Dimensionality Reduction

Part II

PCA: Part V

$$z_1 = w^1 \cdot x$$

$$\bar{z_1} = \frac{1}{N}\sum z_1 = \frac{1}{N}\sum w^1 \cdot x = w^1 \cdot \frac{1}{N}\sum x = w^1 \cdot \bar{x}$$

$$Var(z_1) = \frac{1}{N}\sum_{z_1}(z_1 - \bar{z_1})^2$$

$$= \frac{1}{N}\sum_{x}(w^1 \cdot x - w^1 \cdot \bar{x})^2$$

$$= \frac{1}{N}\sum\left(w^1 \cdot (x - \bar{x})\right)^2$$

$$= \frac{1}{N}\sum (w^1)^T (x - \bar{x})(x - \bar{x})^T w^1$$

$$= (w^1)^T \frac{1}{N}\boxed{\sum (x - \bar{x})(x - \bar{x})^T w^1}$$

$$= (w^1)^T Cov(x) w^1$$

$$S = Cov(x)$$

$$(a \cdot b)^2 = (a^T b)^2 = a^T b a^T b$$

$$= a^T b (a^T b)^T = a^T b b^T a$$

Find w^1 maximizing

$$(w^1)^T S w^1$$

$$\|w^1\|_2 = (w^1)^T w^1 = 1$$

PCA: Part VI

Find w^1 maximizing $(w^1)^T S w^1$ $\quad (w^1)^T w^1 = 1$

$S = Cov(x)$ Symmetric Positive-semidefinite
(non-negative eigenvalues)

Using Lagrange multiplier

$$g(w^1) = (w^1)^T S w^1 - \alpha\left((w^1)^T w^1 - 1\right)$$

$$\left.\begin{array}{l} \partial g(w^1)/\partial w_1^1 = 0 \\ \partial g(w^1)/\partial w_2^1 = 0 \\ \vdots \end{array}\right\}$$

$$Sw^1 - \alpha w^1 = 0$$

$$Sw^1 = \alpha w^1$$

w^1 : eigenvector

$$(w^1)^T S w^1 = \alpha (w^1)^T w^1$$

$$= \alpha$$

Choose the maximum one

w^1 is the eigenvector of the covariance matrix S

Corresponding to the largest eigenvalue

PCA: Part VII

Find w^2 maximizing $(w^2)^T S w^2 \quad (w^2)^T w^2 = 1 \quad (w^2)^T w^1 = 0$

$$g(w^2) = (w^2)^T S w^2 - \alpha\big((w^2)^T w^2 - 1\big) \; -\beta\big((w^2)^T w^1 - 0\big)$$

$$\partial g(w^2)/\partial w_1^2 = 0$$

$$\partial g(w^2)/\partial w_2^2 = 0$$

$$\vdots$$

$$Sw^2 - \alpha w^2 - \beta w^1 = 0$$

$$\boxed{0} - \alpha \boxed{0} - \beta \boxed{1} = 0$$

$$= \big((w^1)^T S w^2\big)^T = (w^2)^T S^T w^1$$

$$= (w^2)^T S w^1 \; = \lambda_1 (w^2)^T w^1 \; = 0$$

$$Sw^1 = \lambda_1 w^1$$

$\beta = 0$: $\quad Sw^2 - \alpha w^2 = 0 \quad Sw^2 = \alpha w^2$

w^2 is the eigenvector of the covariance matrix S

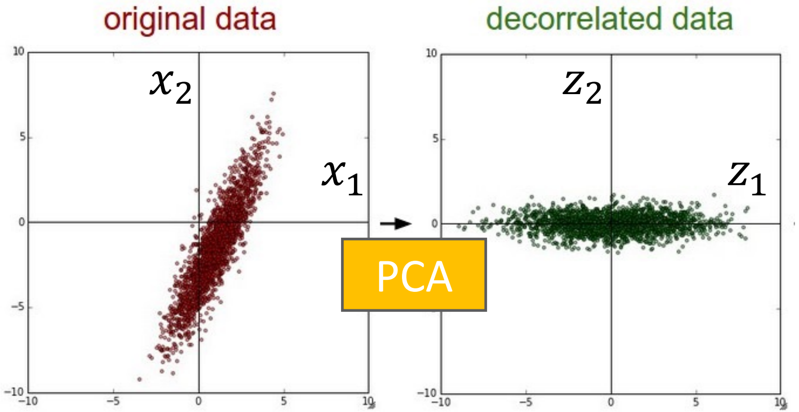
Corresponding to the 2nd largest eigenvalue λ_2

