Kernel Estimator (Parzen windows)

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

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

$\hookrightarrow$ kernel function

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

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

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

$\mu = 0 \quad \sigma = 1 \Rightarrow$ standard normal distr.

Generalization to Multivariate Data

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

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

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

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

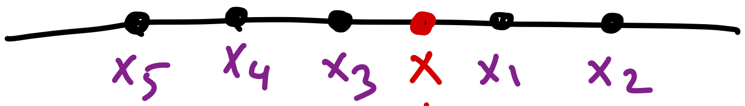
$$K(u) = \frac{1}{\sqrt{(2\pi)^D}} \exp\left[-\frac{u^T \cdot u}{2}\right]$$

$\mu = 0 \quad \Sigma = I$

$$\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1/4 \end{bmatrix}$$

$\mu = 0 \quad \Sigma = S$

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

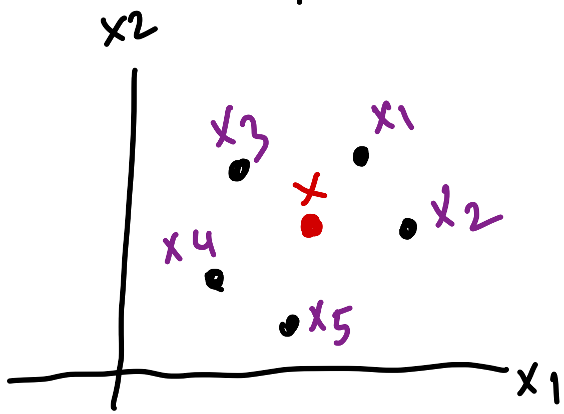


$$\hat{p}(x) = \frac{1}{Nh} \cdot \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right) = \frac{1}{Nh} \cdot \sum_{i=1}^{N} K(u_i)$$

$$\frac{x - x_1}{h} = u_1 \;,\; ----\;,\; \frac{x - x_5}{h} = u_5 \;\Rightarrow\; \frac{1}{h} dx = du \;\Rightarrow\; \underline{dx = h \cdot du}$$

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

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



$$\hat{p}(x) = \frac{1}{Nh^2} \cdot \sum_{i=1}^{N} K\Bigg(\boxed{\overset{u_i}{\frac{x - x_i}{h}}}\Bigg)$$

$$= \frac{1}{Nh^2} \cdot \sum_{i=1}^{N} K(\underline{\underline{u_i}})$$

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

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

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

NON PARAMETRIC CLASSIFICATION

$$\underbrace{\hat{p}(x \mid y=c)}_{\text{class conditional density}} = \frac{1}{\underbrace{N_c}_{\substack{\#\text{ of data} \\ \text{points in class } c}} h^D} \sum_{i=1}^{N} \Big[K\Big(\frac{x - x_i}{h}\Big) \cdot \underbrace{y_{ic}}_{\to 1(y_i = c)} \Big]$$

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

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

N = # of data points

N_c = # of data points in class c

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

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

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

$\hat{P}(y=c) \Rightarrow N_c / N$

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

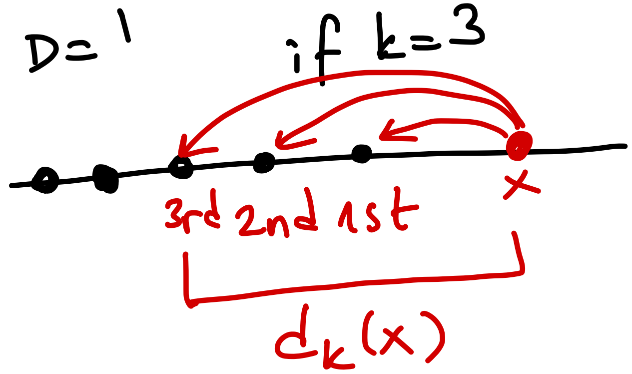
$$g_c(x) \propto \frac{1}{\cancel{N_c} h^D} \cdot \sum_{i=1}^{N} \Big[K\Big(\frac{x - x_i}{h}\Big) \cdot y_{ic} \Big] \cdot \frac{\cancel{N_c}}{N}$$

constant for all "c"

