Principal Component Analysis (PCA)

X – data matrix $(N \times D)$

$x_1, x_2, x_3 \ldots$ data instances

$x_i \in \mathbb{R}^D$

any dimensionality reduction (DR) method.

$$\underset{(N \times D)}{X} \longrightarrow \underset{(N \times L)}{Z} \qquad \underline{L << D}$$

Linear DR

Weights:

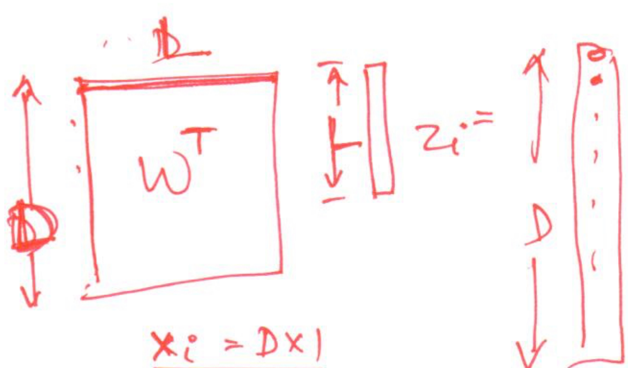
$$\boxed{X = Z\underset{(L \times D)}{W}}$$

~~learn~~ use W and Z to get X

DR $\rightarrow$ $\boxed{\text{Given } X \longrightarrow \text{learn } Z, W}$

$W \rightarrow$ loading matrix.

Consider z_i



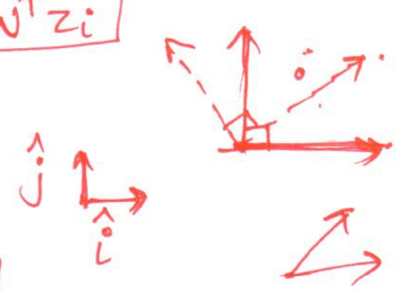
$x_i = D \times 1$

$$\boxed{x_i = W^T z_i}$$

Each column of W is a basis vector

$$\Rightarrow \boxed{w_j^T \cdot w_k = 0}$$

$$\boxed{w_j^T w_j = 1}$$



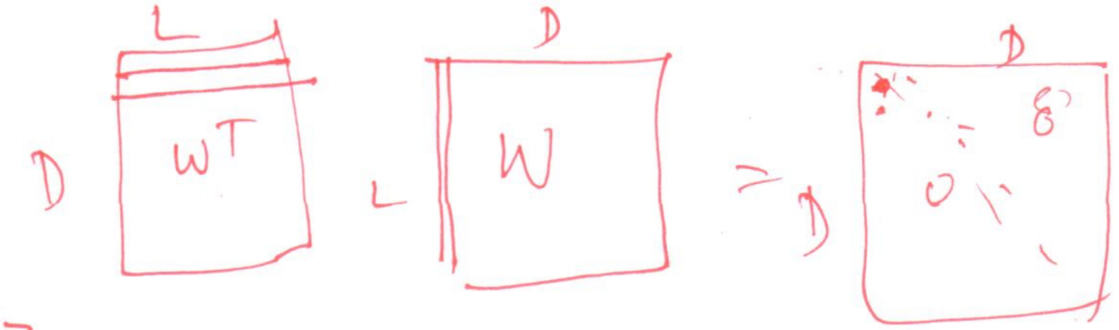
$W \rightarrow$ orthonormal bases.

$$\boxed{J(W, Z) = \frac{1}{N} \sum_{i=1}^{N} \| x_i - W^T z_i \|_2^2}$$

s.t. $W^T W = I$

reconstruction loss

growl, but do not do so often. We have a male and a female here.



find W and Z

Let us solve this for $L=1$ $\Rightarrow$ W is a $1 \times D$ vector w_1

Z is a $N \times 1$ vector. z_1 N

$$J(W,Z) = \frac{1}{N}\sum_{i=1}^{N}\left[(x_i - z_{i1} w_1)^T (x_i - z_{i1} w_1)\right]$$

scalar

$$= \frac{1}{N}\sum_{i=1}^{N}\left[x_i^T x_i - 2 z_{i1} w_1^T x_i + z_{i1}^2 \underline{w_1^T w_1}\right]$$

Recall: that W is orthonormal

$$= \frac{1}{N}\sum_{i=1}^{N}\left(\underline{x_i^T x_i} - 2 \underline{z_{i1}}\, w_1^T x_i + z_{i1}^2\right)$$

$$\frac{\partial J}{\partial z_{i1}} = \frac{1}{N} \cancel{\sum_{i=1}^{N}} \left(-2 w_1^T x_i + 2 z_i\right)$$

