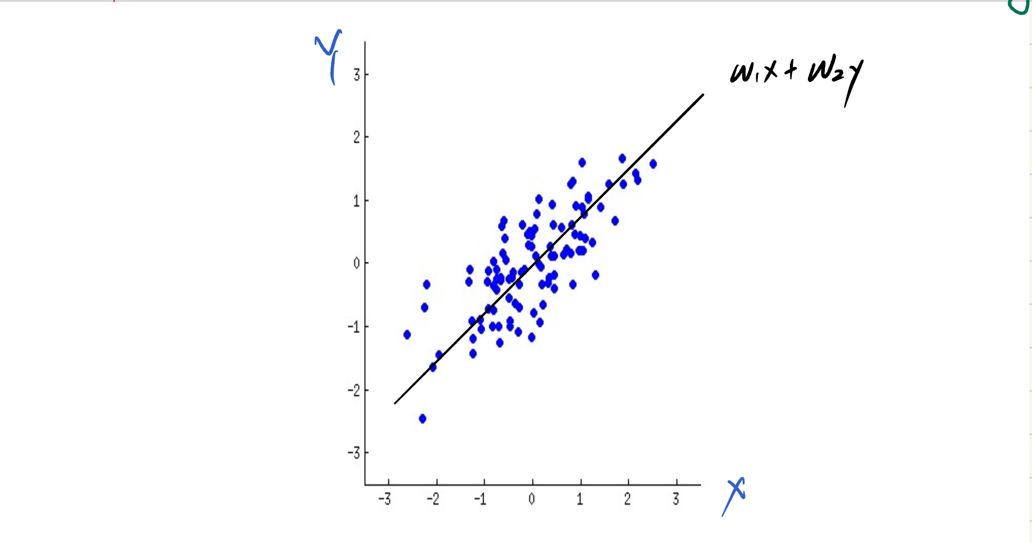
Principal Component Analysis and SVD

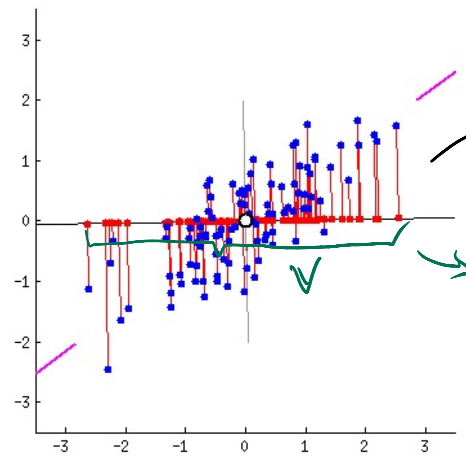
Feb 07 2024

CoralQuant PCA Workshop

Da Gong

Dimensionality Reduction

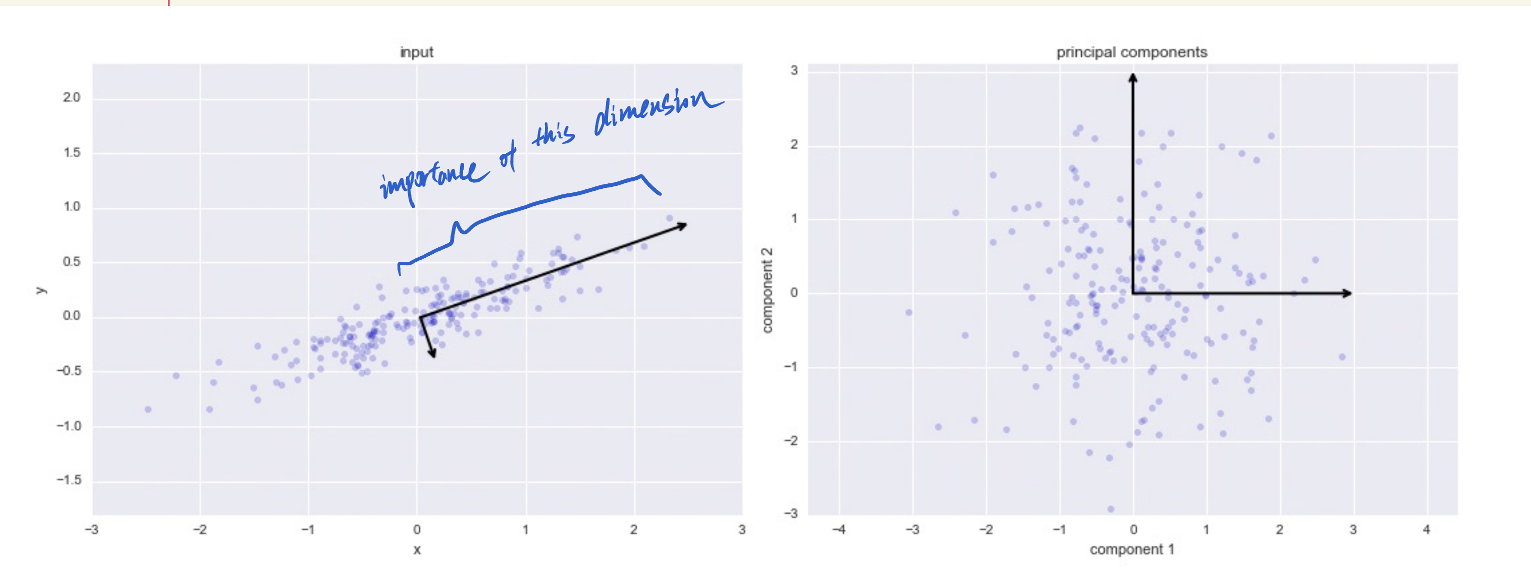




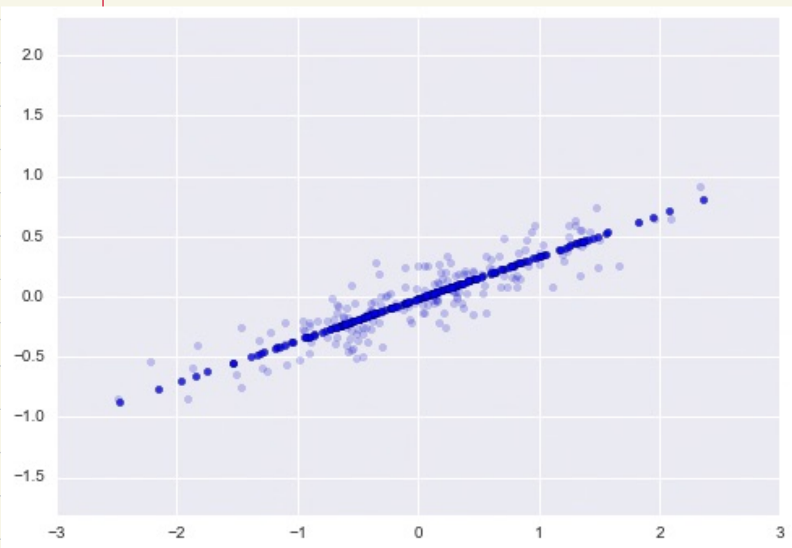
minimum

$\rightarrow J =$ Squared Error

$\rightarrow$ largest V



Dimensionality Reduction



OLS

Eigen value & Eigen vector

$$\underset{n\times n}{A}X = \underset{\text{scalar}}{\lambda}\underset{\text{eigenvector}}{X} \Rightarrow (\lambda I - A)X = 0$$

assume $X \neq 0 \Rightarrow \lambda I - A$ is not invertible

$\Rightarrow \det(\lambda I - A) = 0$

e.g. $A = \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix}$ & $\det(\lambda I - A) = 0$

$\Rightarrow \det\left(\lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix} \right) = 0$

