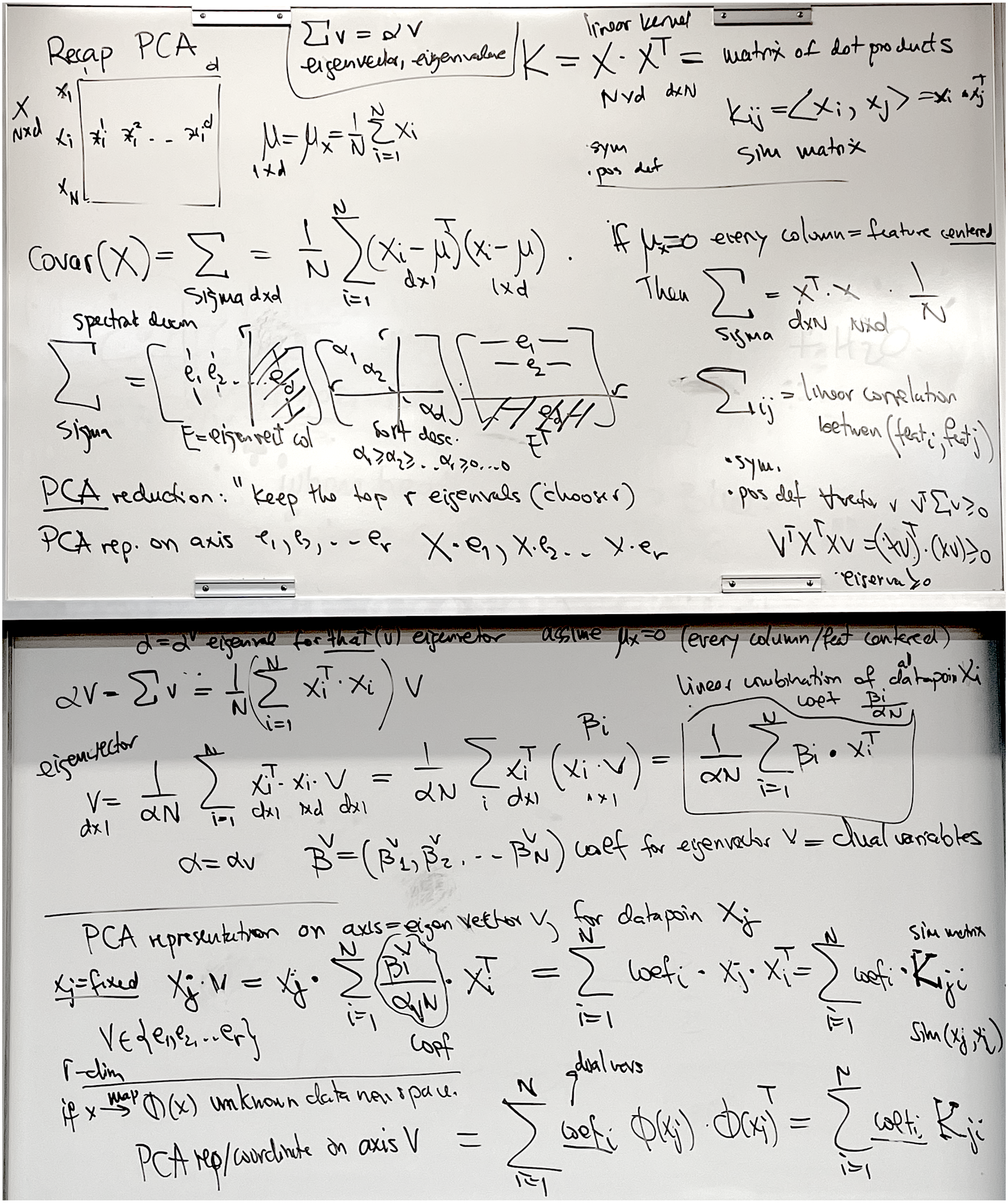


Recap PCA

$$\underset{N\times d}{X}\quad \begin{array}{c|ccc|} & 1 & & d \\ x_1 & & & \\ x_i & x_i^1 & x_i^2 \ldots & x_i^d \\ x_N & & & \end{array}$$

$$\boxed{\Sigma v = \alpha v}$$
eigenvector, eigenvalue

linear kernel

$$K = \underset{N\times d}{X} \cdot \underset{d\times N}{X^T} = \text{matrix of dot products}$$

$$K_{ij} = \langle x_i, x_j \rangle = x_i \cdot x_j^T$$

- sym
- pos def

sim matrix

$$\underset{1\times d}{\mu} = \mu_x = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\text{Covar}(X) = \underset{\text{Sigma}\ d\times d}{\Sigma} = \frac{1}{N}\sum_{i=1}^{N} \underset{d\times 1}{(x_i-\mu)^T}\underset{1\times d}{(x_i-\mu)}\ .$$

