Dimensionality Reduction:

→ Computational efficiency.

age	sal	dob	...	Approve Y/N

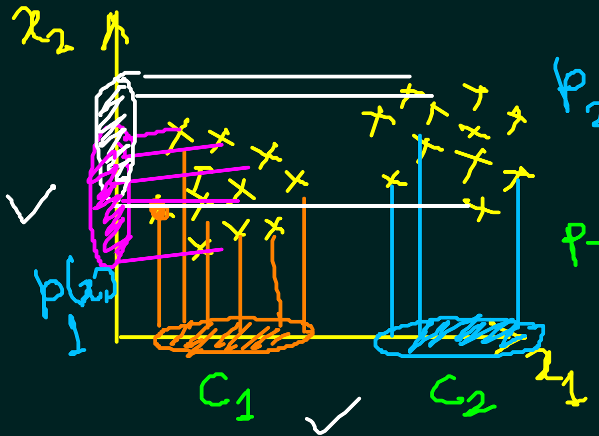
$\langle a_1, a_2, \dots, a_d \rangle$

Reduce

$\langle a_{k_1}, a_{k_2}, \dots, a_{k_m} \rangle$

$1 \le k_i \le d$

$1 \le m < d$



Projection

→ Evaluate: (Class Separability)

Kullback-Liebler Divergence (KL-Div)

$$KLD(p_1, p_2) = \sum_{x \in T} p_1(x) \log \frac{p_1(x)}{p_2(x)} + \sum_{x} p_2(x) \log \frac{p_2(x)}{p_1(x)}$$

(log-likelihood)

Projection

{Attribute Subset Selection}

$\boxed{D}$ attributes

$\boxed{d} < D$ select } ${}^{D}C_d \approx O(D^d)$

$\boxed{D}$

(d) ← 2^D (Search)

→ Feature $\boxed{\text{Selection}}$

① Forward Search (seq)

Backward Search (seq)

① $x_1, \dots, x_D$

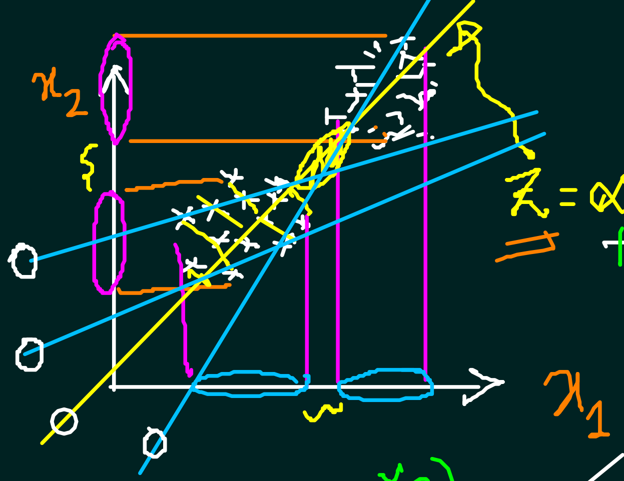
one $\{x_4\}$ ⇝ $\{x_4, x_1\}$, $\{x_4, x_3\}$ ⇝ $\{x_4, x_2, \dots\}$

$\{x_4, x_2\}$

$\vdots$ $\{x_4, x_D\}$

?

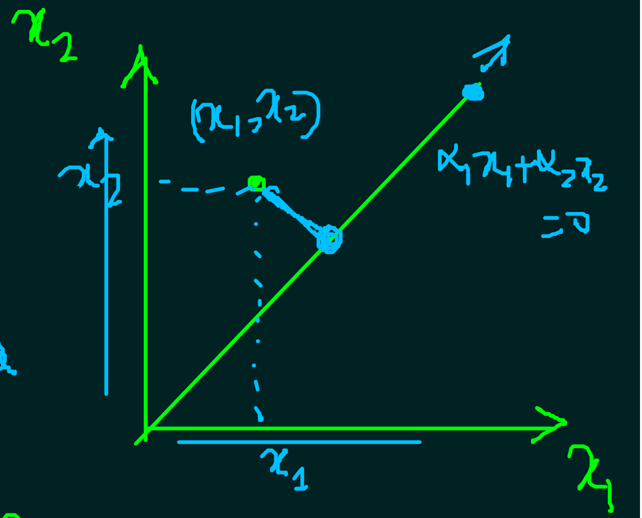
$O(dD)$ ← $D + D-1 + D-2 + \dots + (D-d)$

