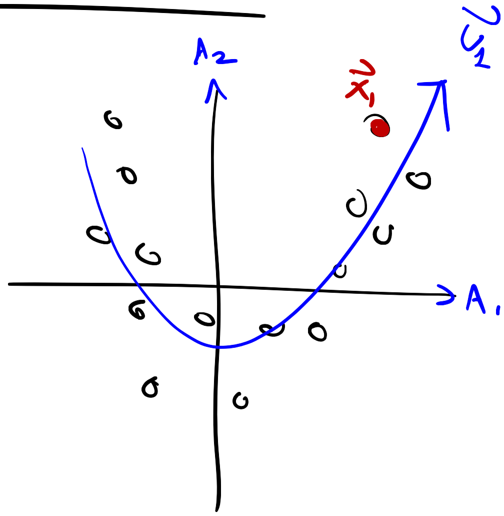
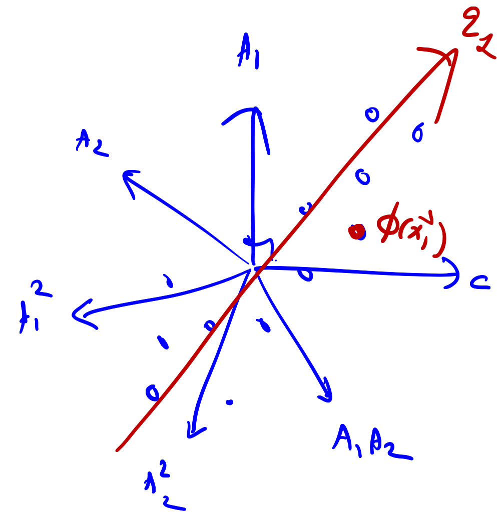
Kernel PCA



Input space

$\xRightarrow[\text{transformation}]{\phi}$

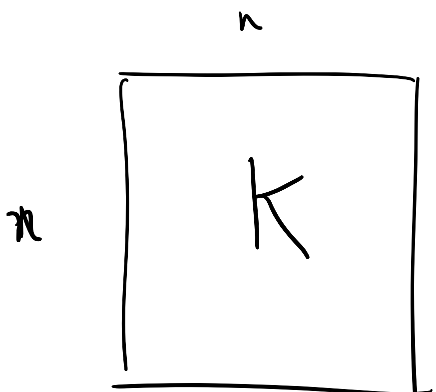


feature space

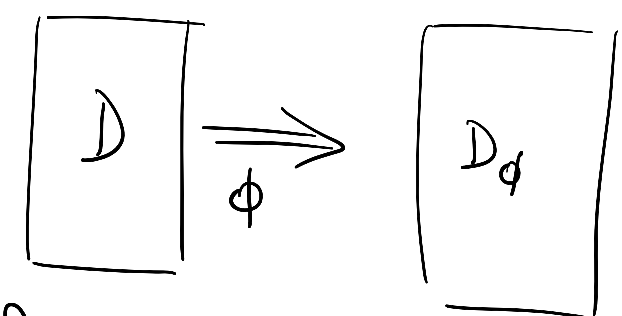
AVOID

$\Downarrow$

$k(\vec{x}_i, \vec{x}_j)$, symmetric, PSD : pos semi definite kernel



Kernel matrix



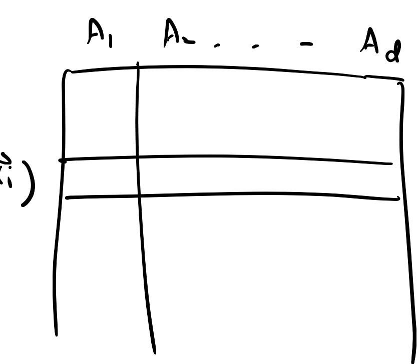
Input Space

Feature Space

$$k(\vec{x}_i, \vec{x}_j) = \boxed{\phi(\vec{x}_i)^T \phi(\vec{x}_j)}$$

$\Downarrow$ Center

D_ϕ



$$\bar{k}(\vec{x}_i, \vec{x}_j) = \underbrace{\overline{\phi(x_i)}^T \overline{\phi(x_j)}}_{\text{Conceptual}}$$

$\bar{K}$ c/o eigen decomposition

$$\begin{array}{cccc} \alpha_1 & \alpha_2 & \cdots & \alpha_r \\ \downarrow & \downarrow & \cdots & \downarrow \\ \vec{c}_1 & \vec{c}_2 & & \vec{c}_r \end{array}$$

