Principal Component Analysis.

Dimensionality Reduction

Latent Variable Modeling
(Hidden)

$x_i \to$ one data instance $x_i \in R^D$

$z_i \to$ hidden / latent unknown variable.

If $z_i \in \{1, \ldots, K\}$
↳ clustering
↳ Mixture of models

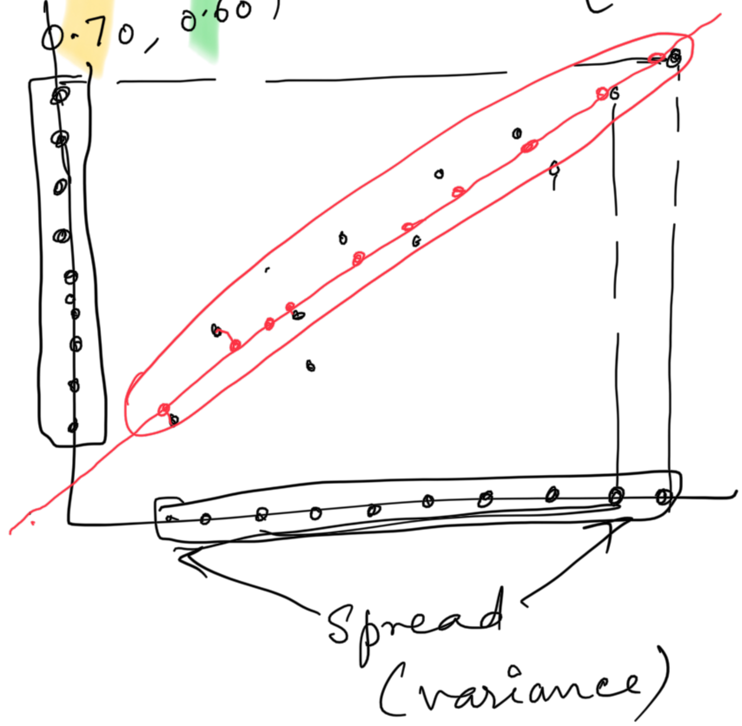
If $z_i \in R^d$
where $d << D$
↳ This is a dimensionality reduction problem.

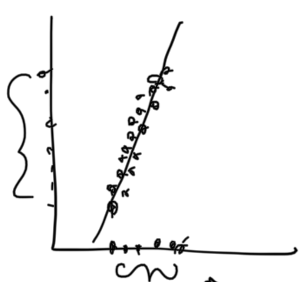
Given X: $\begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \end{bmatrix}$ find Z: $\begin{bmatrix} z_1^T \\ z_2^T \\ \vdots \end{bmatrix}$

$[x_N^T]$ (z_N)

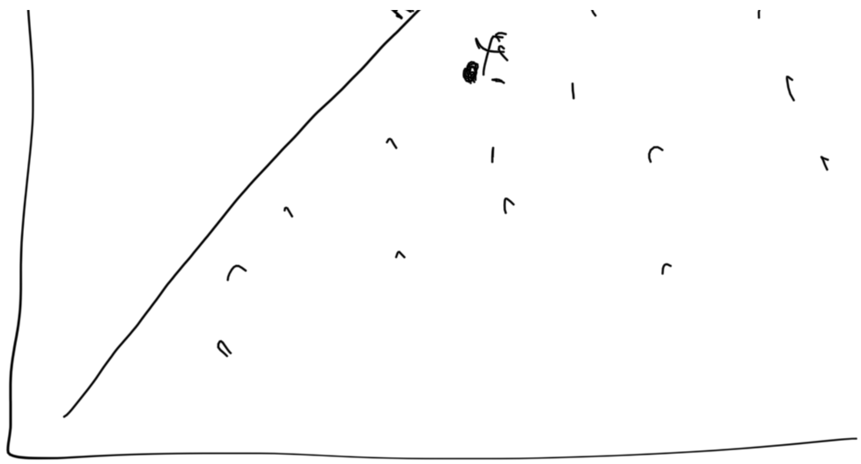
Such that ~~#~~ some property is preserved.

$$X = \begin{bmatrix} 0.20, & 0.10 \\ 0.35, & 0.40 \\ 0.50, & 0.20 \\ 0.65, & 0.10 \\ 0.70, & 0.60 \end{bmatrix} \qquad Z = \begin{bmatrix} ? \end{bmatrix}$$





unit vector



$x_i^T \hat{u}$ — Projection of i^{th} data instance

$$Z = \begin{bmatrix} x_1^T \hat{u} \\ x_2^T \hat{u} \\ \vdots \\ x_N^T \hat{u} \end{bmatrix}$$

Let us assume that the data has 0 mean.

$$\Rightarrow \sum_{i=1}^{N} x_i = 0$$

$$\text{var}: \frac{1}{N} \sum_{i=1}^{N} (x_i^T \hat{u})^2$$

$$= \frac{1}{N} \sum_{i=1} \underline{\hat{u}}^T x_i x_i^T \underline{\hat{u}}$$

$$= \hat{u}^T \underbrace{\left(\frac{1}{N} \sum_{i=1}^{N} x_i x_i^T \right)} \hat{u}$$

