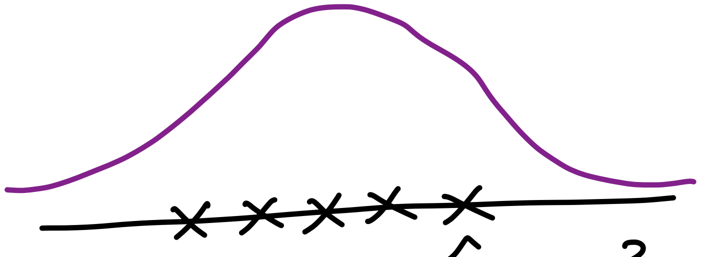
Nonparametric Methods

Linear regression $\Rightarrow$ $f(x) = w^T . x + w_0$

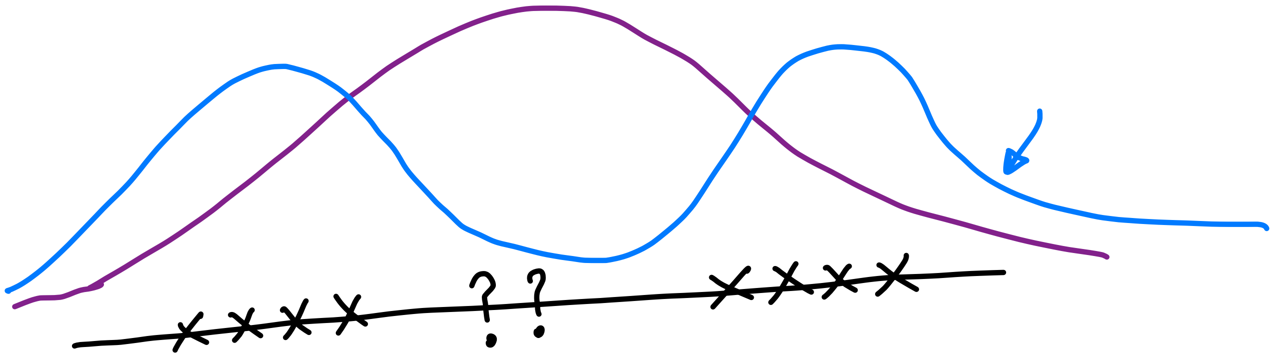
Logistic regression $\Rightarrow$ $\delta(w^T . x + x_0) = \begin{cases} 1 & \text{if } w^T . x + w_0 > 0 \\ 0 & \text{otherwise} \end{cases}$

Density estimation $\Rightarrow$ $N(x; \mu, \sigma^2)$

$N(x; \mu, \Sigma)$



$\hat{\mu} = \bar{x}, \quad \hat{\sigma}^2 = s^2$

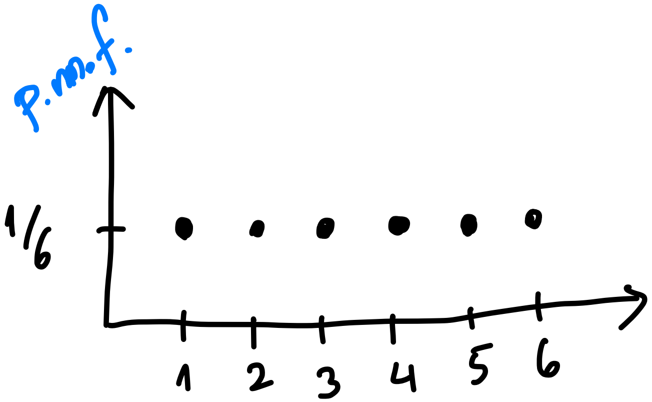


SIMILAR INPUTS $\Rightarrow$ SIMILAR OUTPUTS

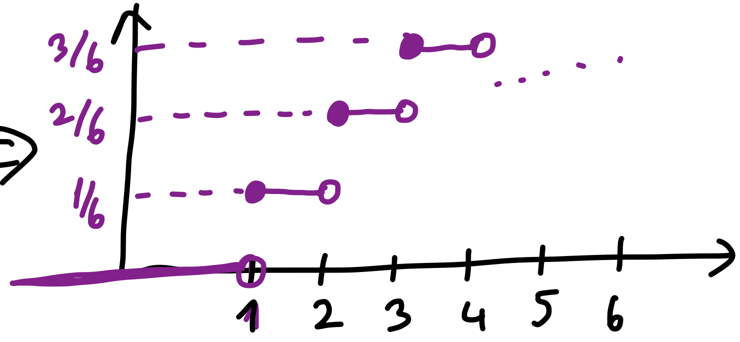
How do we measure similarity?