1) Center D_ϕ

$\hookrightarrow \mu_\phi = \frac{1}{n}\sum_{i=1}^{n} \phi(\vec{x}_i)$

$\forall i, \quad \boxed{\overline{\phi}(x_i) = \phi(\vec{x}_i) - \vec{\mu}_\phi}$

assume that each $\phi(\vec{x}_i)$ has been Centered

$\vec{c}_1$ is NOT the direction

$$\vec{c}_1 = \begin{pmatrix} c_{11} \\ c_{12} \\ c_{13} \\ \vdots \\ c_{1n} \end{pmatrix}$$

$\vec{u}_1 = \sum c_{1i}\, \phi(x_i)$

We need to plug $\vec{c}_1$ into eq above to determine $\vec{u}_1$, 1st PC

2) $\Sigma_\phi \equiv$ Cov matrix

$$= \frac{1}{n}\sum_{i=1}^{n} \underbrace{\overline{\phi}(\vec{x}_i)\, \overline{\phi}(\vec{x}_i)^T}_{\text{outer product}}$$

3) $\boxed{\Sigma_\phi \vec{u}_1 = \lambda_1 \vec{u}_1}$ solve ?

$$\left(\frac{1}{n}\sum_{i=1}^{n} \overline{\phi}(\vec{x}_i)\, \overline{\phi}(\vec{x}_i)^T\right) \vec{u}_1 = \lambda_1 \vec{u}_1$$

$$\sum_{i=1}^{n}\left[\overline{\phi}(\vec{x}_i)\underbrace{\left(\frac{\overline{\phi}(x_i)^T \vec{u}_1}{n\lambda_1}\right)}_{c_i}\right] = \vec{u}_1$$

$$\boxed{\vec{u}_1 = \sum_{i=1}^{n} c_i\, \overline{\phi}(\vec{x}_i)}$$

Representer Theorem

$$\vec{u}_1 = c_1 \cdot \overline{\phi(\vec{x}_1)} + c_2 \overline{\phi(x_2)} + c_3 \overline{\phi(x_3)} + \cdots + c_n \overline{\phi(\vec{x}_n)}$$

$$\vec{c} = (c_1\ c_2\ \cdots\ c_n)^T \in \mathbb{R}^n$$

weight vector, mixing coefficients, etc

$$(\Sigma_\phi)\, \vec{u}_1 = \lambda_1 \vec{u}_1$$

$$\frac{1}{n}\left(\sum_{i=1}^{n} \overline{\phi(x_i)}\, \overline{\phi(x_i)}^T\right)\left(\sum_{j=1}^{n} c_j \overline{\phi(x_j)}\right) = \lambda_1 \sum_{i=1}^{n} c_i \overline{\phi(\vec{x}_i)}$$

$\bar{K}$: Centered Kernel

$$\underbrace{\frac{1}{n} \sum_{i=1}^{n}\sum_{j=1}^{n} c_j \overline{\phi(x_i)}\, \underbrace{\overline{\phi(x_i)}^T \overline{\phi(x_j)}}_{\bar{k}(\vec{x}_i, \vec{x}_j)}}_{} = \lambda_1 \sum_{i=1}^{n} c_i \overline{\phi(\vec{x}_i)}$$

$$\forall_k : \quad \overline{\phi(x_k)}^T \Big(\qquad \Big) = \lambda_1 \overline{\phi(x_k)}^T \left(\sum \overline{\phi(x_i)}\, c_i\right)$$

n different equations one per point $\phi(x_k)$

$$\frac{1}{n} \sum_i \sum_j c_j\, k(x_k, x_i) \cdot k(x_i, x_j) = \lambda_1 \sum c_i k(x_k, x_i)$$

$$\boxed{\bar{K}^2 \vec{c} = n\lambda_1 \bar{K} \vec{c}}$$

$$\boxed{\bar{K} \vec{c} = n\lambda_1 \vec{c}}$$

$$\boxed{\bar{K}\vec{c}_1 = \underline{\alpha_1}\,\vec{c}_1}$$

The weight vector $\vec{c}_1$ is the dominant eigenvector of $\bar{K}$, centered kernel matrix

$\bar{K}$ is PSD matrix, symmetric

Scalars $\rightarrow$ $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_n \geq 0$

$$\underbrace{\vec{c}_1 \quad \vec{c}_2 \quad \dots \quad \vec{c}_n}_{\mathbb{R}^n}$$

Q1: what is $\vec{u}_1$, 1^{st} PC

$$\vec{u}_1 = \sum_{i=1}^{n} c_{1i}\, \overline{\phi(x_i)}$$

scalar ?

We cannot compute $\vec{u}_1$!