PCA - Decorrelation

$$z = Wx$$

$$Cov(z) = D$$

Diagonal matrix



$$Cov(z) = \frac{1}{N}\sum (z-\bar{z})(z-\bar{z})^T = WSW^T \qquad S = Cov(x)$$

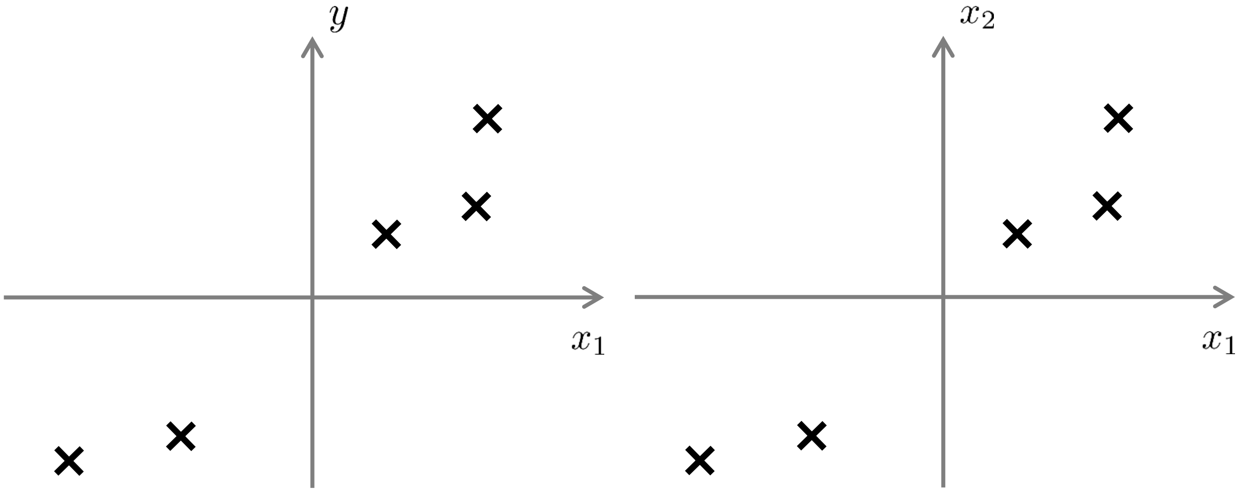
$$= WS\begin{bmatrix} w^1 & \cdots & w^K \end{bmatrix} = W\begin{bmatrix} Sw^1 & \cdots & Sw^K \end{bmatrix}$$

$$= W\begin{bmatrix} \lambda_1 w^1 & \cdots & \lambda_K w^K \end{bmatrix} = \begin{bmatrix} \lambda_1 W w^1 & \cdots & \lambda_K W w^K \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 e_1 & \cdots & \lambda_K e_K \end{bmatrix} = D$$