$\det \begin{bmatrix} \lambda+5 & -2 \\ 7 & \lambda-4 \end{bmatrix} = 0 \Rightarrow \lambda^2 + \lambda - 6 = 0$

$\Rightarrow \begin{cases} \lambda_1 = 2 \\ \lambda_2 = -3 \end{cases}$

$\Rightarrow$ find all $X \neq 0$ for each λ

$\lambda_1 = 2$: $AX = 2X \Rightarrow \left(2\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \begin{bmatrix} -5 & 2 \\ -7 & 4 \end{bmatrix} \right) \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\Rightarrow X_1 = \begin{bmatrix} 2 \\ 7 \end{bmatrix}$

$\lambda_2 = -3$: $\Rightarrow X_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

e.g. $A = \begin{bmatrix} 0, & 2 \\ 2, & 3 \end{bmatrix} \Rightarrow X_1 = \begin{bmatrix} -0.89 \\ -0.44 \end{bmatrix} \quad X_2 = \begin{bmatrix} 0.44 \\ -0.89 \end{bmatrix}$

Symetric Matrix

$X_1^T \cdot X_2 = 0$

"Python Example"

Example: A : square matrix

$A \cdot \vec{u} = \lambda \vec{u}$, λ is eigenvalue and

$\vec{u}$ is eigenvector, assume $u^T u = 1$

Suppose there are n solutions:

$(\lambda_1, u_1), (\lambda_2, u_2) \text{ ---- } (\lambda_n, u_n)$

largest ↗ (for λ_1), smallest ↗ (for λ_n)

$$A \cdot \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} | & | & & | \\ \lambda_1 u_1 & \lambda_2 u_2 & \cdots & \lambda_n u_n \\ | & | & & | \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

$$A \cdot U = U \cdot \Lambda$$

if U is invertible, AKA, $u_1 \cdots u_n$ span $\mathbb{R}^n$. (they are linearly independent)

orthogonal to each other

$$\Rightarrow A = U \cdot \Lambda \cdot U^{-1}$$

if A is symmetric and real $\Rightarrow \lambda \in \mathbb{R}$, $U^{-1} = U^T$ ☆ $\left(u_i^T u_j = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{o/w} \end{cases} \right)$

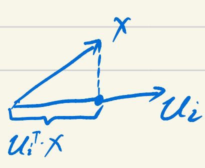
$$A = \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{bmatrix} \begin{bmatrix} - & u_1^T & - \\ & \vdots & \\ - & u_n^T & - \end{bmatrix}$$

$$\begin{bmatrix} | & | & & | \\ \lambda_1 u_1 & \lambda_2 u_2 & \cdots & \lambda_n u_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} - & u_1^T & - \\ & | & \\ - & u_n^T & - \end{bmatrix}$$

$$= \sum_i \lambda_i u_i u_i^T$$

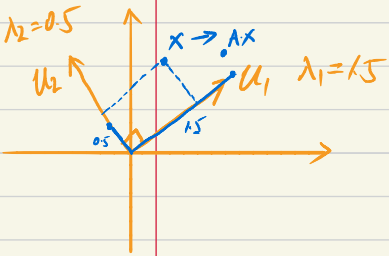
this inner production is projection of X on u_i

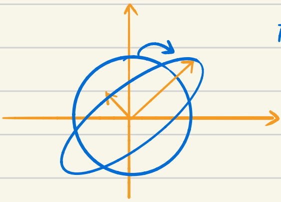
$$\Rightarrow AX = \left(\sum_i \lambda_i \cdot u_i \cdot u_i^T \right) X = \sum_i \lambda_i \cdot u_i \cdot (u_i^T \cdot X)$$



$$= \sum u_i \left(\lambda_i \cdot (u_i^T \cdot x)\right)$$

Scale by λ_i





the circle transform to oval by eigenvector and eigenvalue

Dimensionality Reduction

Kevin Murphy (20.1)

