MBI

Molecular Biomedical Informatics

分 子 生 醫 資 訊 實 驗 室

Machine Learning and Bioinformatics

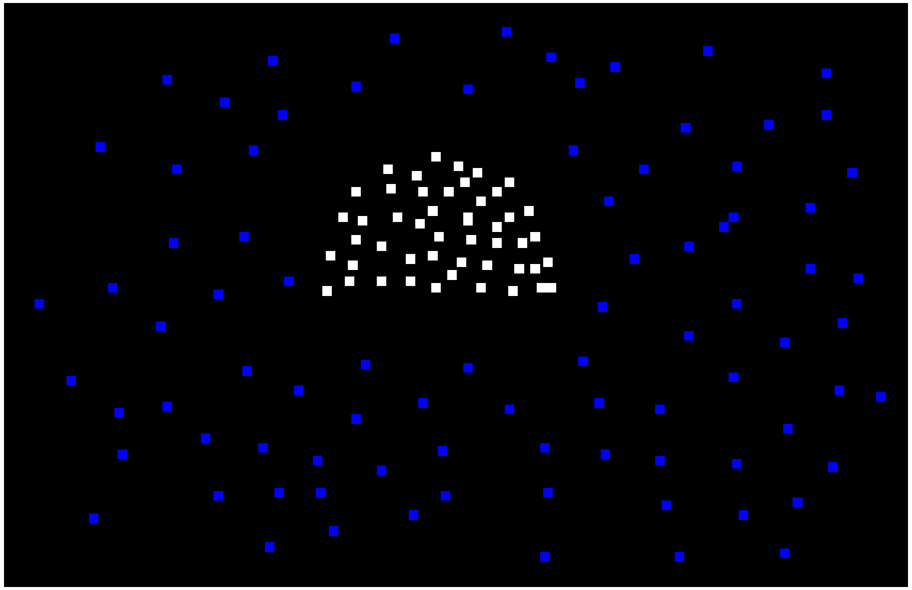
機 器 學 習 與 生 物 資 訊 學

Enhancing instance-based classification using kernel density estimation

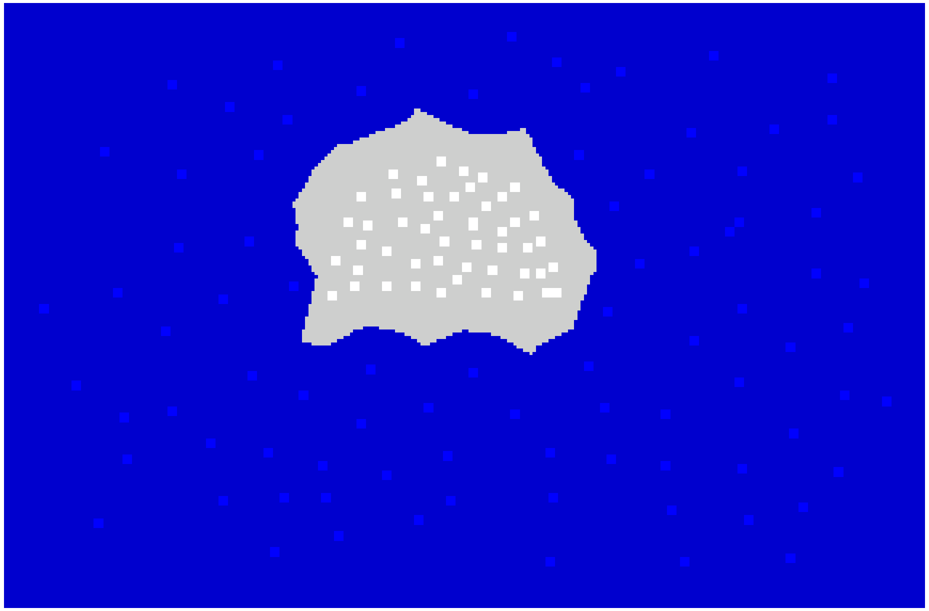
Outline

- Kernel approximation
- Kernel approximation in $O(n)$
- Multi-dimensional kernel approximation
- Kernel density estimation
- Application in data classification

Identifying class boundary



Boundary identified



Kernel approximation

Problem definition

- Given the values of function $f(\mathbf{v})$ at a set of samples $S = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\}$
- We want to find a set of symmetric kernel functions $K(\mathbf{v}, c_i, b_i)$ and the corresponding weights w_i such that $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot K(\mathbf{v}, c_i, b_i) \cong f(\mathbf{v})$

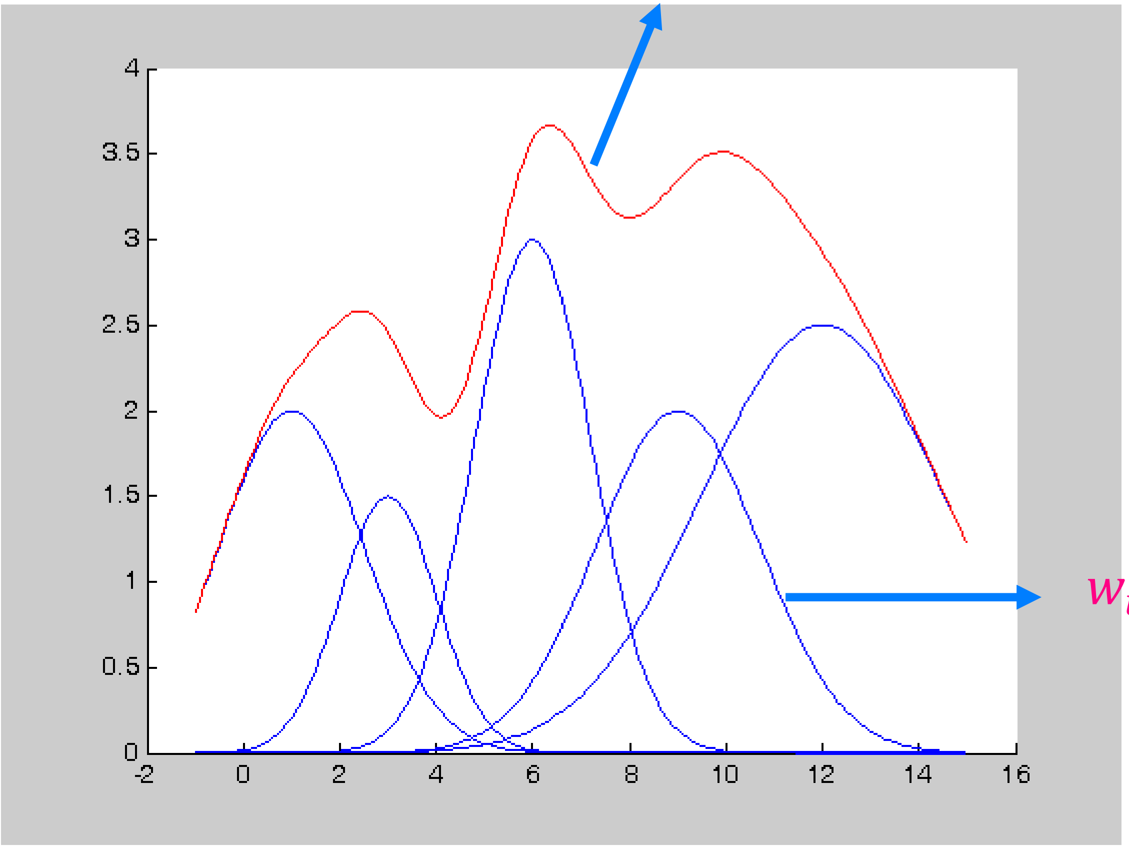
Equations

- $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot K(\mathbf{v}, c_i, b_i) \cong f(\mathbf{v})$ (1)