"data-dependent" or "local" models $\Rightarrow$ no parametric form

c.d.f. vs p.d.f. (p.m.f.)



$\Longleftrightarrow$



$F(2.0) = \frac{2}{6}$

$F(1.9) = \frac{1}{6}$

$F(3.0) = \frac{3}{6}$

$$\underbrace{F(x=a)}_{\text{c.d.f.}} = \int_{-\infty}^{a} \underbrace{p(x)}_{\text{p.d.f.}}\, dx$$

$\Uparrow$

Continuous R.V.

$$F(x=a) = P(X \leq a) = \sum_{-\infty}^{a} P(X=a)$$

$\Uparrow$

Discrete R.V

counting function

$$\hat{F}(x) = \frac{\#\{x_i \leq x\}}{N}$$

estimate for c.d.f.

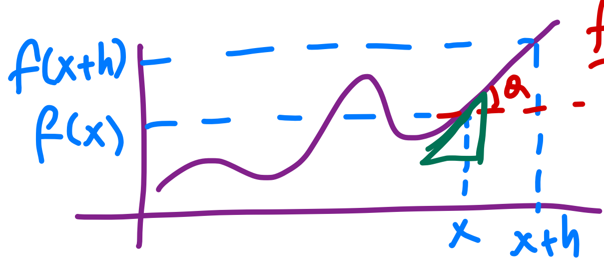
$$\hat{p}(x) = \frac{1}{h}\left[\hat{F}(x+h) - \hat{F}(x)\right]$$

$$= \frac{1}{h}\left[\frac{\#\{x_i \leq x+h\} - \#\{x_i \leq x\}}{N}\right]$$

estimate for p.d.f.

$$\#\{x_i \leq x\} = \sum_{i=1}^{N} 1(x_i \leq x)$$

$$\lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$



$$\frac{f(x+h) - f(x)}{x+h - x}$$

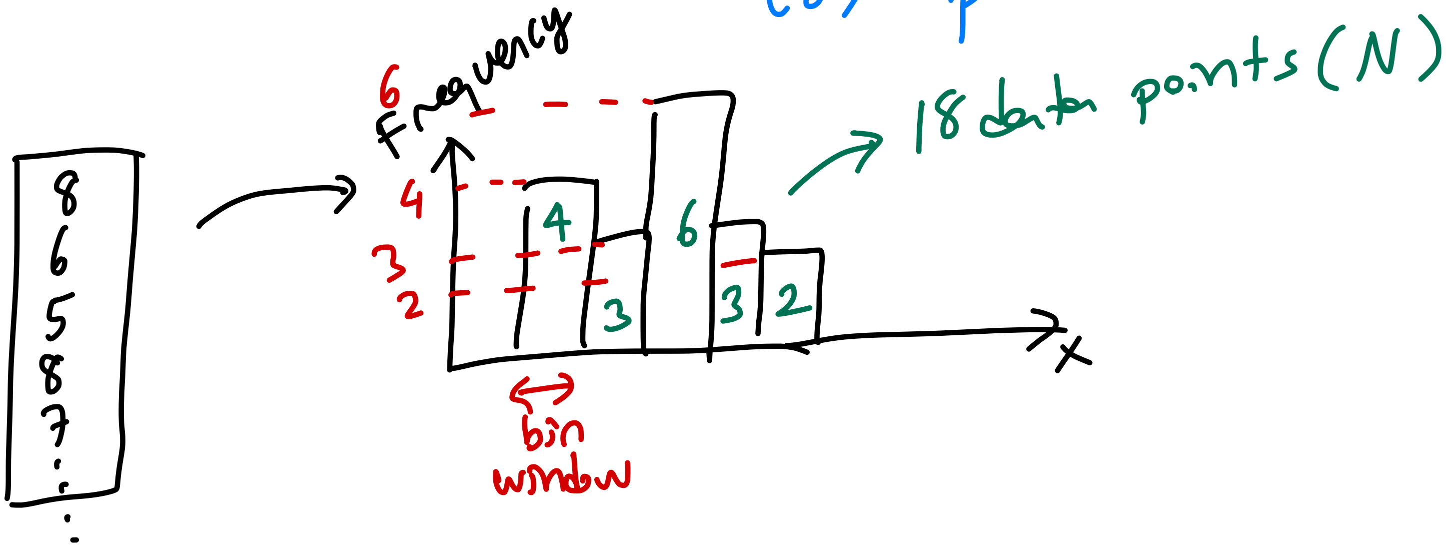
$p(x) \Rightarrow$ I would like to check whether $p(x)$ is a valid density function or not.