If $\mu_x = 0$ every column = feature centered

Then $\underset{\text{Sigma}}{\Sigma} = \underset{d\times N}{X^T} \cdot \underset{N\times d}{X} \cdot \frac{1}{N}$

spectral decom

$$\underset{\text{Sigma}}{\Sigma} = \underset{E = \text{eigenvect col}}{\left[\, e_1\ e_2 \ldots \overset{r}{|}\ \cancel{e_d} \,\right]} \underset{\text{sort desc.}}{\begin{bmatrix} \alpha_1 & & \overset{r}{|} & \\ & \alpha_2 & | & \\ \hline & & & \\ & & & \alpha_d \end{bmatrix}} \underset{E^T}{\begin{bmatrix} -e_1- \\ -e_2- \\ \hline \cancel{e_d} \end{bmatrix}} r$$

$\alpha_1 \geq \alpha_2 \geq \ldots \alpha_r \geq 0 \ldots 0$

Σ_{ij} = linear correlation between $(\text{feat}_i, \text{feat}_j)$

- sym.
- pos def ∀vector v $v^T \Sigma v \geq 0$

$$v^T X^T X v = (Xv)^T \cdot (Xv) \geq 0$$

- eigenval ≥ 0

<u>PCA</u> reduction: "keep the top r eigenvals (choose r)

PCA rep. on axis $e_1, e_2, \ldots e_r$ $X\cdot e_1, X\cdot e_2 \ldots X\cdot e_r$

$\alpha = \alpha^v$ eigenval for <u>that</u> (v) eigenvector   assume $\mu_x = 0$ (every column/feat centered)

$$\alpha v = \Sigma v = \frac{1}{N}\left(\sum_{i=1}^{N} x_i^T \cdot x_i\right) v$$

linear combination of all datapoin x_i
coef $\frac{\beta_i}{\alpha N}$

eigenvector

$$\underset{d\times 1}{v} = \frac{1}{\alpha N}\sum_{i=1}^{N} \underset{d\times 1}{x_i^T} \cdot \underset{1\times d}{x_i} \cdot \underset{d\times 1}{v} = \frac{1}{\alpha N}\sum_{i} \underset{d\times 1}{x_i^T} \underset{1\times 1}{\overset{\beta_i}{(x_i \cdot v)}} = \boxed{\frac{1}{\alpha N}\sum_{i=1}^{N} \beta_i \cdot x_i^T}$$

$\alpha = \alpha_v$ $\beta^v = (\beta_1^v, \beta_2^v, \ldots \beta_N^v)$ coef for eigenvector v = dual variables

---

PCA representation on axis = eigen vector v, for datapoin x_j

$$\underline{x_j = \text{fixed}}\quad x_j \cdot v = x_j \cdot \sum_{i=1}^{N} \underbrace{\frac{\beta_i^v}{\alpha_v N}}_{\text{coef}} \cdot x_i^T = \sum_{i=1}^{N} \text{coef}_i \cdot x_j \cdot x_i^T = \sum_{i=1}^{N} \text{coef}_i \cdot \overset{\text{sim matrix}}{\underset{\text{sim}(x_j, x_i)}{K_{ji}}}$$