Kernel approximation

An 1-D example

$$\hat{f}(\mathbf{v}) = \sum_i w_i \cdot K(\boldsymbol{v}, c_i, b_i) \cong f(\mathbf{v})$$



$w_i \cdot K(\boldsymbol{v}, c_i, b_i)$

Spherical Gaussian functions

- Hartman *et al.* showed that a linear combination of spherical Gaussian functions can approximate any function with arbitrarily small error
- "Layered neural networks with Gaussian hidden units as universal approximations", Neural Computation, Vol. 2, No. 2, 1990
- With the Gaussian kernel functions, we want to find $w_i, \boldsymbol{\mu}_i,$ and σ_i such that

$$\hat{f}(\mathbf{v}) = \sum_i w_i \cdot \exp\left(-\frac{||\mathbf{v} - \boldsymbol{\mu}_i||^2}{2\sigma_i^2}\right) \cong f(\mathbf{v})$$

Equations

- $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot K(\mathbf{v}, c_i, b_i) \cong f(\mathbf{v})$ (1)

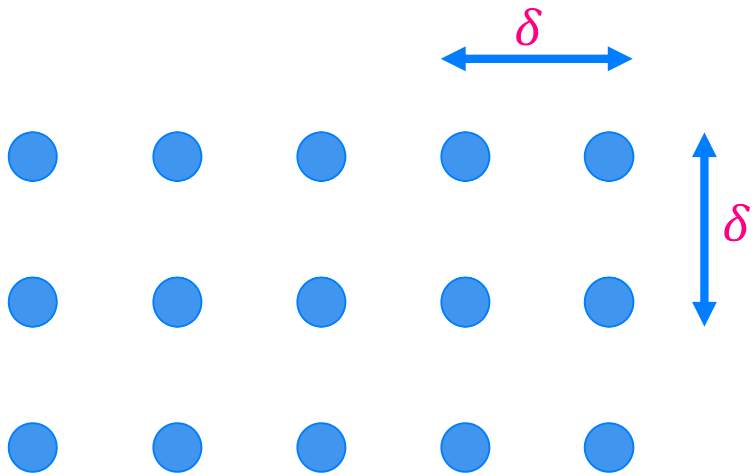
- $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot \exp\left(-\frac{\|\mathbf{v}-\boldsymbol{\mu}_i\|^2}{2\sigma_i^2}\right) \cong f(\mathbf{v})$ (2)

Kernel approximation in $O(n)$

Kernel approximation in $O(n)$

- In the learning algorithm, we assume uniform sampling
  - namely, samples are located at the crosses of an evenly-spaced grid in the d-dimensional vector space
- Let δ denote the distance between two adjacent samples
- If the assumption of uniform sampling does not hold, then some sort of interpolation can be conducted to obtain the approximate function values at the crosses of the grid