if x is a discrete R.V. $\Rightarrow$ i) $\sum_{-\infty}^{+\infty} P(X=x) = 1$

ii) $P(X=x) \geq 0 \quad \forall x$

if x is a continuous R.V. $\Rightarrow$ i) $\int_{-\infty}^{+\infty} p(x)\, dx = 1$

ii) $p(x) \geq 0 \quad \forall x$

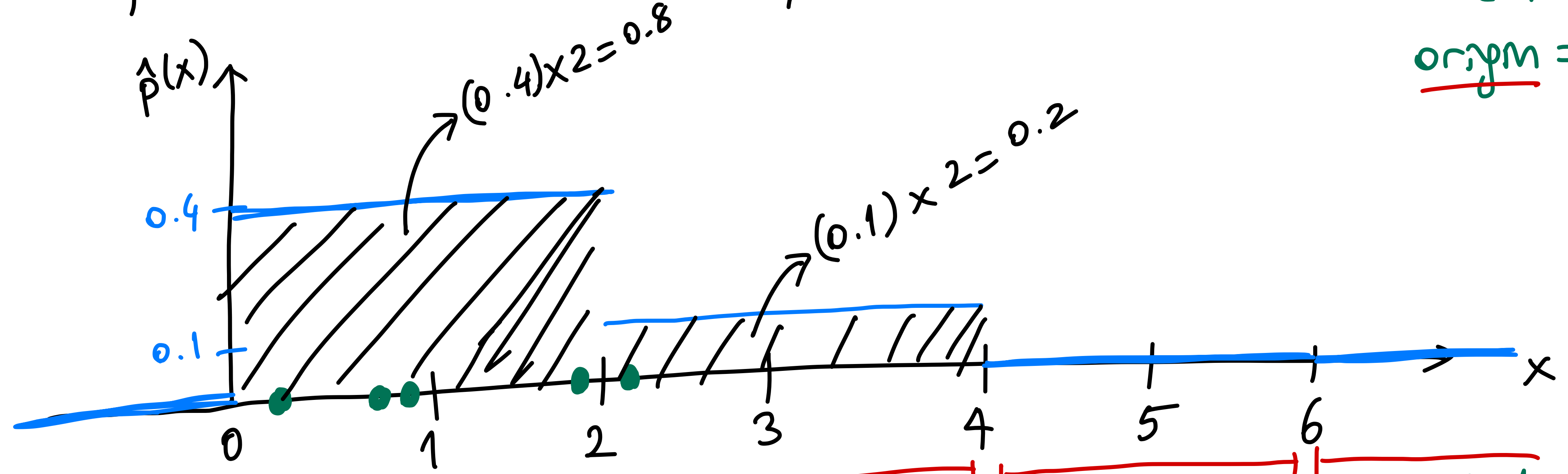


Histogram Estimator

$$\hat{p}(x) = \frac{\#\{x_i\text{'s in the same bin as } x\}}{N} \cdot \frac{1}{h}$$

