$K = X \cdot X^T$ linear kernel

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

linear sim
linear corelation.

- primary

$h(x) = \boxed{X} \cdot w \ldots$

primary form (data)

- dual form has input $sim(x_i, x_j) = K_{ij}$, not the data (x)

$h(x) = \sum \ldots \underbrace{\langle x_j \cdot x_i \rangle}_{sim(x_i, x_j) = K_{ij}}$

---

example x_i = patient in hospital.

$h(x)$ = function (patient data
blood pressure,
age,
symptoms)

dual:
$h(x)$ = fuction of similarity of (x, other patients)

Kernel Trick

- $h(x) \Rightarrow$ dual form $\quad h(x) = \text{function}(K_{ij})$

  $\langle x_i \cdot x_j \rangle = \text{sim}(x_i, x_j)$

- Th h works with any <u>valid</u> kernel (not just linear kernel)

- plug in a domain-specific/sim-designed kernel (valid)

  <u>ex</u> gaussian kernel $K_{ij} = e^{\|x_i - x_j\|^2 / \beta}$ (prove valid)

- use the classifier $h(\,)$ with non-linear kernel.

- effect: $h(x)$ linear classifier (ex. regression, SVM) $\Rightarrow$ dual $h(x) = \text{function}\left(\begin{matrix}\text{non-lin} \\ K\end{matrix}\right)$ can do non-linear separation!

examples

- Lin Regression (L_2) : $(X^TX + \lambda I_D)\, w = X^T y$

  α = dual var $= [X^TX + \lambda I_N]^{-1} y$

  $\Rightarrow w = X^T\alpha$ = lin combination of datapoints

  check: $(X^TX + \lambda I_D) \cdot X^T\alpha = X^TXX^T\alpha + \lambda X^T\alpha = X^T(XX^T + \lambda I_N) \cdot \underbrace{(X^TX + \lambda I_N)^{-1} \cdot y}_{\alpha} = X^Ty$

  $h(x_j) = x_j w = \boxed{x_j \cdot X^T} \alpha = \boxed{K_{ij}}\, \alpha$ dual form.

- PCA coordinates are $X \cdot e$ $\quad e$ = eigenvector $\left(\overset{covar}{\Sigma}\right)$

  derivation... $e = X^T \cdot \beta^e$ = lin combination of datapoints $\left(\underset{\beta}{\text{coef}}\right)$

  dual form PCA: $X \cdot [e_1\ e_2\ \dots\ e_D] = \underbrace{\boxed{X \cdot X^T}}_{K} \begin{bmatrix} \beta^{e_1} & \beta^{e_2} & \dots & \beta^{e_D} \end{bmatrix}$

  $= K \cdot \boxed{\beta}$ matrix