Kernel Estimator (Parzen Windows)

$$\mathcal{X} = \{x_i\}_{i=1}^{N} \quad x_i \in \mathbb{R}$$

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

$\hookrightarrow$ kernel function

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

$$\frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

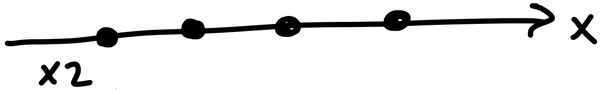
$$\Downarrow$$

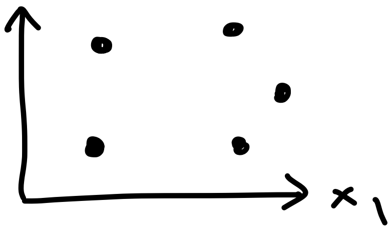
$$x = u, \ \mu = 0, \ \sigma^2 = 1$$

Generalization to Multivariate Data

$$\mathcal{X} = \{x_i\}_{i=1}^{N} \quad x_i \in \mathbb{R}^D$$

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





$$\hookrightarrow K(u) = \frac{1}{\sqrt{(2\pi)^D}} \exp\left[-\frac{u^T \cdot u}{2}\right]$$

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

$$\hookrightarrow K(u) = \frac{1}{\sqrt{(2\pi)^D |S|}} \cdot \exp\left[-\frac{1}{2} u^T . S^{-1} . u\right]$$

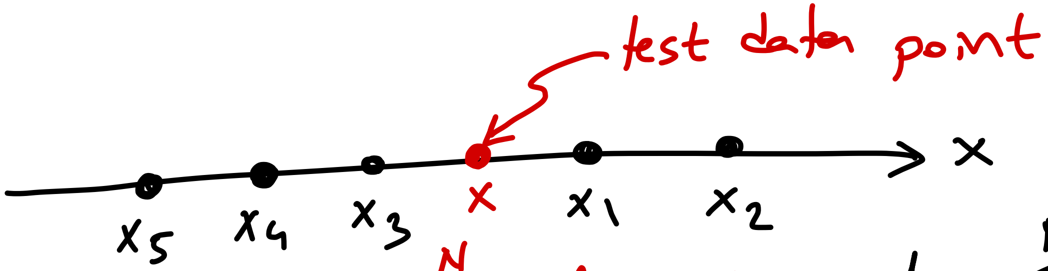
$$\frac{1}{\sqrt{(2\pi)^D |\Sigma|}} \cdot \exp\left[-\frac{1}{2} (x-\mu)^T . \Sigma^{-1} . (x-\mu)\right]$$

$$x = u, \ \mu = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ \Sigma = \begin{bmatrix} 1 & & 0 \\ & 1 & \\ 0 & & \ddots 1 \end{bmatrix}$$

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

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

replace u_2 with $\frac{u_2}{2}$



$$\frac{1}{h}\cdot dx = du$$

$$dx = h\cdot du$$

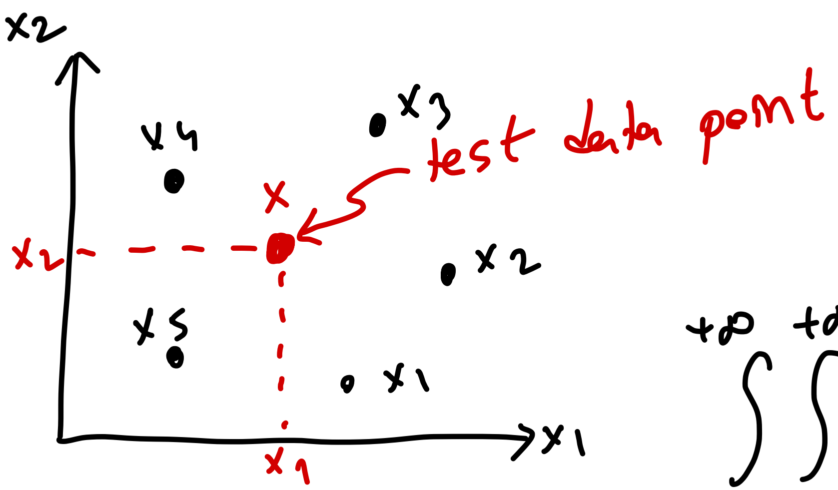
$$\hat{p}(x) = \frac{1}{Nh}\cdot\sum_{i=1}^{N} K\left(\frac{x-x_i}{h}\right) = \frac{1}{Nh}\cdot\sum_{i=1}^{N} K(u_i)$$