$h \rightarrow$ bin width ($h=2$)

origin $= 0$



$\hat{p}(x) = \frac{0}{5} \cdot \frac{1}{2} = 0$

$\hat{p}(x) = \frac{4}{5} \cdot \frac{1}{2} = \underline{\underline{0.4}}$

$\hat{p}(x) = \frac{1}{5} \cdot \frac{1}{2} = 0.1$

$\hat{p}(x) = \frac{0}{5} \cdot \frac{1}{2} = 0$

$\hat{p}(x) = \frac{0}{5} \cdot \frac{1}{2} = 0$

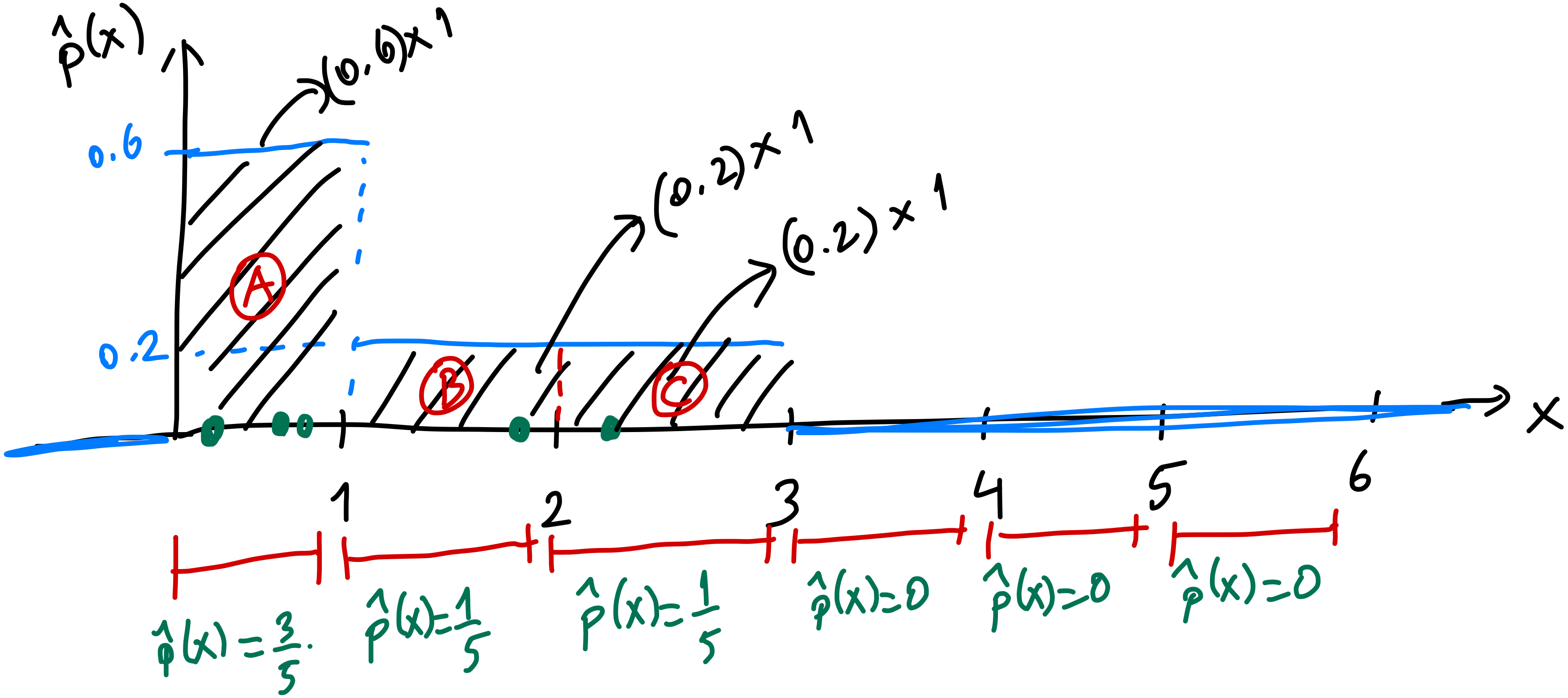
is $\hat{p}(x) \geq 0 \quad \forall x$? ✓

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

Histogram estimator is a valid density estimator!

