Lecture 9

Wednesday, May 23, 2018 4:59 PM

Kernel-based methods

kernel $k(x, x')$

similarity measure

linear kernel $k(x, x') = x^T x'$

$x \perp x' \quad k(x, x') = 0$

$$\boxed{f(x) = \sum_{i=1}^{n} \alpha_i \, k(x_i, x)}$$

weights

datapoints
training set

x is a vector vectorial data

$x \in \mathbb{R}^d$

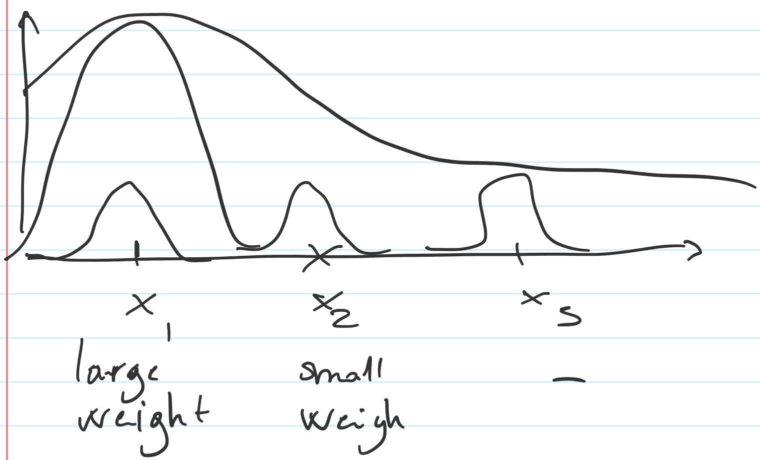
Polynomial kernel:

$$k(x, x') = \{ x^T x' + b \}^p$$

Gaussian RBF

$$k(x, x') = \exp\left(-\frac{\|x - x'\|_2^2}{\sigma^2}\right)$$

$$k(x, x') = \exp\left(- \frac{\|x - x'\|_2}{\sigma^2}\right)$$



$$\min_{f} \ \frac{1}{n} \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2 + \lambda \|f\|^2$$

$$f(x) = \sum_{j=1}^{n} \alpha_j k(x_j, x)$$

$$f(x_i) = \sum_{j=1}^{n} \alpha_j k(x_j, x_i).$$

$$K = \begin{array}{cc} & j \\ i & \begin{pmatrix} & \vdots & \\ \cdots & k(x_i, x_j) & \\ & & \end{pmatrix} \end{array}$$

$$f(x_i) = (K\alpha)_i$$

$$= \begin{pmatrix} \vdots \\ \sum_{j=1}^{n} \alpha_j k(x_j, x_i) \\ \vdots \end{pmatrix}$$

$$\|f\|^2 = \alpha^T K \alpha$$

Min f as parameterized by α

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - (K\alpha)_i\right)^2 + \lambda\,\alpha^T K \alpha$$

$$\min_{\boxed{\alpha \in \mathbb{R}^n}} \left(\frac{1}{n}\sum_{i=1}^{n}\left(y_i - (K\alpha)_i\right)^2 + \lambda \boxed{\alpha^T K \alpha} \right)$$

K is positive semi-definite

$(\boxed{\forall n \in \mathbb{N}},\ \boxed{\forall x_1, \dots, x_n},\ \boxed{\forall \beta_1, \dots, \beta_n}$

$$\beta^T K \beta \geq 0$$

$$f(x) = \sum_{j=1}^{m} \boxed{\alpha_j}\, k\left(\boxed{x'_j}, x\right)$$

training set $\quad x_1, \dots, x_n$

$\quad x'_1, \dots, x'_m$

$$f(x) = \sum^{m} \boxed{\alpha_i}\left(x - \boxed{x'_i}\right)^2$$

$$f(x) = \sum_{j=1}^{m} \boxed{\alpha_j}\left(x - \boxed{x'_j}\right)^2 + \sum_{j=1}^{m} \boxed{\alpha_j^2}\left(x - \boxed{x'_j}\right)^3$$