We can always project scalar onto $\vec{u}_1$

take each point $\phi(x_k)$ $\forall\, k = 1 \dots n$ & project onto $\vec{u}_1$

$$\overline{\phi(x_k)}^T \underline{\vec{u}_1} = \sum_{i=1}^{n} c_{1i}\, \underbrace{\overline{\phi(x_k)}^T\, \overline{\phi(x_i)}}$$

$$a_k^1 = \sum_{i=1}^{n} \underline{c_{1i}}\, \bar{K}\underline{(\vec{x}_k, \vec{x}_i)} = \vec{c}_1^{\,T} \bar{K}_k$$

$\underline{\underline{a_k^1}}$ is the scalar projection of $\overline{\phi(x_k)}$ onto $\vec{u}_1$

k-th row of $\overline{K}$ $\longrightarrow$ $\vec{K}_k$

$$\begin{bmatrix} & & \\ k(x_k,x_1) & k(x_k,x_2) & \cdots \\ & & \end{bmatrix}$$

Q2

What's the variance along $\vec{u}_1$?

$$\boxed{\overline{K}\vec{c}_1 = \alpha_1 \vec{c}_1}$$

λ_1 is the variance

$$\alpha_1 = n\lambda_1$$

$$\Rightarrow \lambda_1 = \alpha_1/n$$

Q3

$\vec{c}_1$ is a unit vector, but $\vec{u}_1$ may not be!

$$\vec{u}_1 = \sum_i c_{1i}\,\overline{\phi(x_i)}$$

we have to scale $\vec{u}_1$ to make it unit length

$$\vec{u}_1^T \vec{u}_1 = 1$$

$$\left(\sum_i c_{1i}\,\overline{\phi(x_i)}\right)^T \left(\sum_j c_{1j}\,\overline{\phi(x_j)}\right) = 1$$

$$\underbrace{\sum_i \sum_j c_{1i}\, \underbrace{\overline{\phi(x_i)}^T\overline{\phi(x_j)}}_{k(x_i,x_j)}\, c_{1j}} = 1$$

$$\boxed{\vec{c}_1^{\,T} \underbrace{\bar{K}\,\vec{c}_1} = 1}$$

$$\vec{c}_1^{\,T}(\alpha_1 \vec{c}_1) = 1$$

$$\vec{c}_1^{\,T} c_1 = 1/\alpha_1$$

$$\|\vec{c}_1\|^2 = 1/\alpha_1$$

Simply multiply c_1 by $1/\sqrt{\alpha_1}$ to obtain a unit $\vec{u}_1$ vector

---

Kernel PCA (D)

1) Compute K

2) Compute $\bar{K}$ (in terms of K)

3) Solve $\boxed{\bar{K}\vec{c} = \alpha \vec{c}}$ eigen decomposition

$$\begin{array}{cccc} \alpha_1 & \alpha_2 & & \alpha_n \\ \downarrow & \downarrow & \cdots & \downarrow \\ \vec{c}_1 & \vec{c}_2 & & \vec{c}_n \end{array}$$

4) Compute variance:

$\underline{\lambda_1} = \alpha_1/n \quad \underline{\lambda_2} = \alpha_2/n \quad \ldots$

5) determine suitable lower dimensionality r based on $\theta = 0.975$

6) Scale $\vec{c}_i$

$\vec{c}_1 = \vec{c}_1/\sqrt{\alpha_1}, \quad \vec{c}_2 = \vec{c}_2/\sqrt{\alpha_2}, \quad \ldots$

$\underbrace{\qquad\qquad}_{\text{weights for } \vec{u}_1}$ $\underbrace{\qquad\qquad}_{\text{weights for } \vec{u}_2}$

7) project each point onto $\vec{u}_1, \vec{u}_2, \ldots \vec{u}_r$

$\forall j$, use j-th row of K, call it $\vec{K}_j$

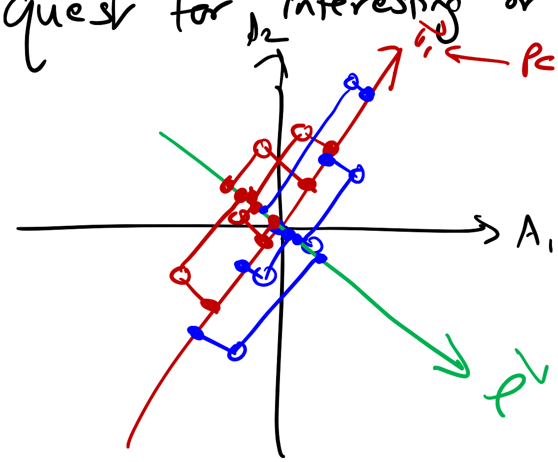
$$a_{j1} = \vec{c}_1^{\,T} \vec{K}_j, \quad a_{j2} = \vec{c}_2^{\,T} \vec{K}_j, \quad \ldots$$

$$\forall j = 1 \ldots n \quad \vec{a}_j = (a_{j1}, a_{j2}, \ldots a_{jr})^T \in \mathbb{R}^r$$