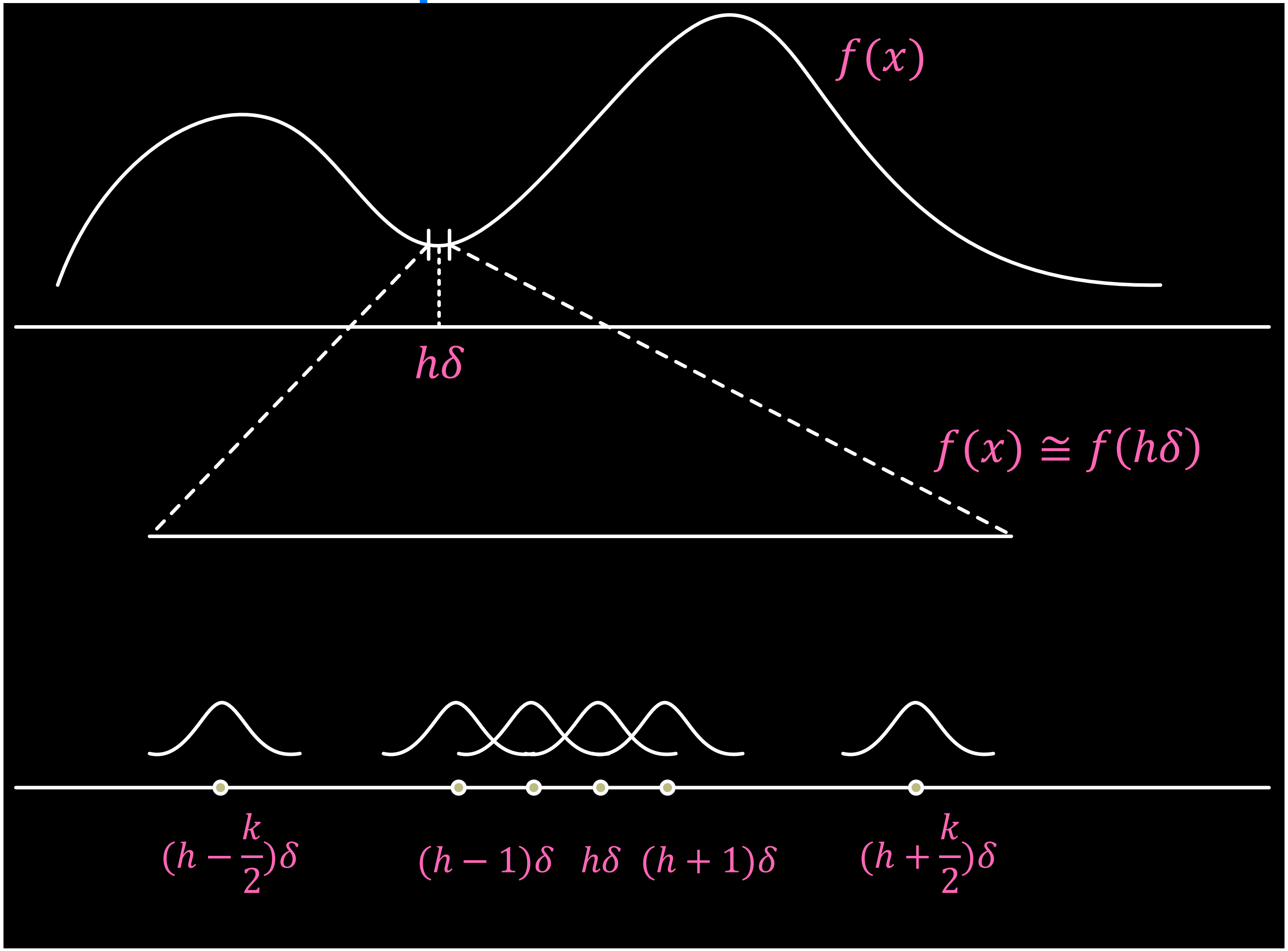
A 2-D example of uniform sampling



The $O(n)$ kernel approximation

- Under the assumption that the sampling density is sufficiently high, *i.e.* $\delta \rightarrow 0$, we have the function values at a sample $\mathbf{s}_h$ and its k nearest samples, $\ddot{\mathbf{s}}_1, \ddot{\mathbf{s}}_2, \ldots, \ddot{\mathbf{s}}_k$, are virtually equal: $f(\mathbf{s}_h) \cong f(\ddot{\mathbf{s}}_1) \cong f(\ddot{\mathbf{s}}_2) \cong \cdots \cong f(\ddot{\mathbf{s}}_k)$
- In other words, $f(\boldsymbol{v})$ is virtually a constant function equal to $f(\mathbf{s}_h)$ in the proximity of $\mathbf{s}_h$
- Accordingly, we can expect that $\ddot{w}_1 \cong \ddot{w}_2 \cong \cdots \cong \ddot{w}_k \cong w_h$; $\ddot{\sigma}_1 \cong \ddot{\sigma}_2 \cong \cdots \cong \ddot{\sigma}_k \cong \sigma_h$

An 1-D example



- Under the assumption that $\delta \to 0$, we have
  - $f(x) \cong f(h\delta)$ for $x \cong h\delta$
  - $w_{h-\frac{k}{2}} \cong w_{h-\frac{k}{2}+1} \cong \cdots \cong w_h \cong \cdots \cong w_{h+\frac{k}{2}}$
  - $\sigma_{h-\frac{k}{2}} \cong \sigma_{h-\frac{k}{2}+1} \cong \cdots \cong \sigma_h \cong \cdots \cong \sigma_{h+\frac{k}{2}}$
- The issue now is to find appropriate w_h and σ_h such that

$$\sum_{i=h-\frac{k}{2}}^{h+\frac{k}{2}} w_h \cdot \exp\left(-\frac{(x-i\delta)^2}{2\sigma_h^2}\right) \cong f(h\delta)$$

in the proximity of $h\delta$

Equations

- $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot K(\mathbf{v}, c_i, b_i) \cong f(\mathbf{v})$ (1)

- $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot \exp\left(-\frac{\|\mathbf{v}-\boldsymbol{\mu}_i\|^2}{2\sigma_i^2}\right) \cong f(\mathbf{v})$ (2)

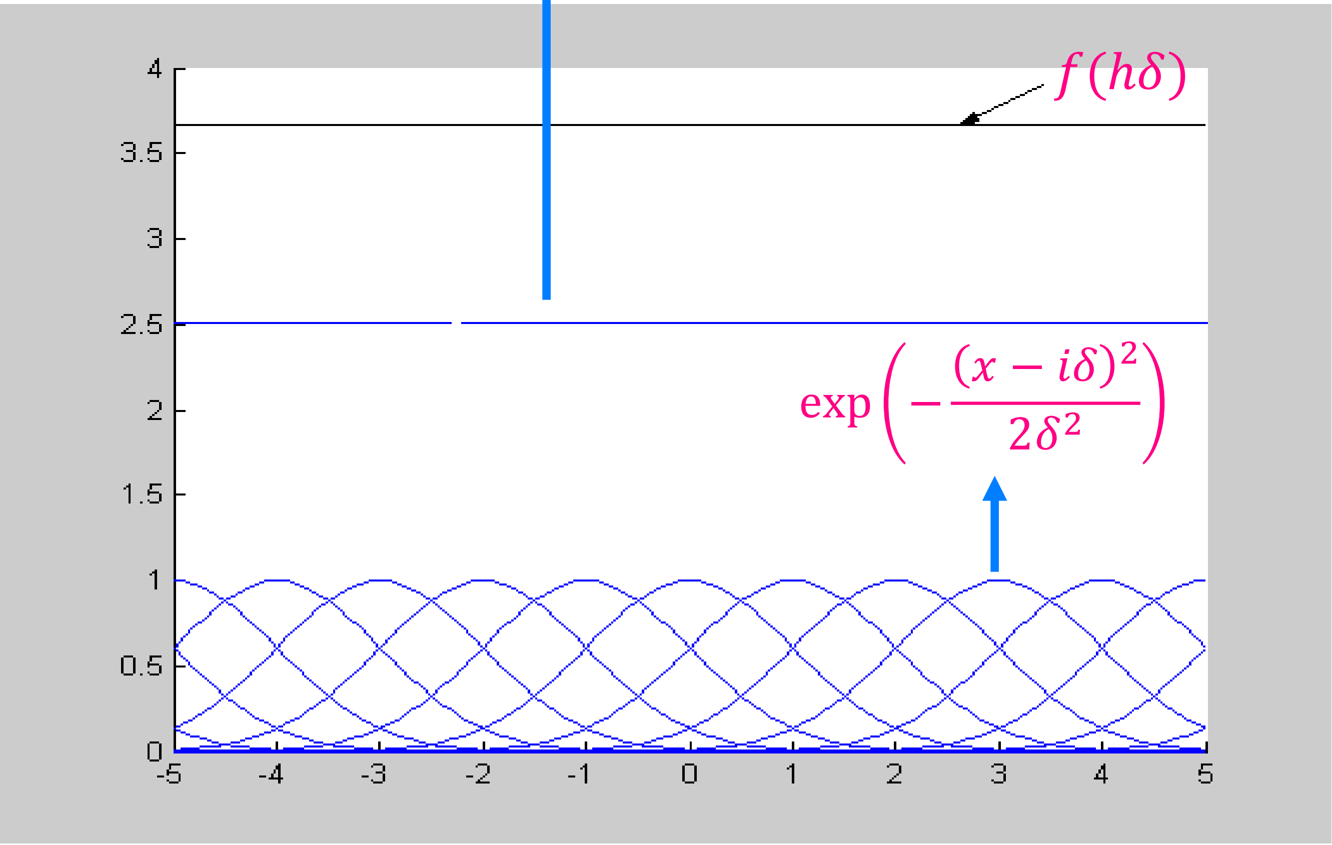
- $\sum_{i=h-\frac{k}{2}}^{h+\frac{k}{2}} w_h \cdot \exp\left(-\frac{(x-i\delta)^2}{2\sigma_h^2}\right) \cong f(h\delta)$ (3)

What's

the difference between equations (2) and (3)?