Setting this to 0 : $\cancel{\frac{1}{N}\sum_{i=1}^{N}}$

$$\boxed{z_{i1} = w_1^T x_i}$$

$$J(W) = \frac{1}{N}\sum_{i=1}^{N}\left[x_i^T x_i - 2 w_1^T x_i w_1^T x_i + (w_1^T x_i)^2\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left((x_i^T x_i) - z_i^2 \cancel{w_1^T x_i}\right)$$

$$= \cancel{\frac{1}{N}}\, \text{const} - \frac{1}{N}\sum_{i=1}^{N} z_i^2.$$

Let us assume that original data is centered

mean$(X) = 0$ $\quad \boxed{X = X - \mu}$ where μ is the col. mean

$$\mathrm{var}[z_{i1}] = E[z_{i1}^2] - \left(E[z_{i1}]\right)^2$$

$$= E[z_{i1}^2] - \left(E[w_1^T x_i]\right)^2$$

$$= E[z_{i1}^2] - \left(w_1^T E[x_i]\right)^2$$

mean of the data.

$E[x_i] = 0$

$$= E[z_{i1}^2]$$

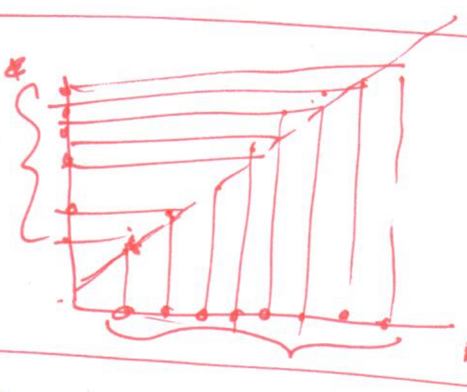
$$\boxed{\mathrm{var}[z_{i1}] = \frac{1}{N}\sum_{i=1}^{N} z_{i1}^2}$$

$$J(w) = \text{const} - \mathrm{var.}[z_{i1}]$$

Instead of maximizing $J(w)$, I can minimize $\mathrm{var}[z_{i1}]$

Principle of maximal variance

We want to identify the ~~direction~~ basis vector w_1 such that the variance of the projected data ($\cancel{z_1, z_2}$ z_{i1}) is maximized.



~~$z_1 = X w_1$~~ $\quad z_{i1} = w_1^T x_i = x_i^T w_1$

$$\mathrm{var}[z_{i1}] = \frac{1}{N}\sum_{i=1}^{N} (x_i^T w_1) * (x_i^T w_1) = \frac{1}{N}\sum (w_1^T x_i x_i^T w_1)$$

$$= w_1^T \underbrace{\left(\frac{1}{N}\left(\sum x_i x_i^T\right)\right)}\, w_1$$

$$X \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_N \end{pmatrix} \qquad \cancel{\Sigma} \rightarrow S = \frac{1}{N}\sum_{i=1}^{N}\left[(x_i-\mu)(x_i-\mu)^T\right]$$

Since $\mu = 0$ [Data is centered]

$$\boxed{S = \frac{1}{N}\sum_{i=1}^{N} x_i x_i^T}$$

$$\boxed{\mathrm{Var}[z_{i1}] = w_1^T S w_1}$$

We want to find w_1 that maximizes $w_1^T S w_1$

Lagrangian method:

$$\boxed{\begin{aligned} \max_{w_1} \; & J(w_1) = w_1^T S w_1 \\ & \text{s.t} \;\; w_1^T w_1 = 1 \end{aligned}}$$

$$\boxed{w_1^T S w_1 - \lambda(w_1^T w_1 - 1)}$$

λ Lagrange multiplier

$$\frac{\partial J}{\partial w_1} = 0 \quad 2 S w_1 - 2\lambda w_1 = 0$$

$$\frac{\partial J}{\partial \lambda} = 0 \quad w_1^T w_1 = 1$$

$$\boxed{S w_1 = \lambda w_1}$$

S: $D \times D$, w_1: $D \times 1$, λ: scalar, w_1: $D \times 1$

A solution to $Sw_1 = \lambda w_1$ $\boxed{Ax = \lambda x}$

would be an eigen-vector of S.

The first eigen-vector will be. w_1

first eigen-value will be the variance along w_1

$w_1^T S w_1 \rightarrow \lambda w_1$

$w_1^T (\lambda w_1)$

$\lambda w_1^T w_1 = \lambda$

The second eigen-vector will be w_2

X

(1) Center X $X = X - \mu$

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

(3) Calculate eigen vectors of S.

(Choose top L eigen vectors) $\rightarrow W$

$\boxed{Z = XW}$

$$X = USV^T$$
$N \times D$

S is a diagonal matrix

U, V are orthogonal matrices.

$$X \approx U_L S_L V_L^T$$

$(N \times D)$

first L left s.v

first L right s.v

first L singular values

$$\boxed{(N \times L) + (D \times L) + L}$$

Allows for compression of X

$\boxed{10^9 \times 10^6}$

$L = 2$

$$(10^9 \times 2 + 10^6 \times 2 + 2)$$

$$\| X - U_L S_L V_L^T \|_F^2$$

Frobenius norm

$$\|A\|_F = \sqrt{\sum_i \sum_j A_{ij}^2}$$