Kernel Estimator (PARZEN WINDOWS)

$x \in \mathbb{R}$

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

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

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

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

Generalization to Multivariate Data

$x \in \mathbb{R}^D \quad x_i \in \mathbb{R}^D$

$$\hat{p}(x) = \frac{1}{N h^{\boxed{D}}} \sum_{i=1}^{N} K\Big(\underbrace{\frac{\boxed{x} - \boxed{x_i}}{h}}_{u}\Big)$$

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

$\Sigma = I \qquad \mu = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}$

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

$$\begin{array}{c|ccc} & x_1 & x_2 & x_3 \\ \hline x_1 & 1 & 0 & 0 \\ x_2 & 0 & 1 & 0 \\ x_3 & 0 & 0 & 1 \end{array}$$

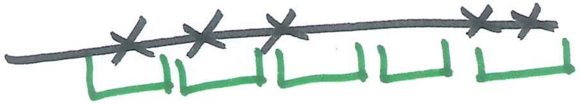
$$K(u) = \left(\frac{1}{\sqrt{(2\pi)^D \cdot |S|}}\right) \cdot \exp\left[-\frac{1}{2} u^T S^{-1} \cdot u\right]$$

$\Sigma = S$

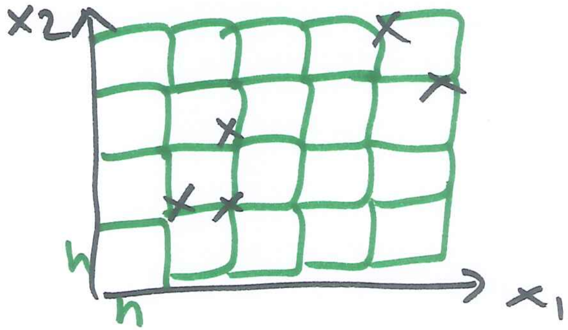
$u = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$

$\underbrace{|S|}$ → determinant of S

$D = 1$



$D = 2$



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

$$\int_{x_2=-\infty}^{+\infty} \int_{x_1=-\infty}^{+\infty} \underset{\substack{\downarrow \\ \text{2-dimensional vector.}}}{\hat{p}(x)}\, dx_1\, dx_2 = 1$$

$\hat{p}(x)$

Nonparametric Classification

$$\hat{p}(x \mid y=c) = \frac{1}{N_c h^D} \sum_{i=1}^{N} \Big[K\Big(\frac{x-x_i}{h}\Big) \cdot \underset{\substack{\uparrow \\ 1(y_i=c)}}{y_{ic}} \Big]$$

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

N = # of data points

N_c = # of data points in class c.

$N = N_1 + N_2 + \ldots + N_K$.

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

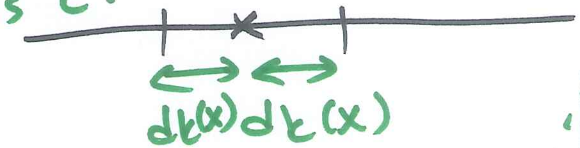
$$g_c(x) = \frac{\hat{p}(x \mid y=c)\ \overbrace{\hat{p}(y=c)}^{N_c/N}}{\underbrace{\cancel{\hat{p}(x)}}_{\to \text{ constant for all } c}} \boxed{= \hat{p}(y=c \mid x)}$$

$$g_c(x) = \overbrace{\frac{1}{\cancel{N_c} h^D} \cdot \sum_{i=1}^{N} \Big[K\Big(\frac{x-x_i}{h}\Big) \cdot y_{ic} \Big]} \cdot \frac{\cancel{N_c}}{N} = \underbrace{\boxed{\frac{1}{N \cdot h^D}}}_{\hookrightarrow \text{ constant for all } c.} \cdot \sum_{i=1}^{N} \Big[K\Big(\frac{x-x_i}{h}\Big) \cdot y_{ic} \Big]$$

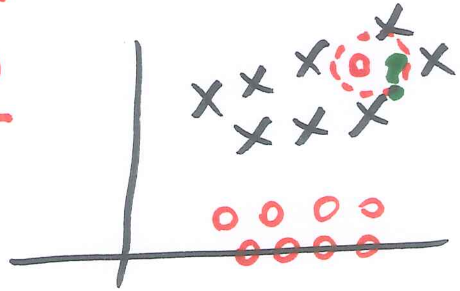
$g_1(x), g_2(x), \ldots, g_K(x) \Rightarrow$ pick maximum value.

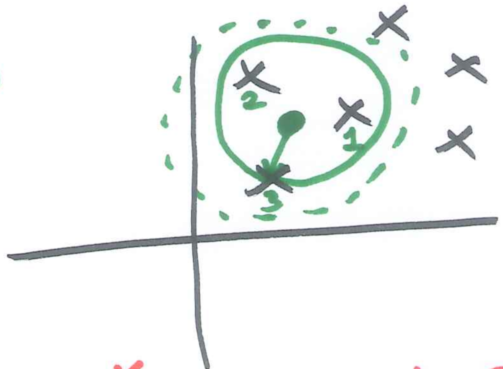
$$\hat{p}(x \mid y=c) = \frac{k_c}{N_c \cdot \underbrace{V_k(x)}_{2 d_k(x)}}$$

$k_c \rightarrow$ # of neighbors from class c.



$V_k(x)$: smallest Volume of D-dimensional hypersphere that covers k nearest neighbors.