If we set $\sigma_h = \delta$, we have

$$\sum_{i=h-\frac{k}{2}}^{h+\frac{k}{2}} \exp\left(-\frac{(x-i\delta)^2}{2\delta^2}\right) \cong 2.5066, \text{ if } k \text{ is sufficiently large}$$



- Therefore, with $\sigma_h = \delta$, we can set $w_h = \frac{f(h\delta)}{2.5066}$ and obtain

$$\sum_{h-\frac{k}{2}}^{h+\frac{k}{2}} \frac{f(h\delta)}{2.5066} \cdot \exp\left(-\frac{(x-i\delta)^2}{2\sigma_h^2}\right) \cong f(h\delta)$$

- In fact, it can be shown that $\sum_{i=-\infty}^{\infty} \exp\left(-\frac{(x-i\delta)^2}{2\delta^2}\right)$ is bounded by $2.5066282745 \pm 1.35 \times 10^{-8}$
- Therefore, we have the following function approximate:

$$\hat{f}(x) = \frac{1}{2.5066} \cdot \sum_{i=-\infty}^{\infty} f(i\delta) \cdot \exp\left(-\frac{(x-i\delta)^2}{2\delta^2}\right)$$

Equations

- $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot K(\mathbf{v}, c_i, b_i) \cong f(\mathbf{v})$ (1)
- $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot \exp\left(-\frac{\|\mathbf{v}-\boldsymbol{\mu}_i\|^2}{2\sigma_i^2}\right) \cong f(\mathbf{v})$ (2)
- $\sum_{i=h-\frac{k}{2}}^{h+\frac{k}{2}} w_h \cdot \exp\left(-\frac{(x-i\delta)^2}{2\sigma_h^2}\right) \cong f(h\delta)$ (3)
- $\sum_{i=h-\frac{k}{2}}^{h+\frac{k}{2}} \frac{f(h\delta)}{2.5066} \cdot \exp\left(-\frac{(x-i\delta)^2}{2\sigma_h^2}\right) \cong f(h\delta)$ (4)
- $\hat{f}(x) = \frac{1}{2.5066} \cdot \sum_{i=-\infty}^{\infty} f(i\delta) \cdot \exp\left(-\frac{(x-i\delta)^2}{2\delta^2}\right)$ (5)

Generalization

- We can generalize the result by setting $\sigma_1 = \sigma_2 = \cdots = \sigma_n = \beta\delta$, where β is a real number

$$\hat{f}(x) = \frac{1}{\lambda} \cdot \sum_{i=-\infty}^{\infty} f(i\delta) \cdot \exp\left(-\frac{(x-i\delta)^2}{2\delta^2}\right),$$

$$\text{where } \lambda = \sum_{j=-\infty}^{\infty} \exp\left(-\frac{j^2}{2\beta^2}\right)$$

- The table shows the bounds of $\sum_{j=-\infty}^{\infty} \exp\left(-\frac{(x-j\delta)^2}{2\beta^2\delta^2}\right)$

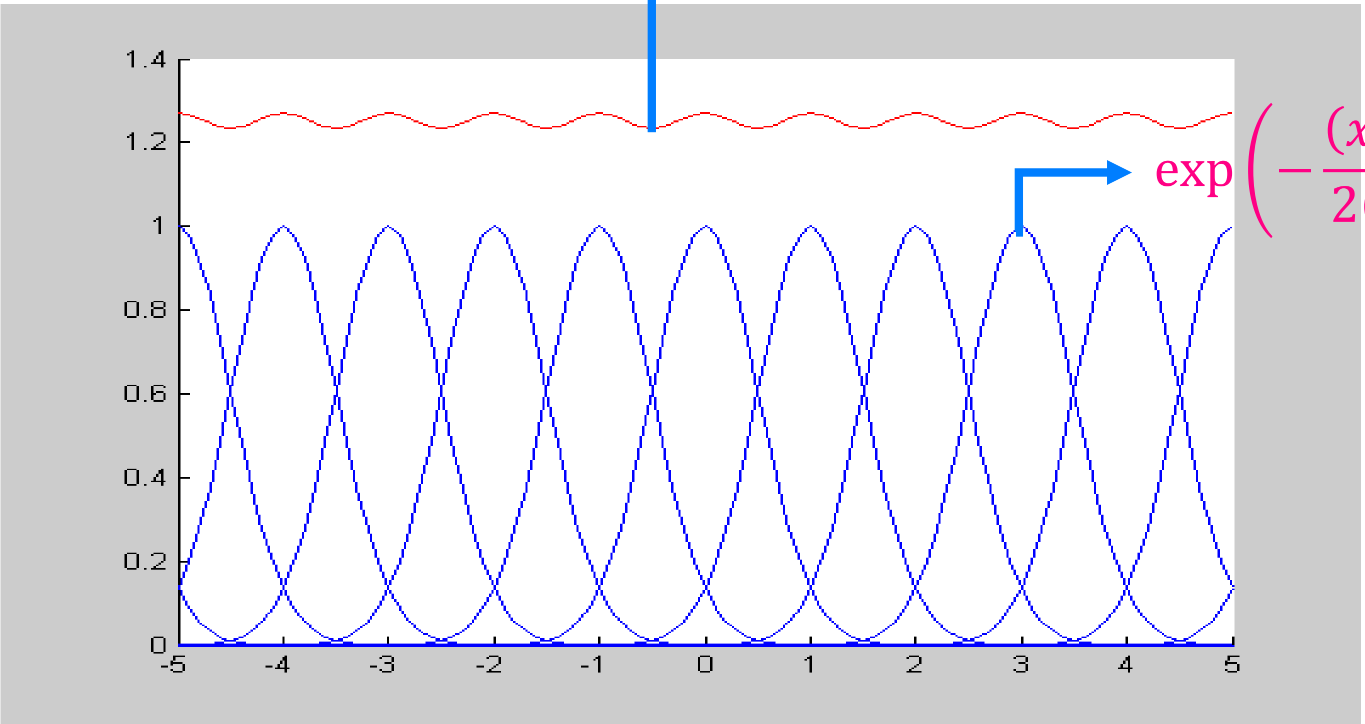
β	**Bounds**
0.5	$1.253 \pm 1.8 \times 10^{-2}$
1.0	$2.5066282745 \pm 1.35 \times 10^{-8}$
1.5	$3.7599424119339 \pm 2.94 \times 10^{-11}$

Equations

- $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot K(\mathbf{v}, c_i, b_i) \cong f(\mathbf{v})$ (1)
- $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot \exp\left(-\frac{\|\mathbf{v}-\boldsymbol{\mu}_i\|^2}{2\sigma_i^2}\right) \cong f(\mathbf{v})$ (2)
- $\sum_{i=h-\frac{k}{2}}^{h+\frac{k}{2}} w_h \cdot \exp\left(-\frac{(x-i\delta)^2}{2\sigma_h^2}\right) \cong f(h\delta)$ (3)
- $\sum_{i=h-\frac{k}{2}}^{h+\frac{k}{2}} \frac{f(h\delta)}{2.5066} \cdot \exp\left(-\frac{(x-i\delta)^2}{2\sigma_h^2}\right) \cong f(h\delta)$ (4)
- $\hat{f}(x) = \frac{1}{2.5066} \cdot \sum_{i=-\infty}^{\infty} f(i\delta) \cdot \exp\left(-\frac{(x-i\delta)^2}{2\delta^2}\right)$ (5)
- $\hat{f}(x) = \frac{1}{\lambda} \cdot \sum_{i=-\infty}^{\infty} f(i\delta) \cdot \exp\left(-\frac{(x-i\delta)^2}{2\delta^2}\right)$ (6)