Principal Component Analysis

- a single vector $\underset{d\times 1}{w}$: vector of reconstruction, z_i is scalar: Latent factor

$\underset{d\times 1}{x_i}$ is the true features of hypos

$$J = \sum_{i=1}^{m} \underbrace{(x_i - z_i w)^T (x_i - z_i w)}_{\text{squared error of difference}}, \quad x_i \text{ is given.}$$

- Assume data are centered (data are substract from mean)

$$J = \sum_{i=1}^{m} \cancelto{\perp z}{x_i^T x_i} - 2 z_i x_i^T w + z_i^2 w^T w$$

plug back

$\frac{\partial J}{\partial z_k} = 0$: $2 x_k^T w = 2 z_k w^T w$, W.L.O.G assume $w^T w = 1$ (length of w is 1)

$\Rightarrow \boxed{x_k^T w = z_k}$, length of x's projection on w is z_k

$\Downarrow$

$\underset{\text{scalar}}{w^T x_k} = x_k^T w = z_k$

$\Rightarrow J = \text{Const} + \sum_i -2\, w^T x_i x_i^T w + w^T x_i \cdot x_i^T w$

$J = \text{const} - \underbrace{w^T(\sum_i x_i x_i^T)\, w}_{\star}$, we wanna have J to be min which means X's projection on W as large as possible.

and we know that, $\sum_i x_i x_i^T = m \cdot \hat{\Sigma}^{ML}$

So minimalze J is equavelent to max $w^T \cdot \hat{\Sigma}^{ML} w$

arg max $\mathcal{L} = w^T \hat{\Sigma} w - \overset{\text{scalar}}{\lambda} (w^T w - 1)$, λ is Lagragian multiplier

$\nabla_w \mathcal{L} = 0$: $2\hat{\Sigma} w - 2\lambda w = 0 \Rightarrow \hat{\Sigma} w = \lambda w$ remind Eigenvalue/vector

$\Rightarrow w^T \hat{\Sigma} w = w^T \lambda w = \lambda\, w^T w = \lambda$ ← Largest Eigenvalue and associated EigenVector!

- You should project on the largest eigenvector of $\hat{\Sigma}$ (of centered data).

the empirical covirance matrix

what if, we don't want to use just one number to describle Hypo??

$\Rightarrow \min \sum_{i=1}^{m} \| x_i - (z_{i1} w_1 + z_{i2} w_2) \|_2^2$ two dimensions.

Result: $w_1 \sim$ eigenvector associate largest eigenvalue of $\hat{\Sigma}$

$w_2 \sim$ eigenvector associate second largest eigenvalue of $\hat{\Sigma}$

$Z_{i1} = X_i^T w_1, \quad Z_{i2} = X_i^T w_2$

General Result:

V are eigenvector of $\hat{\Sigma}$:

$$\begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_d \\ | & | & & | \end{bmatrix}$$

$$W = \begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_k \\ | & | & & | \end{bmatrix} \quad k \le d$$

$$\underset{m \times k}{Z} = \underset{m \times d}{X} \cdot \underset{d \times k}{W}$$, Z is projection X on W

1. W are directions for best squared error reconstruction
2. W are directions w/. largest empirical virances.

Cookbook:

method 1

1. Central data : $X_i - \hat{\mu}_{ML} \Rightarrow X_i$
2. $\hat{\Sigma} = X^T X \cdot \frac{1}{m}$ require a lot computational power
3. get $e\vec{v}$ of $\hat{\Sigma}$ $\xRightarrow{\text{keep largest ones}}$ $V \Rightarrow W \Rightarrow XW = Z$

method 2 (better way and faster and more stable)

1. Same

2. sigular value decomposation

diagonal matrix of sigular values

$$\underset{m\times d}{X} = \underset{m\times d}{U} \cdot \underset{d\times d}{D} \cdot \underset{d\times d}{V^T}$$

matrix of orthogonal columns.

$$\Rightarrow X^TX = VDU^TUDV^T = VD^2V^T$$

$U^TU = I$

eivalue decomposition of X^TX

So after get $X = U \cdot D \cdot V^T$, throw away U.