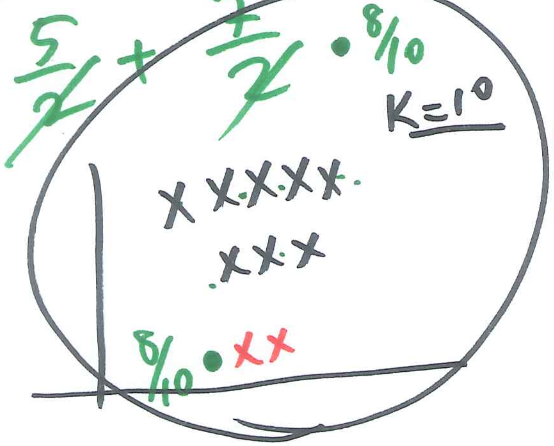


$$p(x) = \sum_{c=1}^{K} p(x \mid y=c) \cdot p(y=c)$$

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

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

$$\frac{\frac{1}{\cancel{2}}}{\frac{5}{\cancel{2}} + \frac{2}{\cancel{2}}}$$

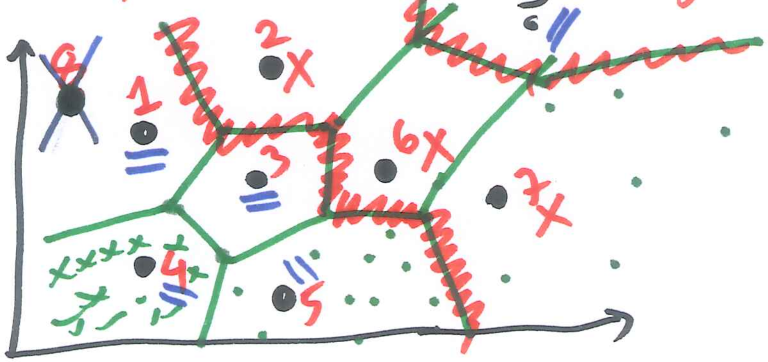


$\frac{k_1}{k}$ $\frac{k_2}{k}$ . . . $\frac{k_K}{k}$

$K \rightarrow$ # of classes

$k \rightarrow$ # of neighbors.

$k=1 \Rightarrow$ (one-nearest neighbor) "Voronoi tesselation"



Condensed Nearest Neighbor

$\underline{\underline{Z}} \leftarrow \underline{\underline{\emptyset}}$ ← condensed neighbors.

Repeat

for all $x \in \underline{\underline{\mathcal{X}}}$ (in random order)

find $x' \in Z$ such that

$\|x - x'\|$ is minimum

if $\underline{class(x) \neq class(x')}$ add x to Z

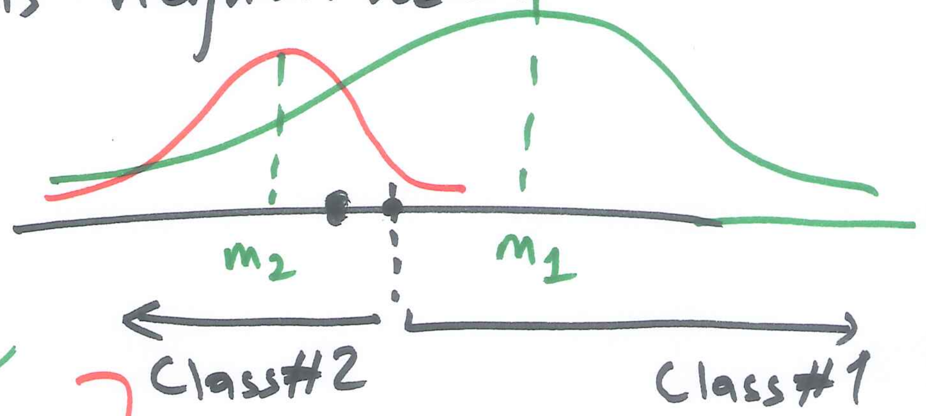
Until Z does not change.

$\mathcal{X} \leftarrow Z$

Distance-Based Classification

$\frac{1}{0.8} \quad \frac{2}{0.2} \quad \frac{3}{0.0}$ vs $\frac{1}{0.75} \quad \frac{2}{0.15} \quad \frac{3}{0.10}$

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



$$D(\underline{x}, \underline{m_c}) = \min_{d=1}^{k} D(\underline{x}, \underline{m_d})$$

Nearest mean classifiers.

$$\cancel{\frac{1}{\sqrt{(2\pi)^D |\Sigma_c|}}} \cdot \exp\left[- \frac{(x-\mu_c)^T \cdot \cancel{\Sigma_c^{-1}} (x-\mu_c)}{2} \right]$$

$$\Sigma_1 = \Sigma_2 = \Sigma_3 = \dots = \Sigma_k = \Sigma$$

$$\min \; (x-\mu_c)^T \cdot (x-\mu_c) = \|x-\mu_c\|_2^2$$

$\|\;\|_1 \to \ell_1$-norm $\quad \|x_1 - x_2\|_1 = \sum_{d=1}^{D} |x_{1d} - x_{2d}|$

$$\|x_1 - x_2\|_2 = \sqrt{\sum_{d=1}^{D} (x_{1d} - x_{2d})^2}$$

$\sqrt{4+9}$

$\|\;\|_2 \to \ell_2$-norm

$\sqrt{4} + \sqrt{9}$

$\|\;\|_p \to \ell_p$-norm

$D(x, m_1) = 2 \Rightarrow 4$

$D(x, m_2) = 3 \Rightarrow 9$

$$D(x, m_c) = \| x - m_c \|_2^2 = (x - m_c)^T (x - m_c)$$

$$D(x, m_c) = (x - m_c)^T \cdot S_c^{-1} \cdot (x - m_c)$$ ← Mahalanobis Distance.

$m_1, m_2, \ldots, m_k$ // $m_1, m_2, \ldots, m_k$

$S_1, S_2, \ldots, S_k$