The effect of β

$$\sum_{j=-\infty}^{\infty} \exp\left(-\frac{(x-j\delta)^2}{2(0.5\delta)^2}\right)$$

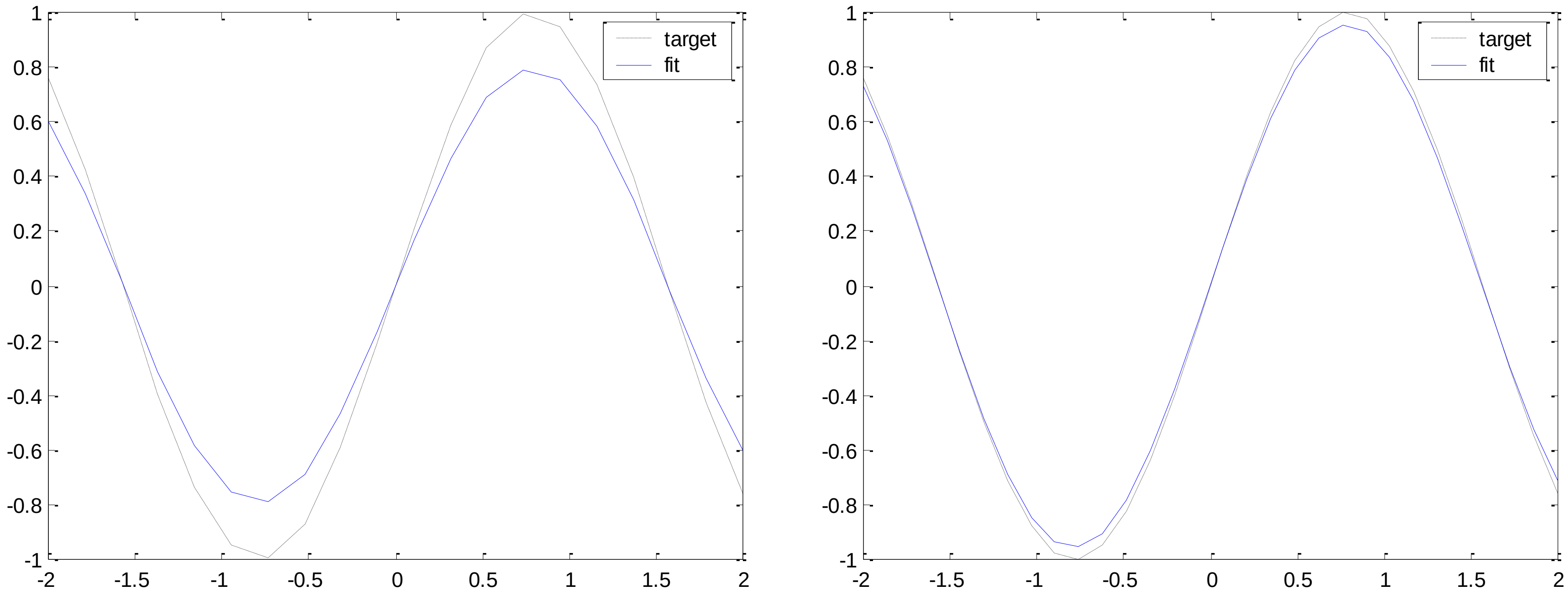


$$\exp\left(-\frac{(x-j\delta)^2}{2(0.5\delta)^2}\right)$$

The smoothing effect

- The kernel approximation function is actually a weighted average of the sampled function values
- Selecting a larger β value implies that the smoothing effect will be more significant
- Our suggestion is set $1 \leq \beta \leq 10$

An example of the smoothing effect



Multi-dimensional kernel approximation

The basic concepts

- Under the assumption that the sampling density is sufficiently high, *i.e.* $\delta \to 0$, we have the function values at a sample $\mathbf{s}_h$ and its k nearest samples, $\ddot{\mathbf{s}}_1, \ddot{\mathbf{s}}_2, \ldots, \ddot{\mathbf{s}}_k$, are virtually equal: $f(\mathbf{s}_h) \cong f(\ddot{\mathbf{s}}_1) \cong f(\ddot{\mathbf{s}}_2) \cong \cdots \cong f(\ddot{\mathbf{s}}_k)$
- As a result, we can expect that $\ddot{w}_1 \cong \ddot{w}_2 \cong \cdots \cong \ddot{w}_k \cong w_h$ and $\ddot{\sigma}_1 \cong \ddot{\sigma}_2 \cong \cdots \cong \ddot{\sigma}_k \cong \sigma_h$, where $\ddot{w}_1, \ddot{w}_2, \ldots, \ddot{w}_k$ and $\ddot{\sigma}_1, \ddot{\sigma}_2, \ldots, \ddot{\sigma}_k$ are the weights and bandwidths of the Gaussian functions located at $\ddot{\mathbf{s}}_1, \ddot{\mathbf{s}}_2, \ldots, \ddot{\mathbf{s}}_k$, respectively

- $\hat{f}(\mathbf{v}) = \sum_{\mathbf{s}_i} w_i \cdot \exp\left(-\frac{\|\mathbf{v}-\mathbf{s}_i\|^2}{2\sigma_i^2}\right)$

$$\cong w_h \cdot \exp\left(-\frac{\|\mathbf{v}-\mathbf{s}_h\|^2}{2\sigma_h^2}\right) + \sum_{\boldsymbol{j=1}}^{\boldsymbol{k}} \ddot{w}_j \cdot \exp\left(-\frac{\left\|\mathbf{v}-\ddot{\mathbf{s}}_j\right\|^2}{2\ddot{\sigma}_j^2}\right)$$

$$\cong w_h \cdot \exp\left(-\frac{\|\mathbf{v}-\mathbf{s}_h\|^2}{2\sigma_h^2}\right) + \sum_{\boldsymbol{j=1}}^{\boldsymbol{k}} w_h \cdot \exp\left(-\frac{\left\|\mathbf{v}-\ddot{\mathbf{s}}_j\right\|^2}{2\sigma_h^2}\right) \cong f(\mathbf{v})$$

- Let $\mathbf{v} = (x_1, x_2, \dots, x_d)$, the scalar form of $-\frac{\left\|\mathbf{v}-\ddot{\mathbf{s}}_j\right\|^2}{2\sigma_h^2}$ is $\sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} \cdots \sum_{j_d=-\infty}^{\infty} \left(-\frac{\|\mathbf{v}-(j_1\delta, j_2\delta, \dots, j_d\delta\ )\|^2}{2\sigma_h^2}\right)$

$$= \sum_{j_1=-\infty}^{\infty} \left(-\frac{(x_1 - j_1\delta)^2}{2\sigma_h^2}\right) \cdots \sum_{j_d=-\infty}^{\infty} \left(-\frac{(x_1 - j_d\delta)^2}{2\sigma_h^2}\right)$$