Diagonal matrix

PCA is not Linear Regression



Data Preprocessing

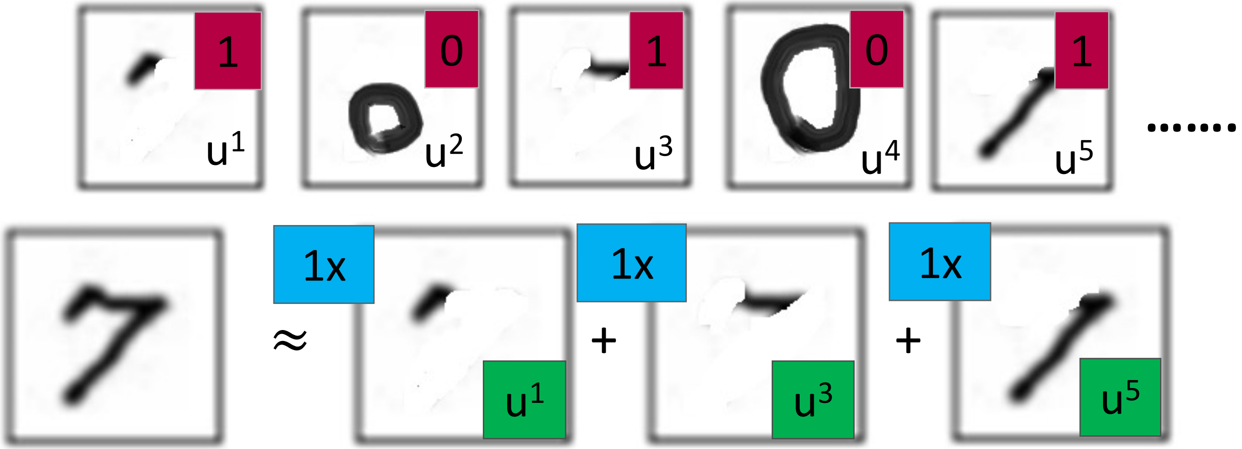
Training set: $x^{(1)}, x^{(2)}, \ldots, x^{(m)}$
Preprocessing (feature scaling/mean normalization):

$$\mu_j = \frac{1}{m} \sum_{i=1}^{m} x_j^{(i)}$$

Replace each $x_j^{(i)}$ with $x_j - \mu_j$.
If different features on different scales (e.g., $x_1 =$size of house, $x_2 =$ number of bedrooms), scale features to have comparable range of values.

PCA – Another Point of View: Part I

Basic Component:



$$\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ \vdots \end{bmatrix}$$

$$x \approx c_1 u^1 + c_2 u^2 + \cdots + c_K u^K + \bar{x}$$

$$\begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_K \end{bmatrix}$$

Represent a digit image

Pixels in a digit image

component

PCA – Another Point of View: Part II

$$x - \bar{x} \approx c_1 u^1 + c_2 u^2 + \cdots + c_K u^K \;\; = \hat{x}$$

Reconstruction error:
$\| (x - \bar{x}) - \hat{x} \|_2$

Find $\{u^1, \dots, u^K\}$ minimizing the error

$$L = \min_{\{u^1,\dots,u^K\}} \sum \left\| (x - \bar{x}) - \underbrace{\left(\sum_{k=1} c_k u^k \right)}_{\hat{x}} \right\|_2$$

PCA: $z = Wx$

$$\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_K \end{bmatrix} = \begin{bmatrix} (w_1)^{\mathrm{T}} \\ (w_2)^{\mathrm{T}} \\ \vdots \\ (w_K)^{\mathrm{T}} \end{bmatrix} x$$

$\{w^1, w^2, \dots w^K\}$ (from PCA) is the component $\{u^1, u^2, \dots u^K\}$ minimizing L

PCA – Another Point of View: Part III

$$x - \bar{x} \approx c_1 u^1 + c_2 u^2 + \cdots + c_K u^K \; = \hat{x}$$

Reconstruction error:
$\| (x - \bar{x}) - \hat{x} \|_2$

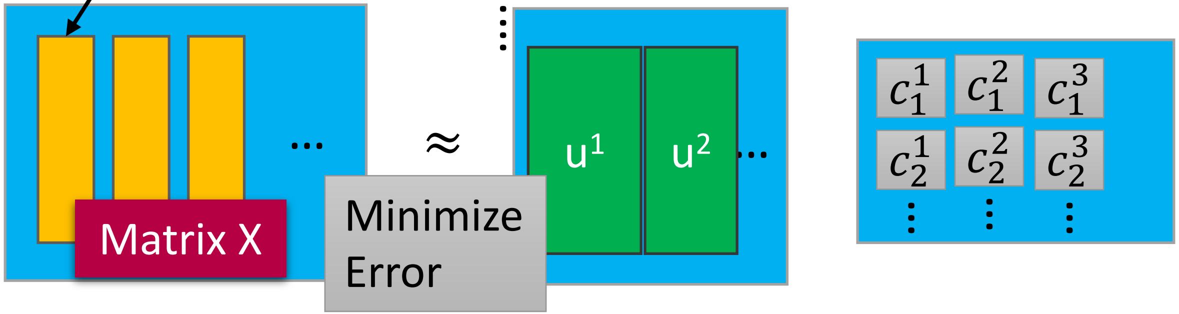
Find $\{u^1, \ldots, u^K\}$ minimizing the error