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

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

1. Calculate $g_1(x), g_2(x), \ldots, g_K(x)$
2. Pick the maximum value.

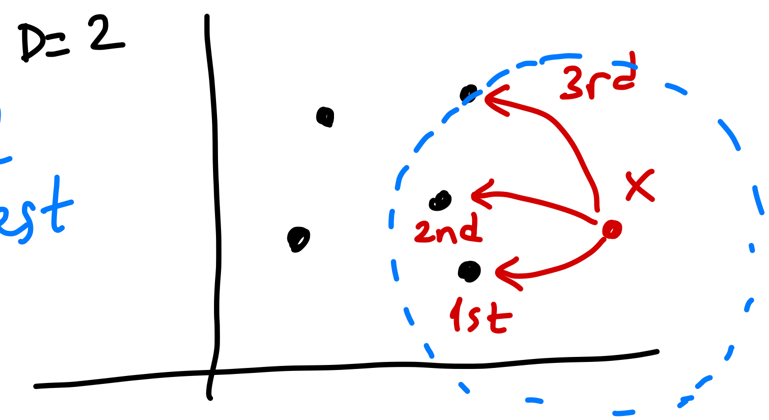


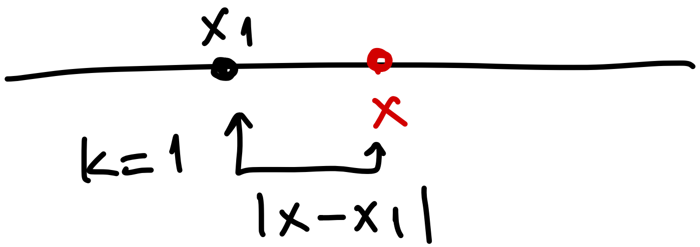
k-nearest Neighbor Estimator

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

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

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





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

$$\int_{-\infty}^{+\infty} \frac{1}{1.2\,|x-x_1|}\,dx \neq 1$$

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

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

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

$$= \frac{k_c}{\underbrace{\sum_{d=1}^{K} k_d}_{\#\text{ of neighbors} = k}}$$

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

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

if $k = N \Rightarrow$

$$k_1 = N_1$$
$$k_2 = N_2$$
$$\vdots$$
$$k_K = N_K$$

$$\frac{k_c}{k} \Rightarrow \frac{N_c}{N}$$

Distence-Based Classification

posteriors

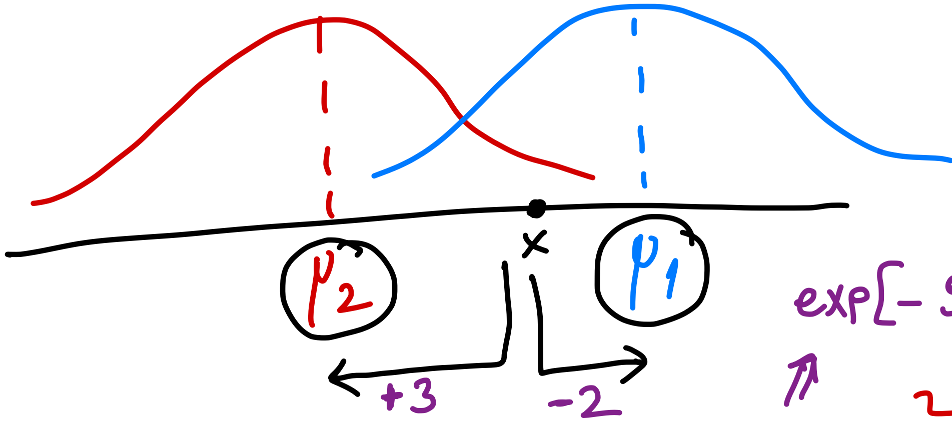
1★	2	3
0.80	0.20	0.00

1★	2	3
0.75	0.15	0.10

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

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

nearest mean classifier



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

$$\cancel{\frac{1}{\sqrt{2\pi\sigma_2^2}}} \exp\left[-\frac{(x-\mu_2)^2}{\cancel{2\sigma_2^2}}\right] \Rightarrow \exp[-9]$$

$$\cancel{\frac{1}{\sqrt{2\pi\sigma_1^2}}} \exp\left[-\frac{(x-\mu_1)^2}{\cancel{2\sigma_1^2}}\right] \Rightarrow \exp[-4]$$

$$\sigma_1 = \sigma_2$$

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

$$\Sigma_1 = \Sigma_2 = \ldots = \Sigma_K = \Sigma$$

$$\underset{d=1}{\overset{K}{\arg\max}} \exp\left[-\frac{1}{2}\overbrace{(x-\mu_d)^T}\cdot\overbrace{(x-\mu_d)}\right]$$

identical

$$\underset{d=1}{\overset{K}{\arg\min}} \underbrace{(x-\mu_d)^T\cdot}_{u^T\cdot} \underbrace{(x-\mu_d)}_{u}$$