- If we set σ_h to δ, then $\sum_{j_1=-\infty}^{\infty} \left(-\frac{(x_1-j_1\delta)^2}{2\sigma_h^2}\right)$ is bounded by $2.5066282745 \pm 1.35 \times 10^{-8}$

- Accordingly, we have $-\frac{\|\mathbf{v}-\ddot{\mathbf{s}}_j\|^2}{2\sigma_h^2}$ is bounded by $(2.5066282745 \pm 1.35 \times 10^{-8})^d$ and, with $\sigma_h = \delta$,

$$w_h = \frac{f(\mathbf{s}_h)}{(2.5066)^d}$$

- In RVKDE, d is relaxed to a parameter, α
- Finally, by setting $\sigma_i, i = 1,2, \ldots, k$ uniformly to δ, we obtain the following function that approximates $f(\mathbf{v})$:
  $\hat{f}(\mathbf{v}) = \sum_{\hat{\mathbf{s}}_i} \frac{f(\ddot{\mathbf{s}}_i)}{(2.5066)^d} \cdot \exp\left(-\frac{\|\mathbf{v}-\ddot{\mathbf{s}}_i\|^2}{2\delta^2}\right)$, where $\ddot{\mathbf{s}}_1, \ddot{\mathbf{s}}_2, \ldots, \ddot{\mathbf{s}}_k$ are the k nearest samples of the object located at $\mathbf{v}$

Equations

- $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot K(\mathbf{v}, c_i, b_i) \cong f(\mathbf{v})$ (1)
- $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot \exp\left(-\frac{\|\mathbf{v}-\boldsymbol{\mu}_i\|^2}{2\sigma_i^2}\right) \cong f(\mathbf{v})$ (2)
- $\sum_{i=h-\frac{k}{2}}^{h+\frac{k}{2}} w_h \cdot \exp\left(-\frac{(x-i\delta)^2}{2\sigma_h^2}\right) \cong f(h\delta)$ (3)
- $\sum_{i=h-\frac{k}{2}}^{h+\frac{k}{2}} \frac{f(h\delta)}{2.5066} \cdot \exp\left(-\frac{(x-i\delta)^2}{2\sigma_h^2}\right) \cong f(h\delta)$ (4)
- $\hat{f}(x) = \frac{1}{2.5066} \cdot \sum_{i=-\infty}^{\infty} f(i\delta) \cdot \exp\left(-\frac{(x-i\delta)^2}{2\delta^2}\right)$ (5)
- $\hat{f}(x) = \frac{1}{\lambda} \cdot \sum_{i=-\infty}^{\infty} f(i\delta) \cdot \exp\left(-\frac{(x-i\delta)^2}{2\delta^2}\right)$ (6)
- $\hat{f}(\mathbf{v}) = \sum_{\hat{\mathbf{s}}_i} \frac{f(\ddot{\mathbf{s}}_i)}{(2.5066)^d} \cdot \exp\left(-\frac{\|\mathbf{v}-\ddot{\mathbf{s}}_i\|^2}{2\delta^2}\right)$ (7)

Generalization

- If we set $\sigma_i, i = 1,2,\dots,k$ uniformly to $\beta\delta$, then $\sum_{j_1=-\infty}^{\infty}\sum_{j_2=-\infty}^{\infty}\cdots\sum_{j_d=-\infty}^{\infty}\left(-\frac{\|\mathbf{v}-(j_1\delta,j_2\delta,\dots,j_d\delta)\|^2}{2\sigma_h^2}\right)$
  $$=\sum_{j_1=-\infty}^{\infty}\left(-\frac{(x_1-j_1\delta)^2}{2\beta^2\delta^2}\right)\cdots\sum_{j_d=-\infty}^{\infty}\left(-\frac{(x_1-j_d\delta)^2}{2\beta^2\delta^2}\right)$$
- As a result, we want to set $w_h = \frac{f(\mathbf{s}_h)}{\lambda^d}$, where $\lambda = \sum_{j=-\infty}^{\infty}\exp\left(-\frac{j^2}{2\beta^2}\right)$ and obtain the following function approximate: $\hat{f}(\mathbf{v}) = \frac{1}{\lambda^d}\cdot\sum_{i=1}^{k} f(\ddot{\mathbf{s}}_i)\cdot\exp\left(-\frac{\|\mathbf{v}-\ddot{\mathbf{s}}_i\|^2}{2\beta^2\delta^2}\right)$, where $\ddot{\mathbf{s}}_1, \ddot{\mathbf{s}}_2, \dots, \ddot{\mathbf{s}}_k$ are the k nearest samples of the object located at $\mathbf{v}$
- In RVKDE, k is a parameter kt

Equations

- $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot K(\mathbf{v}, c_i, b_i) \cong f(\mathbf{v})$ (1)
- $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot \exp\left(-\frac{\|\mathbf{v}-\boldsymbol{\mu}_i\|^2}{2\sigma_i^2}\right) \cong f(\mathbf{v})$ (2)
- $\sum_{i=h-\frac{k}{2}}^{h+\frac{k}{2}} w_h \cdot \exp\left(-\frac{(x-i\delta)^2}{2\sigma_h^2}\right) \cong f(h\delta)$ (3)
- $\sum_{i=h-\frac{k}{2}}^{h+\frac{k}{2}} \frac{f(h\delta)}{2.5066} \cdot \exp\left(-\frac{(x-i\delta)^2}{2\sigma_h^2}\right) \cong f(h\delta)$ (4)
- $\hat{f}(x) = \frac{1}{2.5066} \cdot \sum_{i=-\infty}^{\infty} f(i\delta) \cdot \exp\left(-\frac{(x-i\delta)^2}{2\delta^2}\right)$ (5)
- $\hat{f}(x) = \frac{1}{\lambda} \cdot \sum_{i=-\infty}^{\infty} f(i\delta) \cdot \exp\left(-\frac{(x-i\delta)^2}{2\delta^2}\right)$ (6)
- $\hat{f}(\mathbf{v}) = \sum_{\hat{\mathbf{s}}_i} \frac{f(\ddot{\mathbf{s}}_i)}{(2.5066)^d} \cdot \exp\left(-\frac{\|\mathbf{v}-\ddot{\mathbf{s}}_i\|^2}{2\delta^2}\right)$ (7)
- $\hat{f}(\mathbf{v}) = \frac{1}{\lambda^d} \cdot \sum_{i=1}^{k} f(\ddot{\mathbf{s}}_i) \cdot \exp\left(-\frac{\|\mathbf{v}-\ddot{\mathbf{s}}_i\|^2}{2\beta^2\delta^2}\right)$ (8)