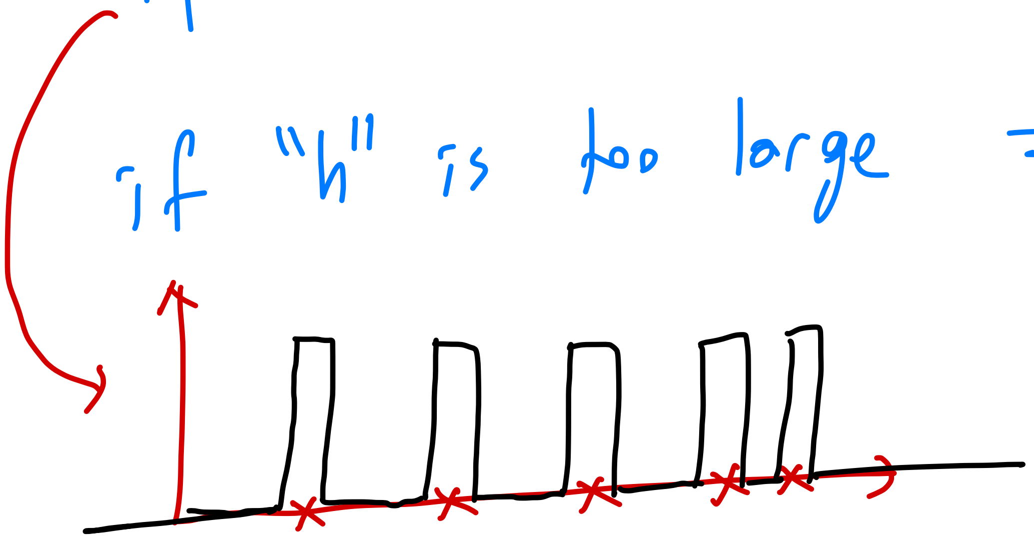
$h = 1$

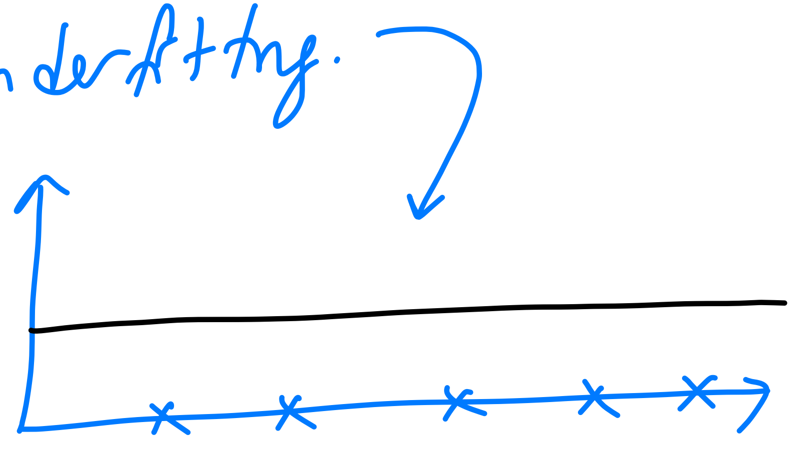
origin $= 0$



if "h" is too small $\Rightarrow$ overfitting

if "h" is too large $\Rightarrow$ underfitting.