V is the same V in method 1.

$\frac{D^2}{m}$ is the eigenvalue of $\hat{\Sigma}$ in method 1.

Steve Brunton Youtube Channel

left S. Vec   S. Value   right S Vec

$$\underset{n \times m}{X} = \begin{bmatrix} | & | & & | \\ x_1, & x_2, & \cdots & x_m \\ | & | & & | \end{bmatrix} = U \Sigma V^T$$

x_1

1 million pixels in 1 vector

$n = 1$ million

$$= \underset{n \times n}{\begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix}} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_m \\ & & 0 & \end{bmatrix} \underset{m \times m}{\begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_m \\ | & | & & | \end{bmatrix}^T}$$

$U U^T = U^T U = I_{n \times n}$ $V V^T = V^T V = I_{m \times m}$ Σ diagonal, $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_m \geq 0$

$$\begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix} \qquad V^T = \begin{bmatrix} v_1^T & \text{---} \\ v_2^T & \text{---} \\ \vdots & \end{bmatrix}$$

eigen face, describe V of cols of x

mixture of all eigenfaces

"economy SVD"

$$\underset{n \times n}{\begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_n \\ | & | & & | \end{bmatrix}} \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_m \\ & & 0 & \end{bmatrix} \underset{m \times m}{\begin{bmatrix} | & | & & | \\ v_1 & v_2 & \cdots & v_m \\ | & | & & | \end{bmatrix}^T} = \underset{n \times m}{\hat{U}} \underset{m \times m}{\hat{\Sigma}} \underset{m \times m}{V^T}, \quad n \gg m$$

$$= \sigma_1 u_1 v_1^T + \sigma_2 u_2 v_2^T + \cdots + \sigma_m u_m v_m^T + 0$$

$$= \sigma_1 \, u_1 \, v_1^T + \cdots + \sigma_m \, u_m \, v_m^T$$

truncate at Rank r

$$\approx \underset{n \times r}{\tilde{U}} \underset{r \times r}{\tilde{\Sigma}} \underset{r \times m}{\tilde{V}}$$

$\tilde{U}^T \tilde{U} = I_{r \times r}$ $\tilde{U} \tilde{U}^T \neq I$

correlation matrix (cov)

$$\underbrace{X^T X}_{m\times m} = \underbrace{\begin{bmatrix} & & \end{bmatrix}}_{m\times n} \underbrace{\begin{bmatrix} x_1 & x_2 & \cdots \end{bmatrix}}_{n\times m} = \begin{bmatrix} x_1^T x_1 & x_1^T x_2 & \cdots & x_1^T x_m \\ \vdots & \vdots & \ddots & \vdots \\ x_m^T x_1 & x_m^T x_2 & \cdots & x_m^T x_m \end{bmatrix}$$

$x_i^T x_j$ ↑ ⇒ similar faces

$= \langle x_i, x_j \rangle$

$$X^T X = V \hat{\Sigma} \underbrace{\hat{U}^T \hat{U}}_{I_{m\times m}} \hat{\Sigma} V^T = V \hat{\Sigma}^2 V^T$$

$$\Rightarrow X^T X \cdot V = V \cdot \hat{\Sigma}^2$$

V → eigen vec, $\hat{\Sigma}^2$ → eigen val

- PCA

$B = X - \bar{X}$, $C = B^T \cdot B$

$CV = V \cdot D$

$T = BV$

principal components

$\lambda = \sigma^2$

$$V \text{ captured by PCA} = \frac{\sum_{k=1}^{r} \lambda_k}{\sum_{k=1}^{n} \lambda_k}$$