This is the sample covariance matrix of X

$$S = \frac{1}{N} \sum_{i=1}^{N} (x_i - \underline{\bar{x}}) (x_i - \underline{\bar{x}})^T$$

$$= \frac{1}{N} \sum_{i=1}^{N} x_i x_i^T \quad [\text{if } \bar{x} \text{ is } 0]$$

$$\text{var} = \underline{\hat{u}^T S \hat{u}}$$

$$\boxed{\begin{aligned} &\arg\max_{\hat{u}} \ \hat{u}^T \mathbf{S} \hat{u} \\ &\text{subject to} \ \underline{\hat{u}^T \hat{u} = 1} \end{aligned}}$$

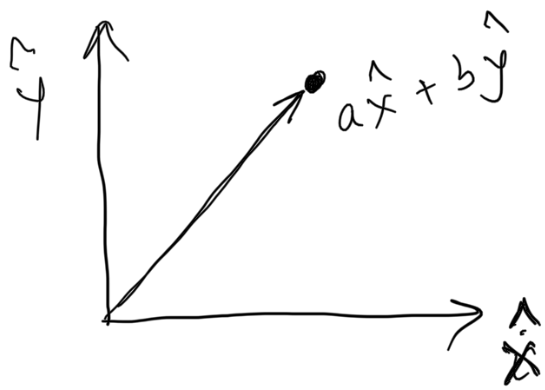
$\hat{u}$ should be unit length

$$\boxed{\hat{u}^T S \hat{u} + \lambda (\hat{u}^T \hat{u} - 1)}$$

Wed May 05

PA3 due on May 9th (Sunday) at 11:59 PM

Final exam for MPS Batch *
is on May 13th at 4:00 PM
4:00 - 6:00 PM



$$x_i^T \hat{u}$$

(D×1) (D×1)

$$\frac{1}{N}\sum_{i=1}^{N}(x_i^T \hat{u})^2$$

$$= \frac{1}{N}\sum^{N}\left(\hat{u}^T x_i x_i^T \hat{u}\right)$$

$$= \hat{u}^T \underbrace{\frac{1}{N}\left[\sum_{i=1}^{N} x_i x_i^T\right]}\hat{u}$$

$$= \hat{u}^T S \hat{u}$$

$$\max_{\hat{u}} \quad \hat{u}^T S \hat{u}$$

$$\text{s.t} \quad \hat{u}^T \hat{u} = 1$$

$$\hat{u}^T S \hat{u} - \lambda(\hat{u}^T \hat{u} - 1)$$

$$\frac{\partial}{\partial \hat{u}}\left[\hat{u}^T S \hat{u} - \lambda(\hat{u}^T \hat{u} - 1)\right]$$

$$2S\hat{u} - 2\lambda\hat{u} = 0$$

$$\boxed{\underset{(D\times D)}{S}\hat{u} = \lambda\hat{u}}$$

The solution for the above eqn. will be the eigen vector of S.

Let $\hat{u}$ be an eigen vector of S.

Variance of projected data $= \hat{u}' \underline{\underline{S \hat{u}}}$

$$= \hat{u}^T \lambda \hat{u}$$
$$= \lambda \hat{u}^T \hat{u}$$
$$= \lambda$$

Variance along any eigen vector of S will be equal to the corresponding eigen value.

<u>First eigen vector</u> gives the direction of maximum variance

$\rightarrow$ First principal component.

$$S \rightarrow \quad \begin{matrix} \boxed{\begin{matrix} \hat{u}_1 \\ \lambda_1 \end{matrix}} & \begin{matrix} \hat{u}_2 \\ \lambda_2 \end{matrix} & \begin{matrix} \hat{u}_3 \\ \lambda_3 \end{matrix} & \begin{matrix} \cdots \\ \cdots \end{matrix} & \begin{matrix} \hat{u}_D \\ \lambda_D \end{matrix} \end{matrix}$$

For top L PCs:

$$\frac{\sum_{i=1}^{L} \lambda_i}{\sum_{i=1}^{D} \lambda_D}$$

$W = [\hat{u}_1 \hat{u}_2 \cdots \hat{u}_L]$

$$W \rightarrow \begin{bmatrix} a_1 & a_2 & \cdots & a_L \\ & & & \\ & & & \end{bmatrix} : D \times L \text{ matrix}$$

$$\underbrace{\underset{(N \times L)}{Z} = \underset{(N \times D)}{X} \; \underset{(D \times L)}{W}}$$

$$\begin{array}{cc} & \begin{array}{cc} & 2 \end{array} \\ X_3 & \begin{bmatrix} 7 & 3 \\ 2 & 4 \\ 1 & 8 \end{bmatrix} \end{array}$$

$$\bar{x} \quad \begin{bmatrix} 10/3 & 5 \end{bmatrix}$$

mean centered $(X) = \begin{bmatrix} \frac{11}{3} & -2 \\ -\frac{4}{3} & -1 \\ -\frac{7}{3} & 3 \end{bmatrix}$

If I have z_i $(L \times 1)$

$$\underset{(D \times 1)}{\hat{x}_i} = W z_i$$