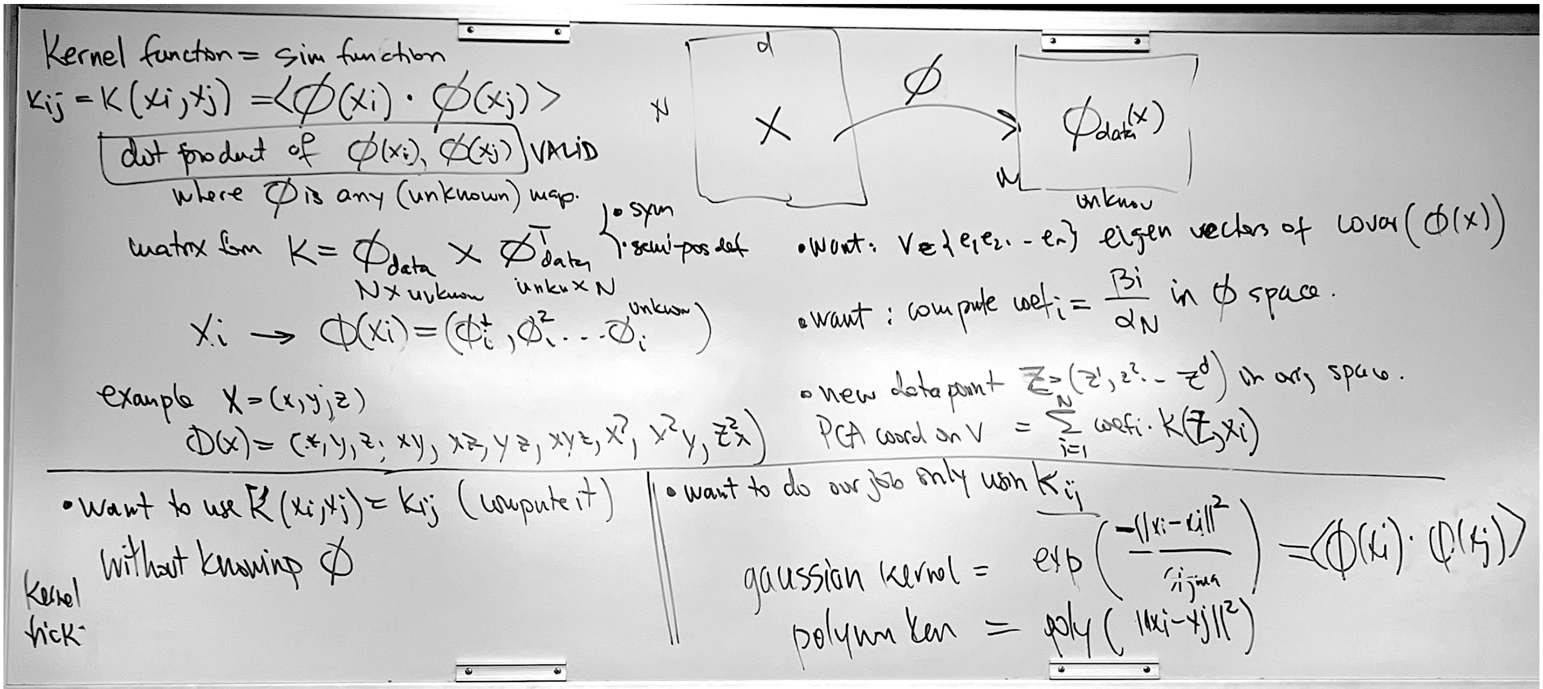
$v \in \{e_1, e_2, \ldots e_r\}$

r-dim

---

if $x \xrightarrow{\text{map}} \phi(x)$ unknown data new space.

$$\text{PCA rep/coordinate on axis } v = \sum_{i=1}^{N} \overset{\text{dual vars}}{\underline{\text{coef}_i}}\ \phi(x_j) \cdot \phi(x_i)^T = \sum_{i=1}^{N} \underline{\text{coef}_i}\ K_{ji}$$



Kernel function = sim function

$k_{ij} = K(x_i, x_j) = \langle \phi(x_i) \cdot \phi(x_j) \rangle$

$\boxed{\text{dot product of } \phi(x_i), \phi(x_j)}$ VALID

where ϕ is any (unknown) map.

matrix form $K = \Phi_{data} \times \phi^T_{data}$ $\left\{ \begin{array}{l} \bullet \text{ sym} \\ \bullet \text{ semi-pos def} \end{array} \right.$

N × unknow   unkn × N

$x_i \rightarrow \Phi(x_i) = (\phi_i^1, \phi_i^2 \ldots \phi_i^{unkown})$

example $X = (x, y, z)$

$\Phi(x) = (x, y, z; xy, xz, yz, xyz, x^2, x^2y, z^2x)$

- want: $V = \{e_1 e_2 .. e_n\}$ eigen vectors of covar$(\Phi(x))$
- want: compute $coef_i = \frac{\beta_i}{\alpha_N}$ in ϕ space.
- new datapoint $Z = (z^1, z^2 .. z^d)$ in orig space.