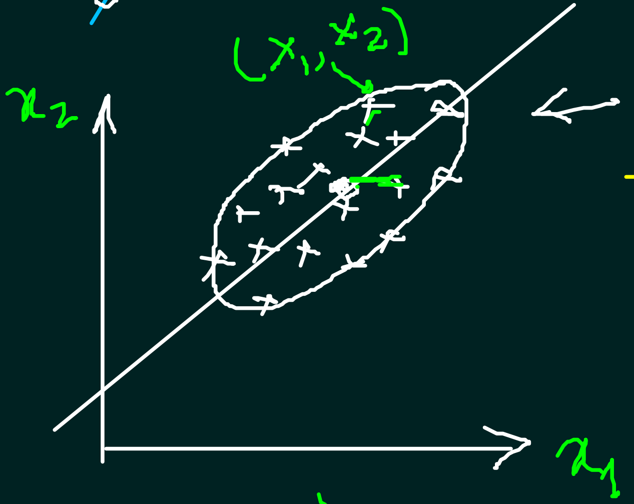


Feature Extraction:

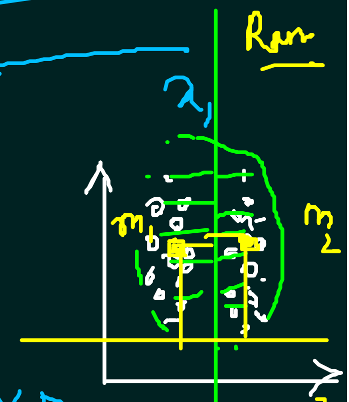
$Z = \alpha_1 x_1 + \alpha_2 x_2 \rightarrow$ 1-D

$Z = \alpha_1 x_1 + \alpha_2 x_2 + \cdots + x_d$





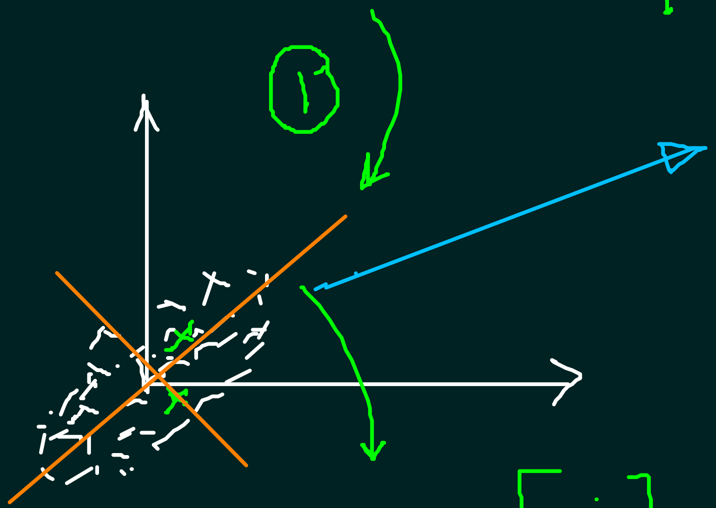
Principal Axis $\rightarrow$ $\boxed{\text{PCA}}$??



1. $x_1' = x_1 - \bar{x}_1$ ← mean of all TE

   $x_2' = x_2 - \bar{x}_2$

   - - . . . .



2. Covar Matrix $C = \begin{bmatrix} \quad \end{bmatrix}$ $D \times D$

$J(s) = \dfrac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$

3. $Cx = \lambda x$ ← all λ (eigenvalues)

highest λ

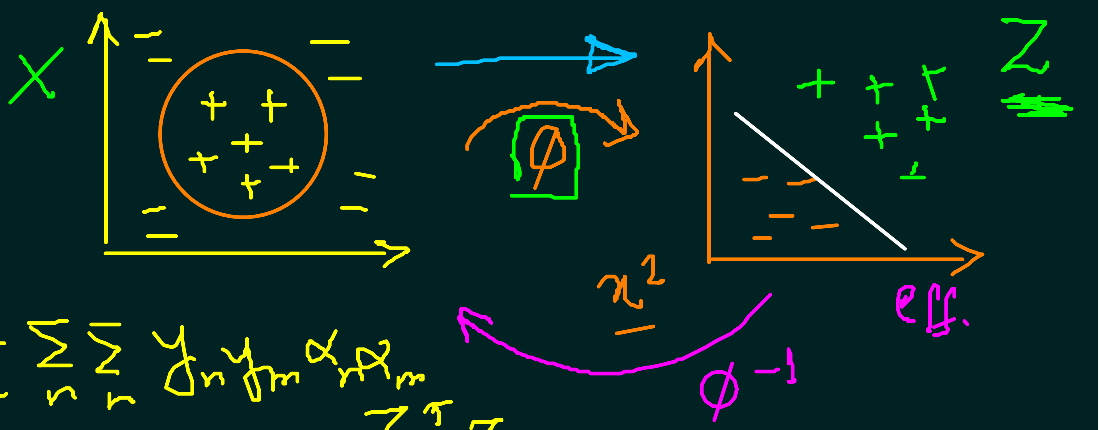
$EV = \begin{bmatrix} \vdots \end{bmatrix}_{d \times 1}$