Simplification

- An interesting observation is that, as long as $\beta \geq 0.5$, then regardless of the value of $\beta = \frac{\sigma_i}{\delta_i}$, we have $\frac{\lambda}{\beta} \cong \sqrt{2\pi}$
- With this observation, we have $\hat{f}(\mathbf{v}) = \sum_{i=1}^{kt} \left(\frac{1}{\sqrt{2\pi} \cdot \ddot{\sigma}_i}\right)^{\alpha} f(\ddot{\mathbf{s}}_i) \cdot \exp\left(-\frac{\|\mathbf{v}-\ddot{\mathbf{s}}_i\|^2}{2\beta^2\delta^2}\right)$

Equations

- $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot K(\mathbf{v}, c_i, b_i) \cong f(\mathbf{v})$ (1)
- $\hat{f}(\mathbf{v}) = \sum_i w_i \cdot \exp\left(-\frac{\|\mathbf{v}-\boldsymbol{\mu}_i\|^2}{2\sigma_i^2}\right) \cong f(\mathbf{v})$ (2)
- $\hat{f}(x) = \frac{1}{2.5066} \cdot \sum_{i=-\infty}^{\infty} f(i\delta) \cdot \exp\left(-\frac{(x-i\delta)^2}{2\delta^2}\right)$ (5)
- $\hat{f}(x) = \frac{1}{\lambda} \cdot \sum_{i=-\infty}^{\infty} f(i\delta) \cdot \exp\left(-\frac{(x-i\delta)^2}{2\delta^2}\right)$ (6)
- $\hat{f}(\mathbf{v}) = \sum_{\hat{\mathbf{s}}_i} \frac{f(\ddot{\mathbf{s}}_i)}{(2.5066)^d} \cdot \exp\left(-\frac{\|\mathbf{v}-\ddot{\mathbf{s}}_i\|^2}{2\delta^2}\right)$ (7)
- $\hat{f}(\mathbf{v}) = \frac{1}{\lambda^d} \cdot \sum_{i=1}^{k} f(\ddot{\mathbf{s}}_i) \cdot \exp\left(-\frac{\|\mathbf{v}-\ddot{\mathbf{s}}_i\|^2}{2\beta^2\delta^2}\right)$ (8)
- $\hat{f}(\mathbf{v}) = \sum_{i=1}^{kt} \left(\frac{1}{\sqrt{2\pi}\cdot\ddot{\sigma}_i}\right)^{\alpha} f(\ddot{\mathbf{s}}_i) \cdot \exp\left(-\frac{\|\mathbf{v}-\ddot{\mathbf{s}}_i\|^2}{2\beta^2\delta^2}\right)$ (9)

Any Questions?

 Lack something?

δ?

Kernel density estimation

Problem definition

- Assume that we are given a set of samples taken from a probability distribution in a *d*-dimensional vector space
- The problem definition of kernel density estimation is to find a linear combination of kernel functions that provides a good approximation of the probability density function

Probability density estimation

- The value of the probability density function at a vector **v** can be estimated as follows:

$$f(\mathbf{v}) \cong \frac{ks}{n} \cdot \left(\frac{R(\mathbf{v})^d \cdot \pi^{d/2}}{\Gamma(\frac{d}{2}+1)} \right)^{-1}$$

- *n* is the total number of samples
- *R*(**v**) is the distance between vector **v** and its *ks*-th nearest samples
- $\frac{R(\mathbf{v})^d \cdot \pi^{d/2}}{\Gamma(\frac{d}{2}+1)}$ is the volume of a sphere with radius = *R*(**v**) in a *d*-dimensional vector space

The Gamma function

- $\Gamma(m) = \int_0^\infty x^{m-1} e^{-x}\, dx$

$$= -x^{m-1}e^{-x}\Big|_0^\infty + (m-1)\int_0^\infty x^{m-2}e^{-x}\, dx$$

$$= (m-1)\int_0^\infty x^{m-2}e^{-x}\, dx$$

$$= (m-1)\cdot\Gamma(m-1)$$

- $\Gamma(1) = \int_0^\infty e^{-x}\, dx = 1$, since $\frac{dx^{1/2}e^{-x}}{dx} = \frac{1}{2}x^{-1/2}e^{-x}$

- $\Gamma\left(\frac{1}{2}\right) = \int_0^\infty x^{-1/2}e^{-x}\, dx = 2\left[x^{1/2}e^{-x}\Big|_0^\infty + \int_0^\infty x^{1/2}e^{-x}\, dx\right] = \int_0^\infty 4y^2e^{-y^2}\, dy$, with $y = \sqrt{x}$

$$\int_0^\infty 4y^2e^{-y^2}\, dy = 2\int_0^\infty e^{-y^2}\, dy - 2ye^{-y^2}\Big|_0^\infty = 2\cdot\sqrt{2\pi}\cdot\left(\frac{1}{\sqrt{2}}\right)\int_0^\infty \frac{1}{\sqrt{2\pi}\cdot\left(\frac{1}{\sqrt{2}}\right)} e^{-\frac{y^2}{2\cdot\left(\frac{1}{\sqrt{2}}\right)^2}}\, dy = \sqrt{\pi},$$

since $\frac{dye^{-y^2}}{dx} = e^{-y^2} - 2y^2e^{-y^2}$

Application in data classification

Application in data classification

- Recent development in data classification focuses on the support vector machines (SVM), due to accuracy concern
- RVKDE
  - delivers the same level of accuracy as the SVM and enjoys some advantages
  - constructs one approximate probability density function for one class of objects

Time complexity

- The average time complexity to construct a classifier is $O(dn\log n + k'n\log n)$, if the k-d tree structure is employed, where n is the number of training samples
- The time complexity to classify m new objects with unknown class is $O(dn\log n + k''m\log n)$

Comparison of classification accuracy

	Classification algorithms			
Data sets	**RVKDE**	**SVM**	**1NN**	**3NN**
1. iris (150)	97.33 $(k' = 24, k'' = 14, d' = 5, \beta = 0.7)$	97.33	94.00	94.67
2. wine (178)	99.44 $(k' = 3, k'' = 16, d' = 1, \beta = 0.7)$	99.44	96.08	94.97
3. vowel (528)	99.62 $(k' = 15, k'' = 1, d' = 1, \beta = 0.7)$	99.05	99.43	97.16
4. segment (2310)	97.27 $(k' = 25, k'' = 1, d' = 1, \beta = 0.7)$	97.40	96.84	95.98
Avg. 1-4	**98.42**	**98.31**	**96.59**	**95.70**
5. glass (214)	75.74 $(k' = 9, k'' = 3, d' = 2, \beta = 0.7)$	71.50	69.65	72.45
6. vehicle (846)	73.53 $(k' = 13, k'' = 8, d' = 2, \beta = 0.7)$	86.64	70.45	71.98
Avg. 1-6	**90.49**	**91.89**	**87.74**	**87.87**