- **Kernel Ridge Regression**

$$\min_{\alpha \in \mathbb{R}^n} \frac{1}{n}\sum_{i=1}^{n}\Big(y_i - \underset{\substack{\uparrow \\ f(x_i)}}{(K\alpha)_i}\Big)^2 + \lambda \underset{\substack{\uparrow \\ \|f\|^2}}{\alpha^T K \alpha}$$

- **Kernel Logistic Regression**

$$\min_{\alpha \in \mathbb{R}^n} \frac{1}{n}\sum_{i=1}^{n} \log\Big(1 + \exp\big(-y_i \underset{\substack{\uparrow \\ f(x_i)}}{(K\alpha)_i}\big)\Big) + \lambda \underset{\substack{\uparrow \\ \|f\|^2}}{\alpha^T K \alpha}$$

- **Kernel Support Vector Machine**

$$\min_{\alpha \in \mathbb{R}^n} \frac{1}{n}\sum_{i=1}^{n} \max\Big(0,\ 1 - y_i \underset{\substack{\uparrow \\ f(x_i)}}{(K\alpha)_i}\Big)^p + \lambda\, \alpha^T K \alpha$$

$$+ \lambda \, \alpha^\top K \alpha$$

$$\uparrow$$

$$\|f\|^2$$

$$f(x) = \sum_{j=1}^{n} \boxed{\alpha_j} \, k(x_j, x)$$

$$f(x^*) = \sum_{j=1}^{n} \alpha_j \, \overbrace{k(x_j, x^*)}^{k(x^*, x_j)}$$

$$K = \begin{array}{c|c} & 1 \longrightarrow n \\ \hline \begin{array}{c} 1 \\ \downarrow \\ n \end{array} & \begin{pmatrix} k(x_1, x_1) & & \cdot^{\cdot^{\cdot}} \\ & \ddots & \\ \cdot^{\cdot^{\cdot}} & & k(x_n, x_n) \end{pmatrix} \end{array}$$

$$k(x^*, x_1) \; \ldots \; k(x^*, x_n)$$

Kernels on strings
trees
graphs
histograms
⋮

x_1 = ACGTTAGC $\qquad y_1 = +1$

x_2 = ACGTAGCT $\qquad y = -1$

⋮ $\qquad$ ⋮

$$k(x, x')$$

$$= \sum_{\substack{s \text{ substrings} \\ \text{of length } L \\ \text{in sequence } x}} \; \sum_{\substack{s' \text{ substrings} \\ \text{of length } L \\ \text{in sequence } x'}} \delta(s, s')$$

$$\delta(s, s') = \begin{cases} 1 & \text{if } s = s' \\ 0 & \text{otherwise} \end{cases}$$