$\boxed{d} \rightarrow$ highest d λ's $\rightarrow$ EVs

Best class sep.

Kernel Methods

$$\underset{\langle x_1, \dots x_d \rangle}{x} \rightarrow \phi \rightarrow \underset{\langle x_{k_1}, \dots, x_{k_m} \rangle}{Z}$$



Kernel Trick:

$$\mathcal{L}(\alpha) = \sum_n \alpha_n - \frac{1}{2} \sum_n \sum_m y_n y_m \alpha_n \alpha_m \, Z_n^T Z_m \rightarrow \text{Hessian}$$

$[x_i, x_j]$ linear

$$Z = \phi(x) = (1, x_1, x_2, x_1^2, x_2^2, x_1 x_2) \qquad H =$$

$$\boxed{Z^T Z'} = \underline{K(x, x')} = 1 + x_1 x_1' + x_2 x_2' + x_1^2 x_1'^2 + x_2^2 x_2'^2 + x_1 x_1' x_2 x_2' \qquad x = (x_1, x_2)$$

Polynomial kernel

$$\underline{K(x, x')} = (1 + x^T x')^2 = (1 + x_1 x_1' + x_2 x_2')^2$$

$$= 1 + x_1^2 x_1'^2 + x_2^2 x_2'^2 + 2 x_1 x_1' + 2 x_2 x_2' + 2 x_1 x_1' x_2 x_2'$$

$$(1, x_1^2, x_2^2, \sqrt{2} x_1, \sqrt{2} x_2, \sqrt{2} x_1 x_2) \leftarrow \phi(x)$$

$$(1, x_1'^2, x_2'^2, \sqrt{2} x_1', \sqrt{2} x_2', \sqrt{2} x_1' x_2') \leftarrow \phi(x')$$

$(1 + x^T x')^Q$

$$\underbrace{K(x,x')} = e^{\underbrace{-\gamma\|x-x'\|^2}} = \underbrace{e^{-x^2}}\,\underbrace{e^{-x'^2}}\,\underbrace{\sum_{k=0}^{\infty}\frac{2^k\,x^k\,x'^k}{k!}}_{e^{2xx'}}$$

$$= \left[e^{-x^2}\sum_{k=0}^{\infty}\frac{(\sqrt{2})^k x^k}{\sqrt{k!}}\right]\left[e^{-x'^2}\sum_{k=0}^{\infty}\frac{(\sqrt{2})^k x'^k}{\sqrt{k!}}\right]$$

$\phi\ (\ -\ -\ -\ -\ -\ -\ -\ )\quad (\ .\ -\ -\ -\ -\ -\ -\ )$

$\infty \quad \phi \rightarrow Z$

$$\mathcal{H} = \begin{bmatrix} y_1 y_1\, x_1 \cdot x_1 & \cdots & y_1 y_k\, x_1 \cdot x_k \\ y_2 y_1\, x_2 \cdot x_1 & & y_2 y_k\, x_2 \cdot x_k \\ x_3 y_1\, x_3 \cdot x_1 & & \vdots \end{bmatrix}$$

$$\mathcal{H} = \begin{bmatrix} y_1 y_1\, K(x_1, x_1) & \cdots \\ y_2 y_1\, K(x_2, x_1) & \cdots \end{bmatrix}$$

$$E\left[E_{out}\right] \leq \frac{E[\#SV]}{N-1} \quad ?$$

$K[\ \ ]$ ✓ Symmetric
✓ Semi-definite

Mercer's cond.

w^* ★

[RADIAL BASIS FUNCTION] → Next class