vs. Fisher's discriminant:
Both projects data on lower dimension
- PCA is un-supervised
- Fisher uses target label y

* **Deriving Equation**

- $X = \begin{bmatrix} \quad \end{bmatrix}$

n points of D dimention

→ Don't start with matrix

→ We want to project one 1 dimension $M = 1$

2016

August

July						2016
S	M	T	W	T	F	S
31					1	2
3	4	5	6	7	8	9
10	11	12	13	14	15	16
17	18	19	20	21	22	23
24	25	26	27	28	29	30

September						2016
S	M	T	W	T	F	S
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

17 Wednesday

→ we have n points each of dimension D

$x_n \quad .. \quad n \in \{1 \dots n\}$

→ We want to project on a direction $\vec{u_1}$

- note $\vec{u_1}$ will also be D dimension
- $|\vec{u_1}| = 1$

$$z_n = \vec{x_n} \cdot \vec{u_1}$$

$$= u_1^T \cdot x_n$$

In matrix notation we generally consider column ~~vector~~

18 Thursday

both u_1 and x_n are column matrices.

→ mean of z_n

$$\bar{z} = \cancel{\frac{1}{N} \sum x_n}$$

$$= u_1^T \cdot \bar{x}$$

→ variance of z_n

$$= \frac{1}{N} \sum \left(u_1^T x - u_1^T \bar{x} \right)^2$$

u_1 D $[\ ]_{D\times n}$

$x-\bar{x}$ $D\times n$

$(x-\bar{x})^T$ $n\times D$

19 Friday

$$= \frac{1}{N}\sum_{i=1}^{n} (u_1^T)^2 (x-\bar{x})^2$$

$$= \frac{1}{N}\sum u_1^T \cdot u_1 \, (x-\bar{x})(x-\bar{x})^T$$

$1\times D$ $D\times D$ $D\times 1$

$$\text{var} = u_1^T S u_1$$

1

$$S = \frac{1}{N}\sum (x_n-\bar{x})(x_n-\bar{x})^2$$

$$x_n-\bar{x} = D\times n$$

20 Saturday

$$[\ ]_{D\times n}$$

We want to maximize

$$\text{var} = u_1^T S u_1$$

s. t. $u_1^T u_1 = 1$

<u>Langrangian</u>

$$L = u_1^T S u_1 + \lambda_1 (1 - u_1^T u_1)$$

21 Sunday

$$\frac{\partial L}{\partial \lambda_1}$$

$=$

$\Rightarrow$

22

2016 August

September 2016

S	M	T	W	T	F	S
				1	2	3
4	5	6	7	8	9	10
11	12	13	14	15	16	17
18	19	20	21	22	23	24
25	26	27	28	29	30	

21 Sunday

$\frac{\partial L}{\partial \lambda_1} = 0$

$\Rightarrow$

$\frac{\partial L}{\partial u_1} = 0$

$\Rightarrow 2Su_1 + \lambda_1(-2u_1) = 0$

$\Rightarrow \boxed{Su_1 = \lambda_1 u_1}$

defination of eigen-vectors
it just scales

$\Rightarrow$ Muliply by u_1^T

$u_1^T S u_1 = u_1^T \lambda_1 u_1$

22 Monday

$= \lambda_1$

because $u_1^T u_1$ is 1

$\Rightarrow$ And from (2)

variance $= \lambda_1$