CS4x80 lec23 kernel cont'd

$x \to \phi(x)$

$\vec{w} = \Sigma \alpha_i \phi(\vec{x}_i)$

Only way to access input is via in terms of inner products $\phi(x_i)^T \phi(x_j)$

define inner product in higher space

How to construct Kernel:

1. $K(x,z) = x^T z$ linear
2. $K(x,z) = c K_1(x,z)$ $c \geq 0$
3. $K(x,z) = K_1(x,z) + K_2(x,z)$
4. $K(x,z) = g(K_1(x,z))$ g polynomial with +ve coff
5. $K(x,z) = K_1(x,z) K_2(x,z)$
6. $f(x) K(x,z) f(z)$
7. exponentiation $e^{K_1(x,z)}$
8. Any +ve semidefinite matrix A
   $K(x,z) = x^T A z$

Proof RBF is kernel

$$K(x,z) = e^{-\frac{(x-z)^T(x-z)}{\sigma^2}}$$

$$= \underbrace{e^{-\frac{x^T x}{\sigma^2}}}_{f(x)} e^{\frac{2x^T z}{\sigma^2}} \underbrace{e^{\frac{-z^T z}{\sigma^2}}}_{f(z)}$$

- $x^T z$ is a linear kernel $\Rightarrow$ multiply by $\frac{2}{\sigma^2} \Rightarrow \frac{2x^T z}{\sigma^2}$
- exponiate it
  $\to e^{\frac{2x^T z}{\sigma^2}}$

- Any function pre post multiply $f(x)$ & $f(z)$
  so RBF is a kernel well-defined one.

Set kernel - kernel b/w sets finite

$$K(S_1, S_2) = e^{|S_1 \cap S_2|}$$ $|S_1 \cap S_2|$ - no. of common element

* Data points that are similar have large inner product
" " dissimilar have small inner product

---

Kernelize a nearest neighbor algorithm:

euclidean distance

$$d(x,z) \quad \|x-z\|_2$$

square it (doesn't change the distance)

$$d(x,z)^2 = \|x-z\|_2^2$$

$$= x^T x - 2x^T z + z^T z$$

inner product

$$= k(x,x) - 2k(x,z) + k(z,z)$$

* Kernelizing Reduces Bias. so not required as N.N has a variance problem, not Bias

---

Closed form sol^n for OLS:

$$w = (XX^T)^{-1} X y$$ $X = [x_1 \dots x_n]$ $Y = [y_1 \dots y_n]^T$

$XX^T$ covariance matrix

In terms where $w = \sum_{i=1}^{n} \alpha_i x_i$

$$\Rightarrow w = X\alpha$$

$$X\alpha = (XX^T)^{-1} XY$$

Multiply $(XX^T)^{-1} X^T$

$$\cancel{(XX^T)^{-1} X^T X}\alpha = (X^T X)^{-1} \underbrace{X^T (XX^T)^{-1}}_{I} Xy$$

Now $X^T (XX^T)^{-1} = I$

so, $\alpha = (\underbrace{X^T X}_{K})^{-1} y$

$K$ where $K_{ij} = x_i^T x_j$

$$\alpha = K^{-1} Y$$

we assumed that $X^T X$ is invertible

when not

$$\alpha = (K + \sigma I)^{-1} Y$$ — Kernelized Ridge Regression