Comparison of classification accuracy

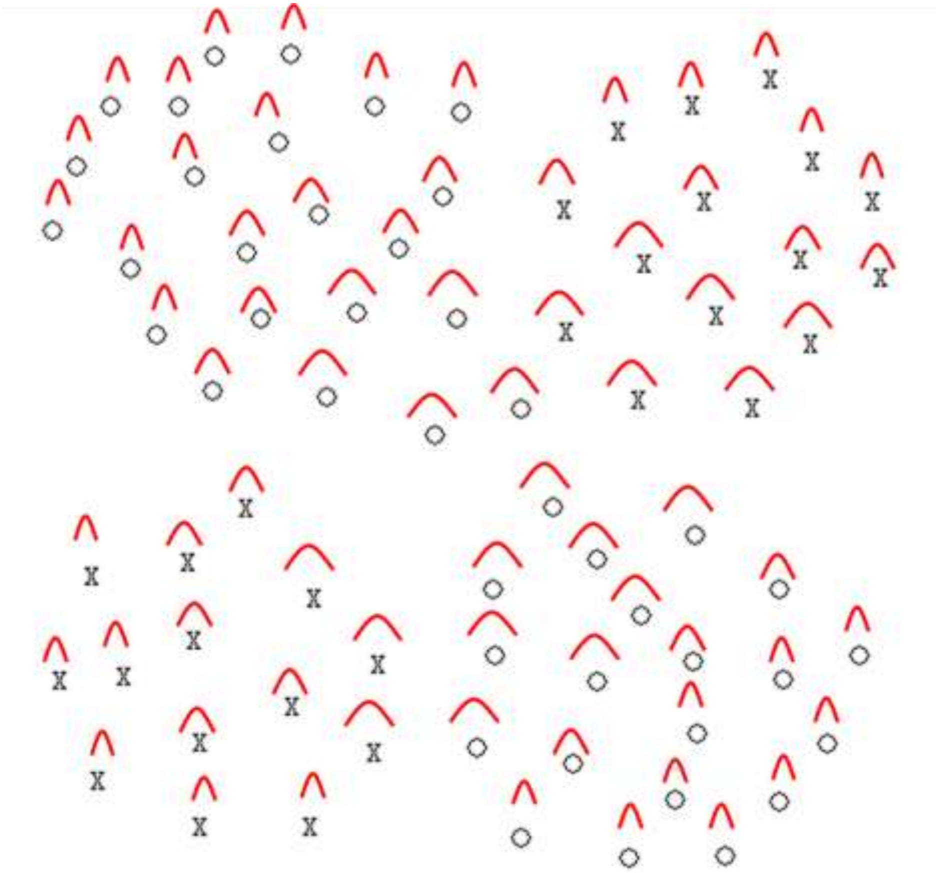
3 larger data sets

	Classification algorithms			
Data sets	**RVKDE**	**SVM**	**1NN**	**3NN**
7. satimage (4435,2000)	92.30 ($k^{'} = 6, k^{''} = 26, d^{'} = 1, \beta = 0.7$)	91.30	89.35	90.60
8. letter (15000,5000)	97.12 ($k^{'} = 28, k^{''} = 28, d^{'} = 2, \beta = 0.7$)	97.98	95.26	95.46
9. shuttle (43500,14500)	99.94 ($k^{'} = 18, k^{''} = 1, d^{'} = 3, \beta = 0.7$)	99.92	99.91	99.92
Avg. 7-9	**96.45**	**96.40**	**94.84**	**95.33**

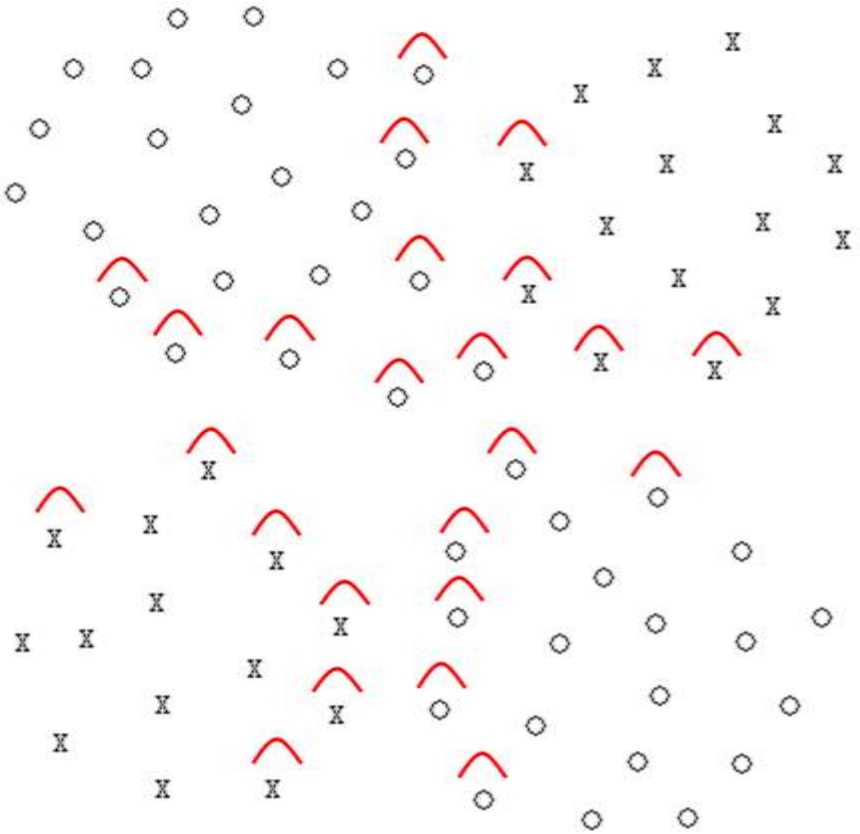
Data reduction

- As the learning algorithm is instance-based, removal of redundant training samples will lower the complexity of the prediction phase
- The effect of a naïve data reduction mechanism was studied
- The naïve mechanism removes a training sample, if all of its 10 nearest samples belong to the same class as this particular sample

With the KDE based statistical learning algorithm, each training sample is associated with a kernel function with typically a varying width



In comparison with the decision function of SVM in which only support vectors are associated with a kernel function with a fixed width



Effect of data reduction

	satimage	letter	shuttle
# of training samples in the original data set	4435	15000	43500
# of training samples after data reduction is applied	1815	7794	627
% of training samples remaining	40.92%	51.96%	1.44%
Classification accuracy after data reduction is applied	92.15%	96.18%	99.32%
Degradation of accuracy due to data reduction	-0.15%	-0.94%	-0.62%

Effect of data reduction

	# of training samples after data reduction is applied	# of support vectors identified by LIBSVM
satimage	1815	1689
letter	7794	8931
shuttle	627	287

Execution times (in seconds)

		Proposed algorithm without data reduction	Proposed algorithm with data reduction	SVM
Cross validation	satimage	670	265	64622
	letter	2825	1724	386814
	shuttle	96795	59.9	467825
Make classifier	satimage	5.91	0.85	21.66
	letter	17.05	6.48	282.05
	shuttle	1745	0.69	129.84
Test	satimage	21.3	7.4	11.53
	letter	128.6	51.74	94.91
	shuttle	996.1	5.58	2.13

Today's exercise

Regression

Use the regression mode of RVKDE to generate transaction decisions. Upload, test and commit them in our stock system. Send TA Lin a report before 23:59 11/20 (Wed).