PCA coord on V $= \sum_{i=1}^{N} coef_i \cdot K(Z, x_i)$

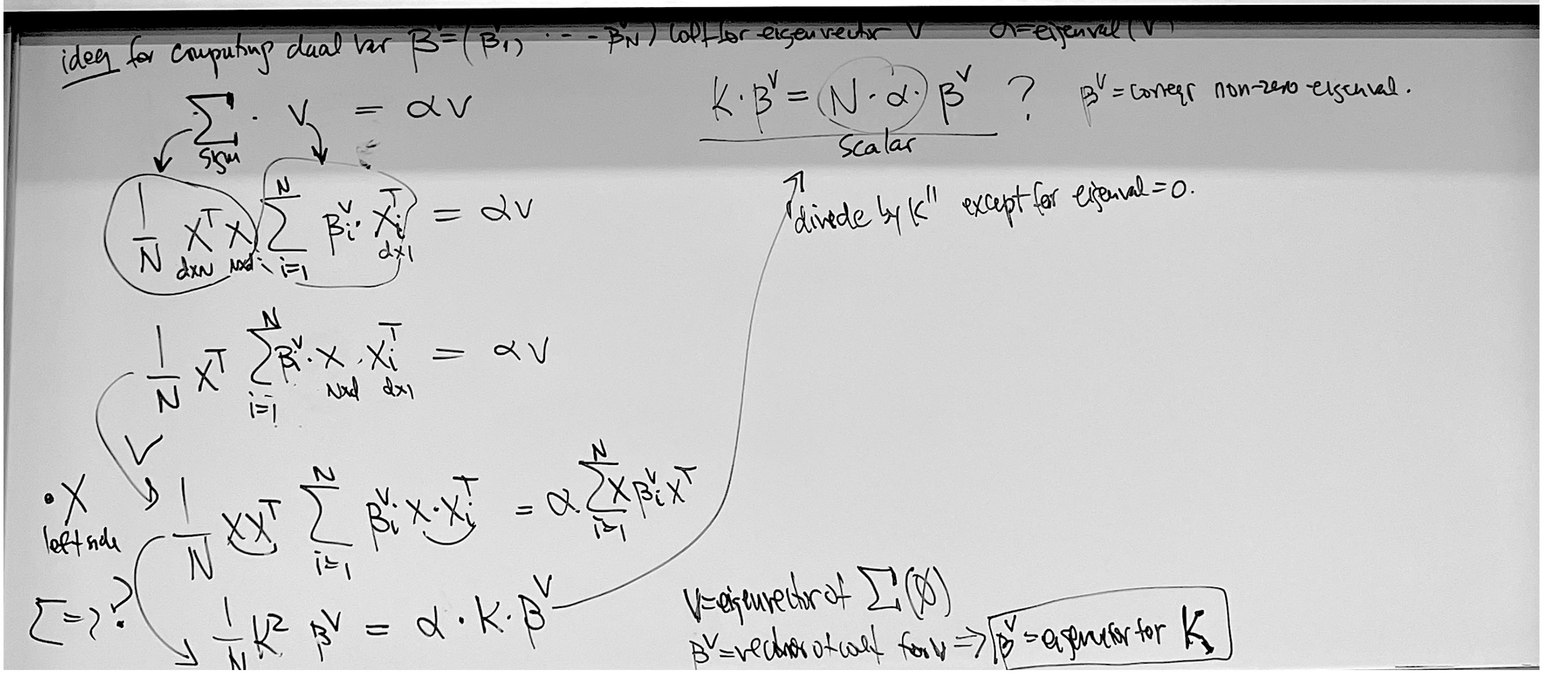
- want to use $K(x_i, x_j) = k_{ij}$ (compute it) without knowing ϕ

Kernel trick.

- want to do our job only usin $\underline{K_{ij}}$

gaussian kernel $= \exp\left(\frac{-\|x_i - x_j\|^2}{\text{sigma}}\right) = \langle \Phi(x_i) \cdot \Phi(x_j) \rangle$

polynm ken $= \text{poly}(\|x_i - x_j\|^2)$



idea for computing dual var $\beta = (\beta_1, \ldots \beta_N)$ coef for eigenvector V α = eigenval(V)

$$\underset{\text{sigm}}{\Sigma} \cdot V = \alpha V$$

$$\frac{1}{N} \underset{d \times N}{X^T} \underset{N \times d}{X} \sum_{i=1}^{N} \beta_i^V \cdot \underset{d \times 1}{X_i^T} = \alpha V$$

$$\frac{1}{N} X^T \sum_{i=1}^{N} \beta_i^V \cdot \underset{N \times d}{X} \cdot \underset{d \times 1}{X_i^T} = \alpha V$$

• X left side

$$\frac{1}{N} X X^T \sum_{i=1}^{N} \beta_i^V X \cdot X_i^T = \alpha \sum_{i=1}^{N} X \beta_i^V X^T$$

$\Sigma = ?$

$$\frac{1}{N} K^2 \beta^V = \alpha \cdot K \cdot \beta^V$$

$\underbrace{K \cdot \beta^V = N \cdot \alpha \cdot \beta^V}_{\text{scalar}}$? β^V = correspr non-zero eigenval.

"divide by K" except for eigenval = 0.

V = eigenvector of $\Sigma(\phi)$

β^V = vector of coef for V ⇒ $\boxed{\beta^V = \text{eigenvector for } K}$