Kernel Estimator (Perzen windows)

$$\hat{p}(\underline{x}) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\boxed{\overset{u \in \mathbb{R}}{\frac{x - x_i}{h}}}\right) \qquad x \in \mathbb{R} \qquad \{x_i \in \mathbb{R}\}_{i=1}^{N}$$

↳ kernel function

$$K : \mathbb{R} \to \mathbb{R} \qquad K(u) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{u^2}{2}\right] \Rightarrow \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$\mu = 0 \quad \sigma = 1$$

Generalization to Multivariate Data

$$\hat{p}(x) = \frac{1}{Nh^{D}} \sum_{i=1}^{N} K\left(\boxed{\overset{u \in \mathbb{R}^D}{\frac{x - x_i}{h}}}\right) \qquad x \in \mathbb{R}^D \qquad \{x_i \in \mathbb{R}^D\}_{i=1}^{N}$$

$$\hookrightarrow \frac{1}{\sqrt{(2\pi)^D |\Sigma|}} \cdot \exp\left[-\frac{1}{2}(x-\mu)^T \Sigma^{-1} \cdot (x-\mu)\right]$$

$$K : \mathbb{R}^D \to \mathbb{R}$$

$$\mu = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & \ddots & \\ & & & 1 \end{bmatrix} \Rightarrow K(u) = \frac{1}{\sqrt{(2\pi)^D}} \cdot \exp\left[-\frac{u^T u}{2}\right]$$

$$S = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$$

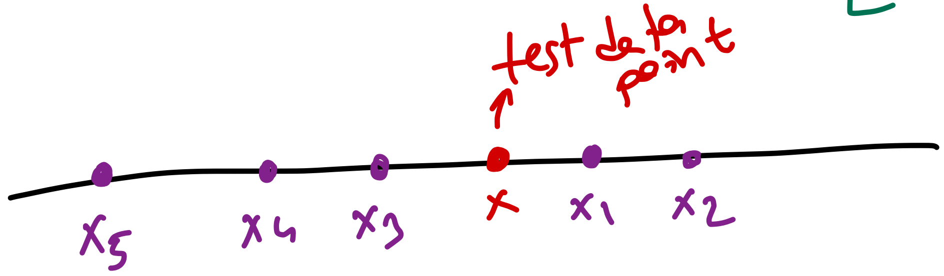
$$\Downarrow$$

$$\mu = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \Sigma = S \Rightarrow K(u) = \frac{1}{\sqrt{(2\pi)^D |S|}} \cdot \exp\left[-\frac{1}{2} \cdot u^T \cdot \underline{\underline{S^{-1}}} \cdot u\right]$$

$$S^{-1} = \begin{bmatrix} 1 & 0 \\ 0 & 1/4 \end{bmatrix}$$

$$\begin{bmatrix} u_1 & u_2 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1/4 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = u_1^2 + \frac{1}{4} u_2^2$$

$$\underbrace{\begin{bmatrix} u_1 & \frac{1}{4} u_2 \end{bmatrix}} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = u_1^2 + \frac{1}{4} u_2^2$$



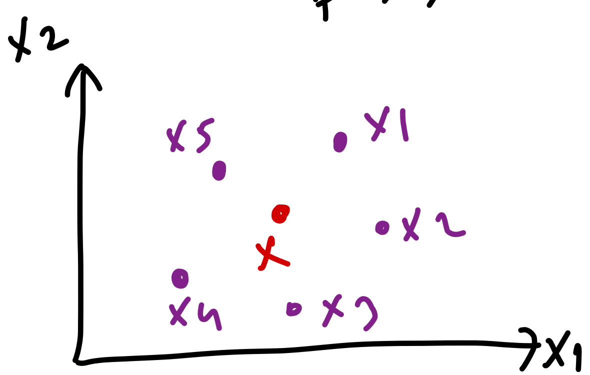
$$\hat{p}(x) = \frac{1}{Nh} \cdot \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right) = \frac{1}{Nh} \cdot \sum_{i=1}^{N} K(u_i)$$

$$\frac{x - x_1}{h} = u_1, \quad \frac{x - x_2}{h} = u_2, \quad ---, \quad \frac{x - x_5}{h} = u_5 \Rightarrow \frac{1}{h} \cdot dx = du$$

$$dx = h \cdot du$$

$$\int_{-\infty}^{+\infty} \hat{p}(x)\, dx = 1 \Rightarrow \int_{-\infty}^{+\infty} \dots\, du = 1$$