$\underline{L = 3}$

ACG	ACG
CGT	CGT
GTT	GTA
TTA	TAG
TAG	AGC
AGC	GCT

$$k(x, x') = 4$$

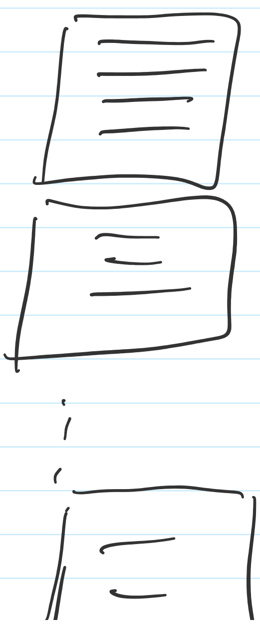
<u>Text</u>

x_1 $\qquad$ $y_1 = +1$

x_2 $\qquad$ $y_2 = -1$

$\vdots$

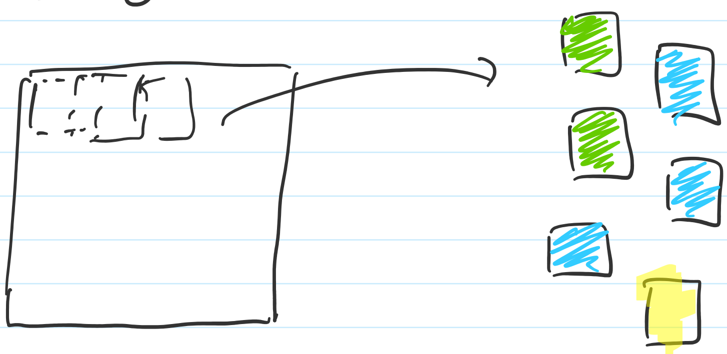
x_n $\qquad$ $y_n = +1$



x_n -

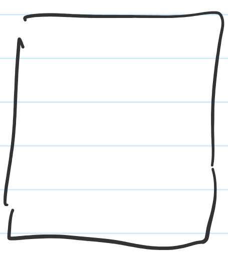
$y_n = +1$

Image



$\frac{3}{6}$. $\frac{2}{6}$. $\frac{1}{6}$

blue green yellow



$$\left[\frac{2}{6}, \frac{2}{6}, \frac{2}{6} \right]$$

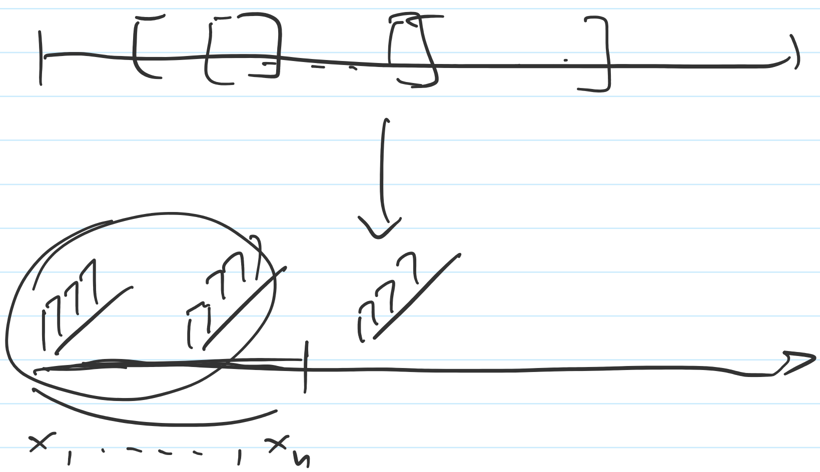
$$k(h, h') = \frac{1}{N} \sum_{j=1}^{N} \min(h_j, h'_j)$$