$$x^1 - \bar{x} \approx c_1^1 u^1 + c_2^1 u^2 + \cdots$$

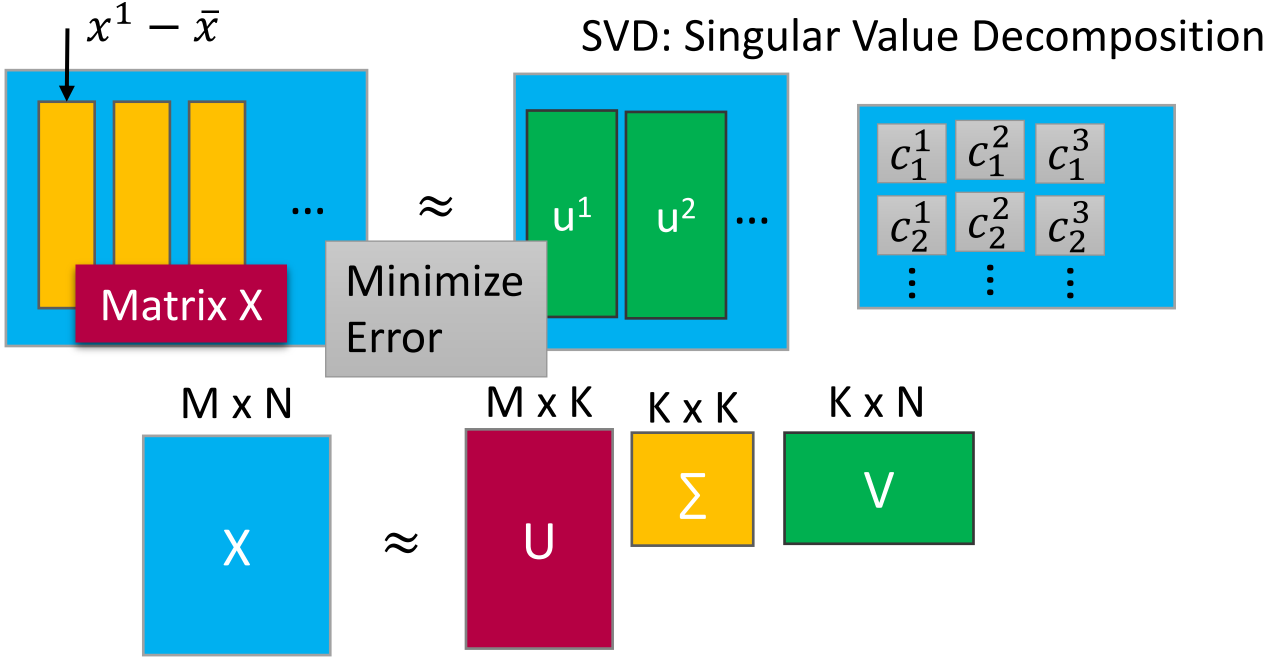
$$x^2 - \bar{x} \approx c_1^2 u^1 + c_2^2 u^2 + \cdots$$

$$x^3 - \bar{x} \approx c_1^3 u^1 + c_2^3 u^2 + \cdots$$

$\vdots$



PCA – Another Point of View: Part IV



K columns of U: a set of orthonormal eigen vectors corresponding to the K largest eigenvalues of XX^T

This is the solution of PCA

Choosing K (Number of Principal Components)

Typically, choose k to be smallest value so that

Average squared projection error

$$\frac{\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)} - x_{approx}^{(i)}\|^2}{\frac{1}{m}\sum_{i=1}^{m}\|x^{(i)}\|^2} \leq 0.01 \qquad (1\%)$$

“99% of variance is retained”