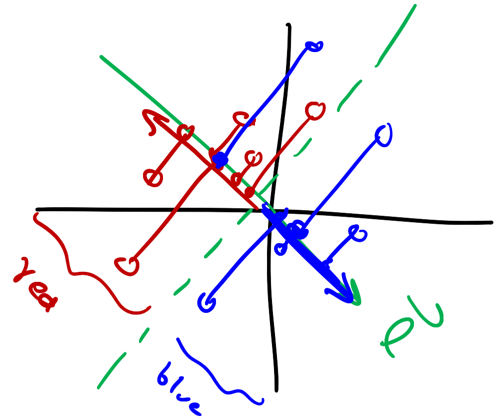
Chapter 20: Linear Discriminant Analysis (LDA)

1^{st} LD : 1^{st} linear discriminant

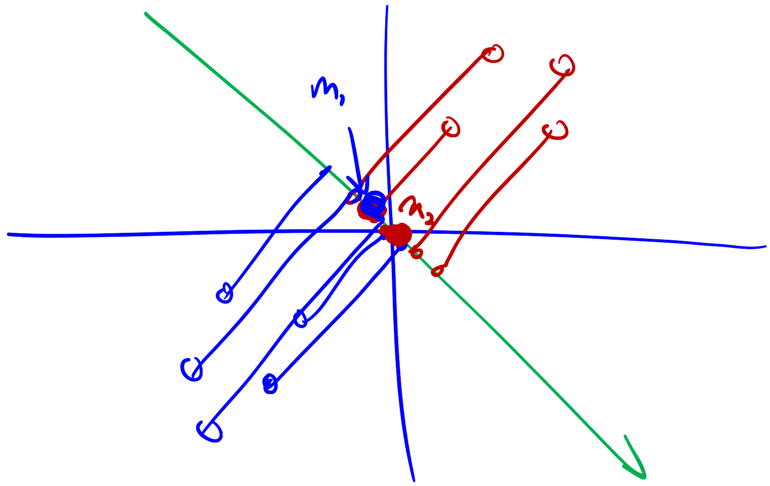
quest for interesting or useful directions

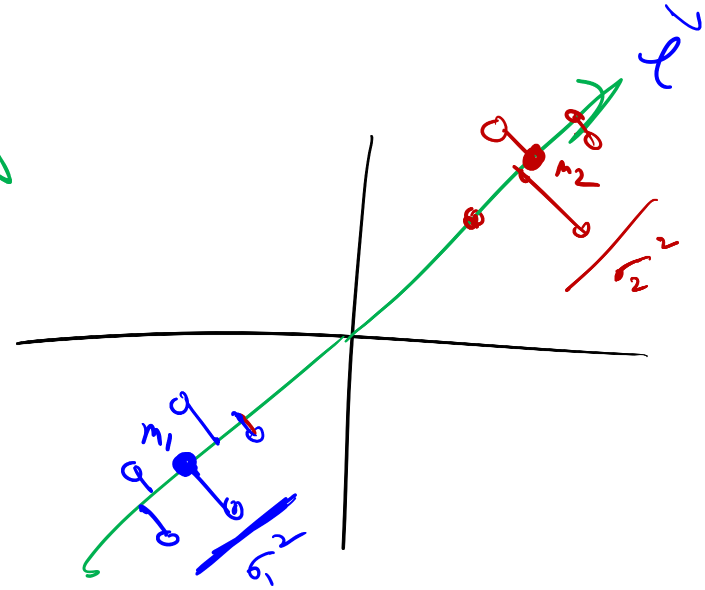


find $\vec{\ell}$, that 'separates' the red vs. blue points



1^{st} PC is bad in separating

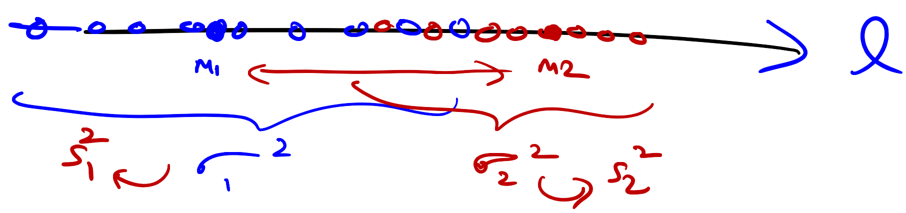




a) distance between projected scalar means m_1 & m_2

$$|m_1 - m_2|$$

$$(m_1 - m_2)^2 \leftarrow \text{maximized}$$



b) minimize sum of the variance

S_1 = scatter for blue

S_2 = " " red

n_i = # points in group i
class i

$$S_1^2 = n_1 \sigma_1^2$$

$$S_2^2 = n_2 \sigma_2^2$$

$$J = \text{maximize} \left[\frac{(m_1 - m_2)^2}{s_1^2 + s_2^2} \right]$$