bins of the histograms

$$= \frac{1}{3}\left\{ \frac{2}{6} + \frac{2}{6} + \frac{1}{6} \right\}$$

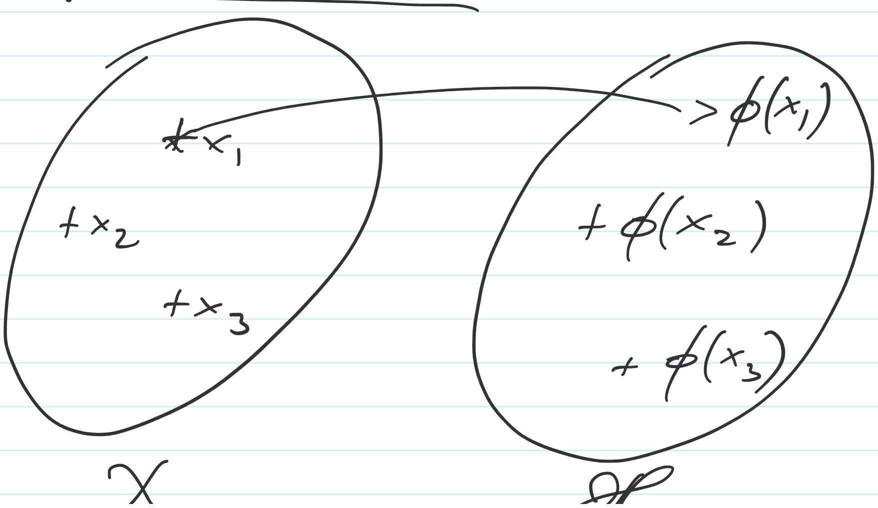
$$= \frac{1}{3} \times \frac{5}{6}$$

Signals



$$\frac{1}{n}\sum_{j=1}^{n} \|x_i - \hat{\mu}\|^2 \quad \text{small}$$

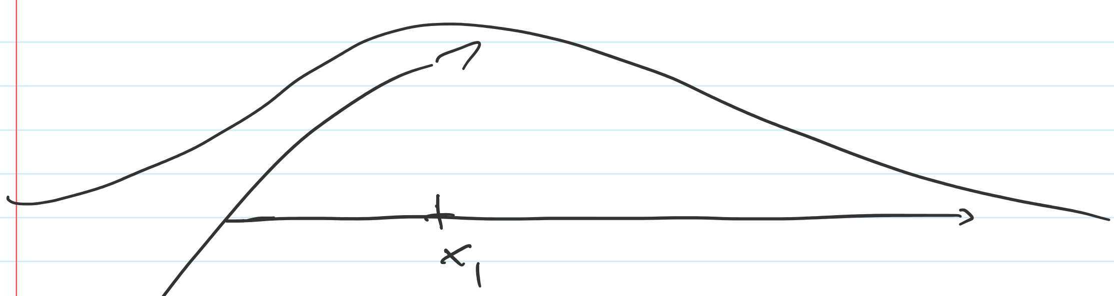
Kernel Embedding



$\mathcal{X}$ $\qquad\qquad$ $\mathcal{H}$

reproducing Kernel Hilbert space

$$\phi(x_1) = k(x_1, \cdot)$$



$$\left[y \mapsto k(x_1, y) \right]$$

$$x_1 = ACG$$

$$\phi(x_1) = \boxed{k(x_1, \cdot)}$$

$x_1, \ldots, x_n$

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

mean vector

$$\hat{\mu}_{\mathcal{H}} = \frac{1}{n}\sum_{i=1}^{n} \phi(x_i)$$

mean element

$$\langle \hat{\mu}_{\mathcal{H}}, \phi(x) \rangle_{\mathcal{H}} = \left\langle \left(\frac{1}{n}\sum_{i=1}^{n} \phi(x_i) \right), \phi(x) \right\rangle_{\mathcal{H}}$$