$$\hat{p}(x) \geqslant 0 \quad \forall x$$



$$\hat{p}(x) = \frac{1}{Nh^2} \sum_{i=1}^{N} K\Big(\underbrace{\frac{x - x_i}{h}}^{u_i}\Big)$$

$$dx_1 = h \cdot du_1$$
$$dx_2 = h \cdot du_2$$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \hat{p}(x)\, dx_1\, dx_2 = 1 \Rightarrow \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \dots\dots du_1\, du_2 = 1$$

$$\hat{p}(x) \geqslant 0 \quad \forall x$$

NONPARAMETRIC CLASSIFICATION

$$\underbrace{\hat{p}(x \mid y=c)}_{\text{class conditional density estimation}} = \frac{1}{N_c h^D} \sum_{i=1}^{N} \left[K\left(\frac{x-x_i}{h}\right) y_{ic} \right]$$

$y_{ic} \rightarrow \mathbb{1}(y_i = c)$

$N_c \rightarrow$ # of data points from class c $\Big\}\ \sum_{i=1}^{N} \mathbb{1}(y_i = c)$

N_c = # of data points from class # c

$c = 1, 2, \ldots, K$

N = # of data points

$N = N_1 + N_2 + \cdots + N_K$

$$y_{ic} = \begin{cases} 1 & \text{if } y_i = c \\ 0 & \text{otherwise} \end{cases}$$

$$g_c(x) \Rightarrow \hat{P}(y=c \mid x) = \frac{\hat{p}(x \mid y=c)\,\underbrace{\hat{P}(y=c)}_{N_c/N}}{\cancel{\hat{p}(x)}}$$

constant for all "c"

$$g_c(x) \propto \frac{1}{\cancel{N_c} h^D} \cdot \sum_{i=1}^{N} \left[K\left(\frac{x-x_i}{h}\right) \cdot y_{ic} \right] \cdot \frac{\cancel{N_c}}{N}$$

$$\propto \frac{1}{\cancel{N h^D}} \sum_{i=1}^{N} \left[K\left(\frac{x-x_i}{h}\right) \cdot y_{ic} \right]$$

constant for all "c"

$$g_c(x) \propto \sum_{i=1}^{N} \left[K\left(\frac{x - x_i}{h}\right) \cdot y_{ic} \right]$$

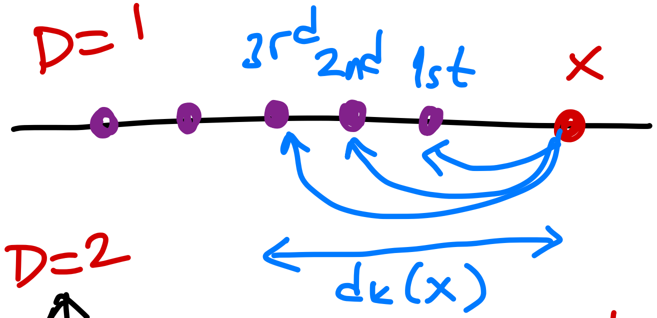
800 ⇒ class 1
100 ⇒ class 2
100 ⇒ class 3

(1) Calculate $g_1(x), g_2(x), \ldots, g_K(x)$

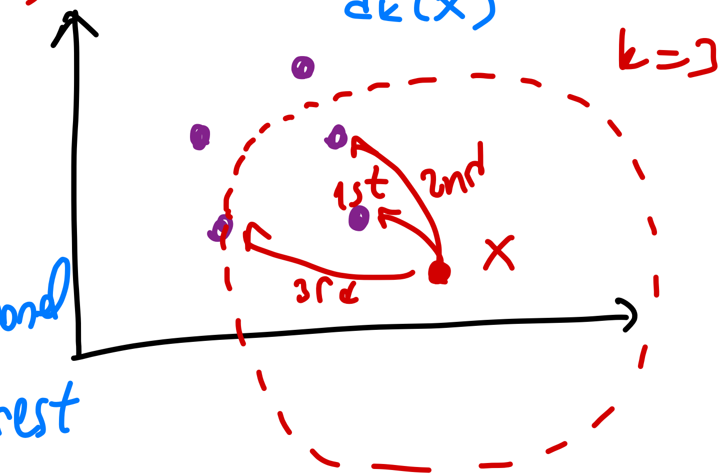
(2) Pick the maximum value.

k-Nearest Neighbor Estimator

$$\hat{p}(x) = \frac{k}{N \, 2 d_k(x)} \qquad x \in \mathbb{R}$$

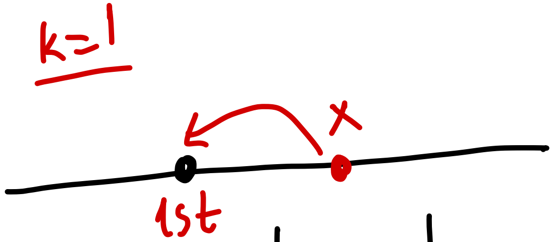


$$\hat{p}(x) = \frac{k}{N \underbrace{V_k(x)}} \qquad x \in \mathbb{R}^D$$



Volume of smallest D-dimensional hypersphere that covers k-nearest neighbors.

