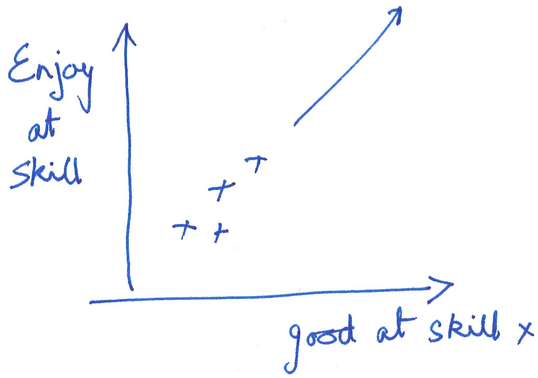
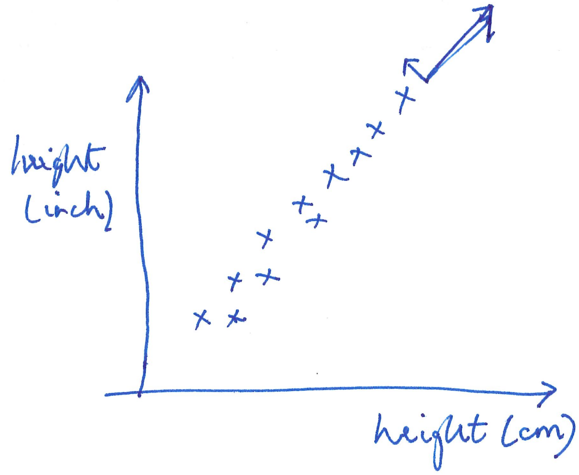
CS 307 Lecture 23

Training set : $S = \{ x^{(i)} \}_{i=1}^{n} \quad x^{(i)} \in \mathbb{R}^d$.

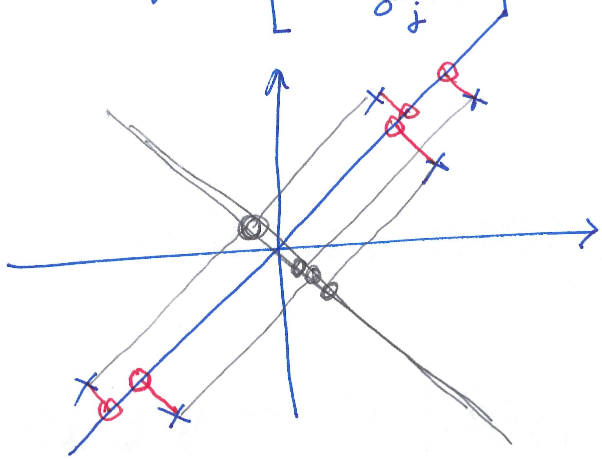




Normalize Data :

For each feature j

1. $\mu_j = \frac{1}{n} \sum x_j^{(i)}$
2. $\sigma_j^2 = \frac{1}{n} \sum (x_j^{(i)} - \mu_j)^2$
3. $\bar{x}_j^{(i)} = \left[\dfrac{x_j^{(i)} - \mu_j}{\sigma_j} \right]$ ⟵ The mean is 0 and variance is 1.



Principal Component Analysis (PCA):

Recall: The projection of vector x along u

$$= \frac{\langle x, u \rangle}{\|u\|_2}$$

If $\|u\|_2 = 1$, length projection of x along u $= \langle x, u \rangle = x^T u$.

Find a vector u. (To reduce the data to one dim)

$$\max. \quad \frac{1}{n} \sum_{i=1}^{n} \left(x^{(i)T} u\right)^2$$

$$\text{s.t} \quad \|u\|_2 = 1$$

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

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

$$= \max_{\|u\|=1} u^T \underbrace{\left(\frac{1}{n} \sum_{i=1}^{n} x^{(i)} x^{(i)T}\right)}_{d \times d} u$$

Let $A = \frac{1}{n} \sum_{i=1}^{n} x^{(i)} x^{(i)T}$

PCA : $\max_{\|u\|=1} u^T A u$.