$$= \frac{1}{n}\sum_{i=1}^{n} \langle \phi(x_i), \phi(x) \rangle_{\mathcal{H}}$$

$$\langle k(x_i, \cdot), k(x, \cdot) \rangle_{\mathcal{H}}$$

$$\parallel$$

$$k(x_i, x)$$

$\boxed{\text{reproducing property}}$

$$\langle \bullet, \circledast \rangle_{\mathcal{H}}$$

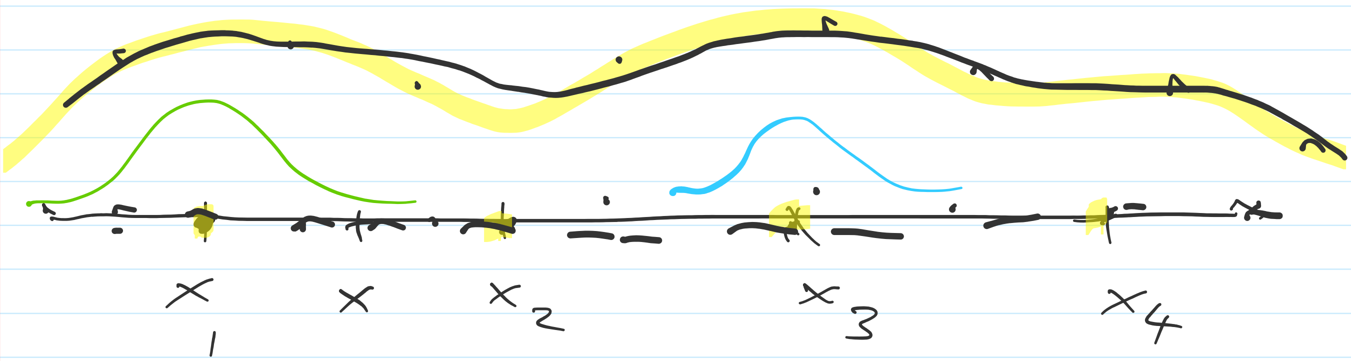
$$\ll \_ \gg$$

$$=$$

$$\bullet^{T} \circledast$$

$$= \boxed{\frac{1}{n}\sum_{i=1}^{n} k(x_i, x)}$$

$$= \boxed{\frac{1}{n}\sum_{i=1} k(x_i, x)}$$



$$k(x, x') = \exp\left(-\frac{\|x-x'\|_2^2}{\sigma^2}\right)$$

$$\left\| \phi(x) - \frac{1}{n}\sum_{i=1}^{n} \phi(x_i) \right\|_{\mathcal{H}}^2$$

$$= \left\langle \phi(x) - \frac{1}{n}\sum_{i=1}^{n} \phi(x_i),\ \phi(x) - \frac{1}{n}\sum_{i=1}^{n} \phi(x_i) \right\rangle_{\mathcal{H}}$$

$$= \langle \phi(x), \phi(x) \rangle_{\mathcal{H}}$$

$$- \left\langle \phi(x), \frac{1}{n}\sum^{n} \phi(x_i) \right\rangle_{\mathcal{H}}$$

$$- \langle \phi(x), \frac{1}{n} \sum_{i=1}^{n} \phi(x_i) \rangle_{\mathcal{H}}$$

$$- \langle \frac{1}{n} \sum_{i=1}^{n} \phi(x_i), \phi(x) \rangle_{\mathcal{H}}$$

$$+ \langle \frac{1}{n} \sum_{i=1}^{n} \phi(x_i), \frac{1}{n} \sum_{j=1}^{n} \phi(x_j) \rangle_{\mathcal{H}}$$

REPRODUCING PROPERTY

$$\langle \phi(x), \phi(x') \rangle_{\mathcal{H}}$$

$$= \langle k(x, \cdot), k(x', \cdot) \rangle_{\mathcal{H}}$$

$$= k(x, x')$$

$$= \boxed{k(x, x)} \qquad k(x, x_i)$$

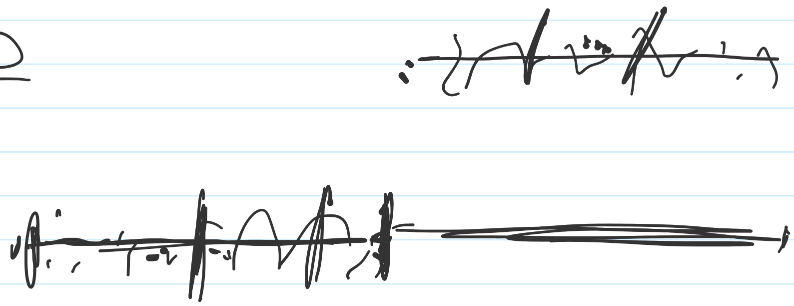
$$- \frac{1}{n} \sum_{i=1}^{n} \overbrace{\langle \phi(x), \phi(x_i) \rangle_{\mathcal{H}}}$$

$$- \frac{1}{n} \sum_{i=1}^{n} \langle \phi(x), \phi(x_i) \rangle_{\mathcal{H}}$$

$$+ \frac{1}{n^2} \sum^{n} \sum^{n} \langle \phi(x_i), \phi(x_j) \rangle_{\mathcal{H}}$$

$$+ \frac{1}{n^2} \sum_{i=1} \sum_{j=1} \langle \underbrace{\phi(x_i), \phi(x_j)}_{k(x_i, x_j)} \rangle_{\mathcal{H}}$$