$$u_1 = \frac{x-x_1}{h} \qquad u_2 = \frac{x-x_2}{h} \qquad \cdots \qquad u_5 = \frac{x-x_5}{h}$$

$$\int_{-\infty}^{+\infty}\hat{p}(x)\,dx = 1 \implies \int_{-\infty}^{+\infty}\frac{1}{N\cancel{h}}\cdot\sum_{i=1}^{N}K(u_i)\cdot\cancel{h}\cdot du = \sum_{i=1}^{N}\underbrace{\left[\int_{-\infty}^{+\infty}K(u_i)\,du\right]}_{1}\frac{1}{N}$$

$$= 1$$

$$\hat{p}(x) \geq 0 \quad \forall x \;\checkmark$$



$$\hat{p}(x) = \frac{1}{N.h^2}\cdot\sum_{i=1}^{N}K\left(\left(\frac{x-x_i}{h}\right)\right)$$

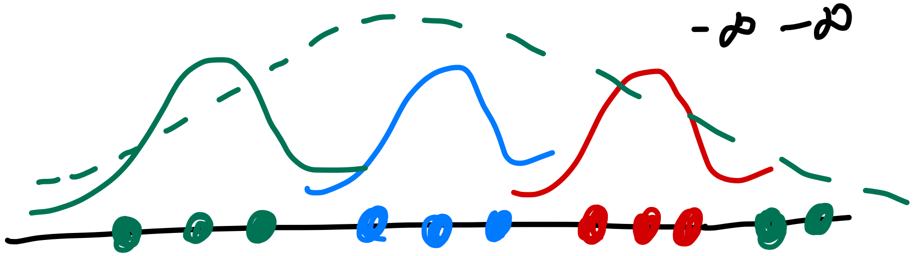
$$dx_1 = h.du_1 \Leftarrow \frac{x_1 - x_{i1}}{h} = u_1$$

$$dx_2 = h.du_2 \Leftarrow \frac{x_2 - x_{i2}}{h} = u_2$$

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\hat{p}(x)\,dx_1\,dx_2 = 1 \implies \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\frac{1}{N\cancel{h^2}}\sum_{i=1}^{N}K\left(\frac{x-x_i}{h}\right).\cancel{h}.du_1\cancel{h}.du_2$$

$$= 1$$

$$\hat{p}(x) \geq 0 \quad \forall x \;\checkmark$$



NONPARAMETRIC CLASSIFICATION

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

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

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

$c = 1, 2, \ldots, K$ $\qquad N_c = $ # of data points from class c

$N = $ # of data points $\qquad N = N_1 + N_2 + \ldots + N_K$

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

$$g_c(x) \Rightarrow \hat{\Pr}(y = c \mid x) = \frac{\hat{p}(x \mid y=c)\, \hat{\Pr}(y=c)}{\hat{p}(x)}$$

$\hat{\Pr}(y=c) \rightarrow$ frequency of class c: $\frac{N_c}{N}$

$\hat{p}(x) \rightarrow$ constant for all c

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

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

N: constant for all c

$\rightarrow$ score function for class c

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

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

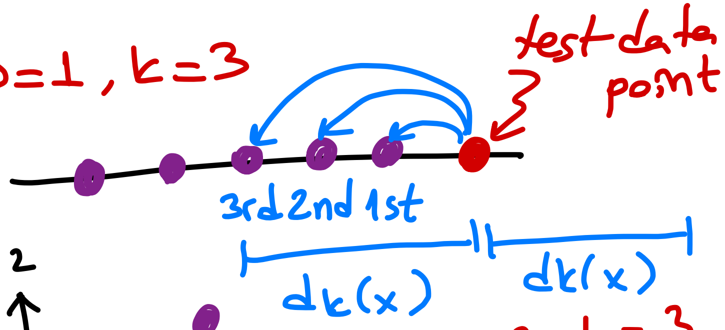
② Pick the maximum value.