Proposition: A is a symmetric matrix, ie $A^T = A$.

$$A = \frac{1}{n} \sum_{i=1}^{n} \underbrace{x^{(i)} x^{(i)T}}_{\text{symmetric}}$$

symmetric (Midterm 1, problem 1)

Linear Algebra Recap:

Eigen value / Eigen vector of A: $(A \in \mathbb{R}^{d \times d})$

$$\underbrace{A u}_{\text{eigen vector}} = \underbrace{\lambda}_{\text{eigen value of } A} u.$$

Theorem: Let A be a symmetric matrix. $(A \in \mathbb{R}^{d \times d})$

All eigenvalues of A are real and there are $u_1, u_2 \ldots u_d$ orthonormal eigen vectors of A.

$\underbrace{\text{ortho}}_{\forall i \neq j.\ \langle u_i, u_j \rangle = 0}\underbrace{\text{normal}}_{\|u_i\| = 1}$

$\langle u_i, u_i \rangle = 1$

Eigen values of symmetric A: $\lambda_1 \geq \lambda_2 \geq \lambda_3 \ldots \lambda_d$.

$\rightarrow u_1, u_2 \ldots u_d$ orthonormal.

Define $U = \begin{bmatrix} | & | & | & & | \\ u_1 & u_2 & u_3 & \cdots & u_d \\ | & | & | & & | \end{bmatrix}$ ($u_1, u_2 \ldots u_d$ are the orthonormal eigen vectors of A)

$$U^T U = \begin{bmatrix} - & u_1^T & - \\ - & u_2^T & - \\ & \vdots & \\ - & u_d^T & - \end{bmatrix} \begin{bmatrix} | & | & & | \\ u_1 & u_2 & \cdots & u_d \\ | & | & & | \end{bmatrix} = \begin{bmatrix} u_1^T u_1 & u_1^T u_2 & \cdots & u_1^T u_d \\ & & & \\ u_d^T u_1 & & & u_d^T u_d \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 0 & & 0 \\ 0 & 1 & 0 & \cdot & 0 \\ & & & \\ 0 & \cdot & \cdot & 0 & 1 \end{bmatrix} = I$$

U – orthonormal eigen vectors of A.

$$U^T A U = \begin{bmatrix} - & u_1^T & - \\ & \vdots & \\ - & u_d^T & - \end{bmatrix} \underbrace{A \begin{bmatrix} | & & | \\ u_1 & \cdots & u_d \\ | & & | \end{bmatrix}} = U^T \begin{bmatrix} | & | & & | \\ A u_1 & A u_2 & \cdots & A u_d \\ | & | & & | \end{bmatrix}$$

$$= U^T \begin{bmatrix} | & | & & | \\ \lambda_1 u_1 & \lambda_2 u_2 & \cdots & \lambda_d u_d \\ | & | & & | \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdot & \cdot & \cdot & \lambda_d \end{bmatrix}$$

$$\Lambda = \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_d \end{bmatrix}$$

$$U^T A U = \Lambda \quad \Rightarrow \quad A = U^{\cancel{T}} \Lambda U^T$$

<u>Proposition</u>: $\forall x \in \mathbb{R}^d$. and A is symmetric matrix with maximum eigenvalue λ_1 and minimum eigen value λ_d. Assume $\|x\| = 1$.

$$\lambda_d \leq x^T A x \leq \lambda_1$$

<u>Proof</u>: $A = U \Lambda U^T$ $\quad \Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \lambda_d \end{bmatrix}$ $\quad U$ is orthonormal

$$x^T A x = x^T U \Lambda U^T x = (U^T x)^T \Lambda (U^T x)$$

$$= (U^T x)^T \begin{bmatrix} \lambda_1 (U^T x)_1 \\ \lambda_2 (U^T x)_2 \\ \vdots \\ \lambda_d (U^T x)_d \end{bmatrix}$$

$$= \sum_{i=1}^{d} \lambda_i (U^T x)_i^2$$

$$x^T A x = \sum_{i=1}^{d} \lambda_i (U^T x)_i^2 \leq \sum_{i=1}^{d} \lambda_1 (U^T x)_i^2 = \lambda_1 \sum_{i=1}^{d} (U^T x)_i^2 = \lambda_1 \|U^T x\|^2$$

$$\geq \lambda_d \|U^T x\|^2.$$

$$\Rightarrow \quad \underbrace{\|U^T x\|^2 = \|x\|^2}_{\text{Exercise}} = 1$$

$$\lambda_d \leq x^T A x \leq \lambda_1$$

u_1 is the vector corresponding to λ_1

$$u_1^T \underline{A u_1} = u_1^T \lambda_1 u_1 = \lambda_1 u_1^T u_1 = \lambda_1$$