D=2 πR^2
D=3 $\frac{4}{3}\pi R^3$



$|x - x_1|$ ⇐ distance ⇒ $d_k(x)$

$$\int_{-\infty}^{+\infty} \hat{p}(x)\,dx \overset{?}{=} 1 \Rightarrow \int_{-\infty}^{+\infty} \frac{1}{1 \cdot 2 \cdot |x - x_1|}\,dx \neq 1$$

$$\hat{p}(y=c \mid x) = \frac{\hat{p}(x \mid y=c) \cdot \hat{P}(y=c)}{\hat{p}(x)} \qquad \hat{p}(x) = \sum_{c=1}^{K} \hat{p}(x \mid y=c)\hat{P}(y=c)$$

$$= \frac{\dfrac{k_c}{\cancel{N_c} \cancel{V_k(x)}} \cdot \dfrac{\cancel{N_c}}{\cancel{N}}}{\sum_{d=1}^{K} \left[\dfrac{k_d}{\cancel{N_d} \cancel{V_k(x)}} \cdot \dfrac{\cancel{N_d}}{\cancel{N}} \right]}$$

$$= \frac{k_c}{\sum_{d=1}^{K} k_d}$$

k_c → # of neighbors from class # c

$\sum_{d=1}^{K} k_d$ → total # of neighbors (k)

$$\frac{k_1}{k} + \frac{k_2}{k} + \cdots + \frac{k_K}{k} = 1$$

if $k = N \Rightarrow \left.\begin{array}{l} k_1 = N_1 \\ k_2 = N_2 \\ \vdots \\ k_K = N_K \end{array}\right\} \Rightarrow \dfrac{k_c}{k} = \dfrac{N_c}{N}$

Distance-Based Classification

posteriors

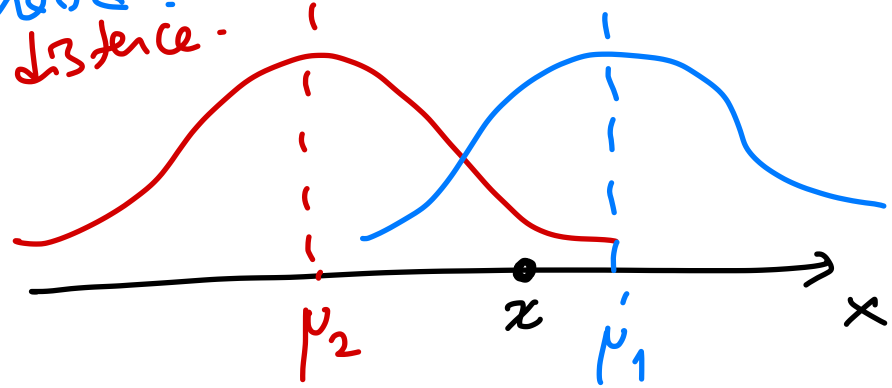
	1*	2	3
$x \Rightarrow$	0.80	0.20	0.00

	1*	2	3
$x \Rightarrow$	0.75	0.15	0.10

$\Rightarrow$ assign a data point to a class, which is heavily represented in its neighborhood.

$$c^* = \underbrace{\arg\min_{d=1}^{K} D(x, \mu_d)}_{\text{nearest mean classifier}}$$

→ picking minimum distance.



$$P(y=1|x) \propto \cancel{\frac{1}{\sqrt{2\pi\sigma_1^2}}} \exp\left[-\frac{(x-\mu_1)^2}{\cancel{2\sigma_1^2}}\right] \cdot \cancel{\frac{N_1}{N}}$$

$$P(y=2|x) \propto \cancel{\frac{1}{\sqrt{2\pi\sigma_2^2}}} \cdot \exp\left[-\frac{(x-\mu_2)^2}{\cancel{2\sigma_2^2}}\right] \cdot \cancel{\frac{N_2}{N}}$$

$$\left.\begin{array}{l} \sigma_1^2 = \sigma_2^2 \\ N_1 = N_2 \end{array}\right\} \text{assumptions}$$

picking the maximum posterior is equivalent to picking smallest distance.