<u>1D</u>



$$\max_{1 < k < n} \| \underbrace{\hat{\mu}_{k+1:n}}_{} - \hat{\mu}_{1:k} \|^2_{\mathcal{H}}$$

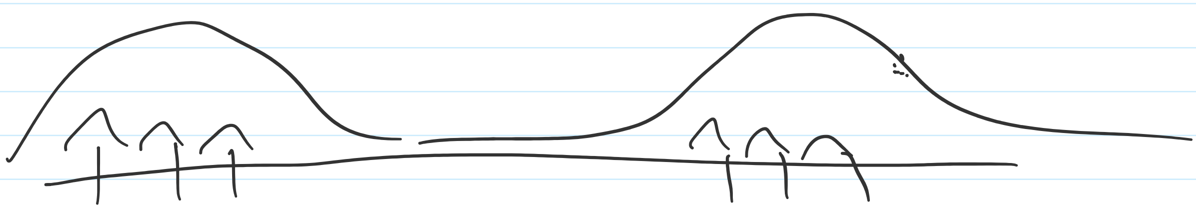
$$\frac{1}{n-k} \sum_{j=k+1}^{n} \phi(x_j) - \frac{1}{k} \sum_{j=1}^{k} \phi(x_j)$$

$$\max_{k} \| \underbrace{\frac{1}{n-k} \sum_{j=k+1}^{n} x_j}_{0} - \underbrace{\frac{1}{k} \sum_{j=1}^{k} x_j}_{1} \|^2$$

$\underline{k = k^*}$

$$\max \| \frac{1}{n-} \sum^{n} \phi(x_i) - \frac{1}{} \sum^{k} \phi(x_j) \|^2$$

$$\max_{k} \left\| \frac{1}{n-k} \sum_{j=k+1}^{n} \phi(x_j) - \frac{1}{k} \sum_{j=1}^{k} \phi(x_j) \right\|_{\mathcal{H}}$$



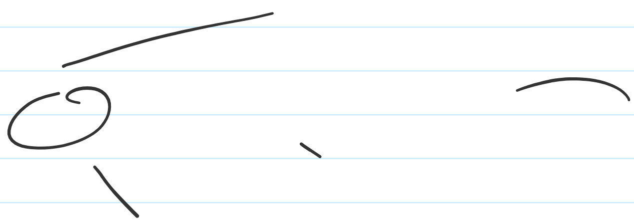
Artificial Neural Networks

$$f(x) = f_k \circ f_{k-1} \circ \dots f_1(x)$$

composition

$$f_\ell = \sigma\Big(\underbrace{\boxed{w_\ell}^T \boxed{x} + b_\ell}_{\text{affine fn of } x}\Big)$$

$$\boxed{\sigma(x) = \frac{1}{1+e^{-x}}}$$



$$\min_{\substack{w_1, \dots, w_k \\ b_1, \dots, b_k}} \overbrace{\frac{1}{n}\sum_{i=1}^{n} \Big(y_i - f_k(f_{k-1}(\cdots f_1(x)))\Big)^2}^{f(w_1, \dots, w_k, b_1, \dots, b_k)} + \boxed{?} \sum_{\ell=1}^{k} \|w_\ell\|_2^2$$

- non-convex
-

K layers

w weights

neurons $w_\ell^T x_1 + b_\ell,$

neurons $\underbrace{w_e^T x}$ $+ b_e$?

non-linearity σ

<u>learning rate</u> η

$$\cancel{\eta_t = \sin(\alpha t)}$$

- <u>Multilayer Perceptrons</u>

$$w_e^T x \longrightarrow w_e \circledast x$$

↳ Convolutional Neural Network

- Recurrent Neural Network

$$x_1, \dots, x_t, \dots$$

$$x_{t+1} = f_k(\cdots f_1(x_t))$$

$$\nabla_{w_e} \longrightarrow \text{chain rule}$$