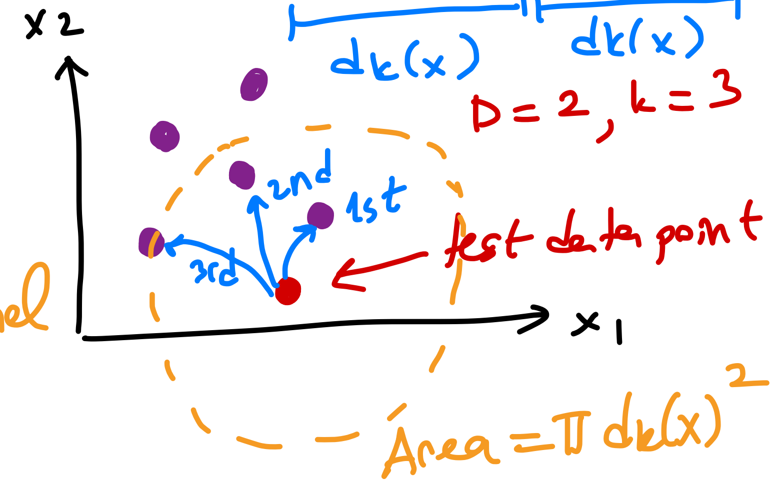
k-Nearest Neighbor Estimator

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

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

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

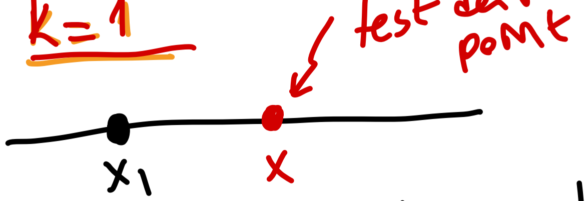




D=3

$$\text{Volume} = \frac{4}{3}\pi\, d_k(x)^3$$

k=1

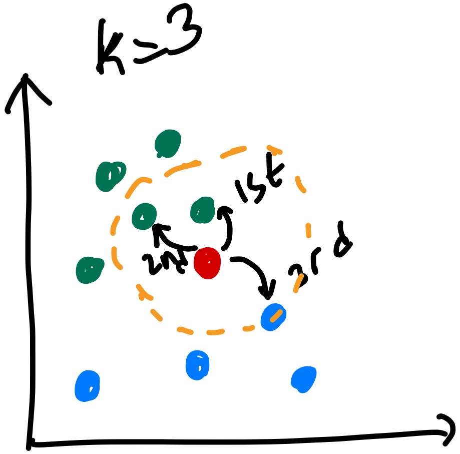


$$|x - x_1| \leftarrow d_k(x)$$

$$\int_{-\infty}^{+\infty} \hat{p}(x)\,dx = 1 \Rightarrow \int_{-\infty}^{+\infty} \frac{1}{1.2.|x-x_1|}\,dx \neq 1$$

$$\hat{\Pr}(y=c \mid x) = \frac{\hat{p}(x \mid y=c) \cdot \hat{\Pr}(y=c)}{\hat{p}(x)}$$

$$\hat{p}(x) = \sum_{c=1}^{K} \hat{p}(x \mid y=c)\hat{\Pr}(y=c)$$



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

$$\hat{\Pr}(\bullet \mid x) = \frac{2}{3}$$

$$\hat{\Pr}(\bullet \mid x) = \frac{1}{3}$$

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

k_c → # of neighbors from class c

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

$$= \frac{k_c}{k}$$

$$\sqrt{(x_{i1} - x_1)^2 + \ldots + (x_{iD} - x_D)^2}$$

$$w_1^T \cdot x + w_{10}$$

Distance-Based Classification

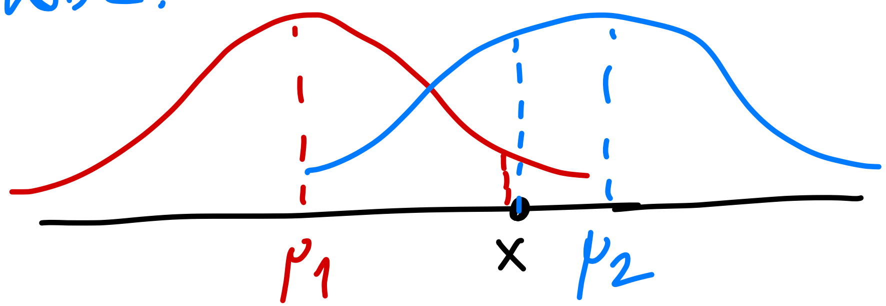
posteriors

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

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

⇒ assign a data point to a class, which is heavily represented in its neighborhood.

$$c^* = \arg\min_{d=1}^{K} D(x, \mu_d)$$



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

$$\sigma_1^2 = \sigma_2^2$$
$$N_1 = N_2$$

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

picking the maximum posterior is equivalent to picking the smallest distance.