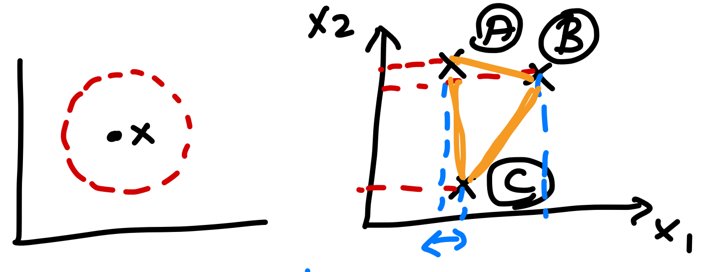


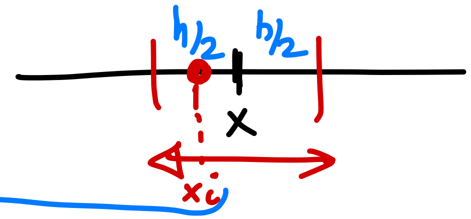
Naive Estimator

$$\hat{p}(x) = \frac{\#\{x - h/2 < x_i \le x + h/2\}}{N} \cdot \frac{1}{h}$$

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

$\downarrow$ weight function





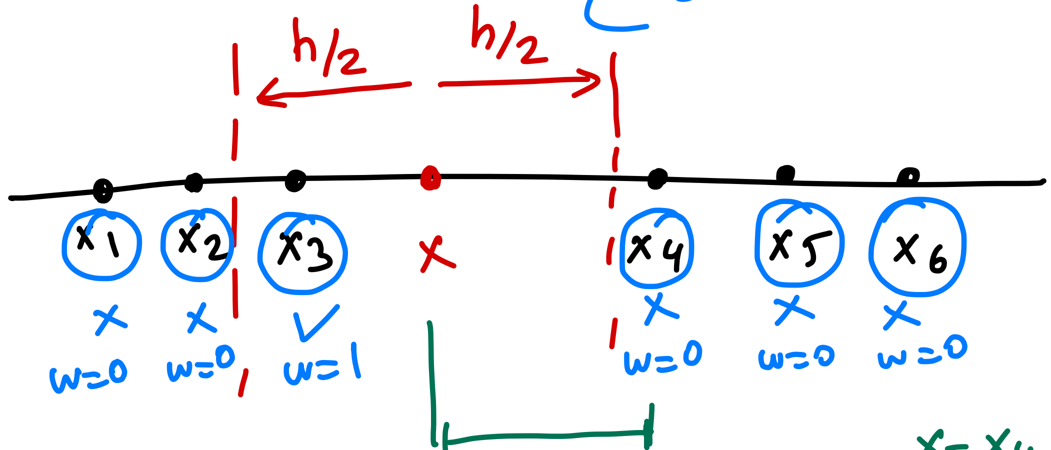
$$w(u) = \begin{cases} 1 & \text{if } |u| \le 1/2 \\ 0 & \text{otherwise} \end{cases}$$

Exercise

Show that $\hat{p}(x)$ is a valid density estimator.

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

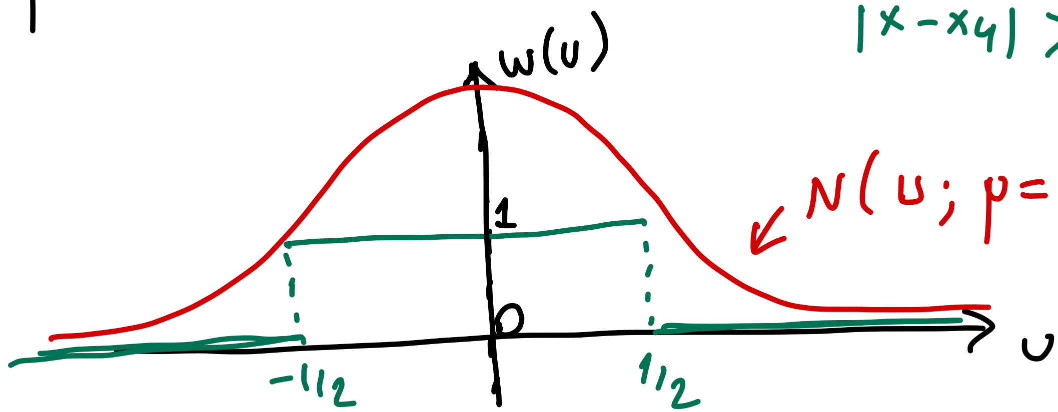
ii) $\int_{-\infty}^{+\infty} \hat{p}(x)\, dx = 1$



$$|x - x_4| > h/2 \Rightarrow \frac{x - x_4}{h} > 1/2$$

$\Downarrow$

$w = 0$



$\swarrow N(u; \mu = 0, \sigma^2 = 1)$

Kernel Estimator (PARZEN WINDOWS) (KDE)

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

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

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

__Exercise:__ Show that Parzen Windows produces a valid density estimator.

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

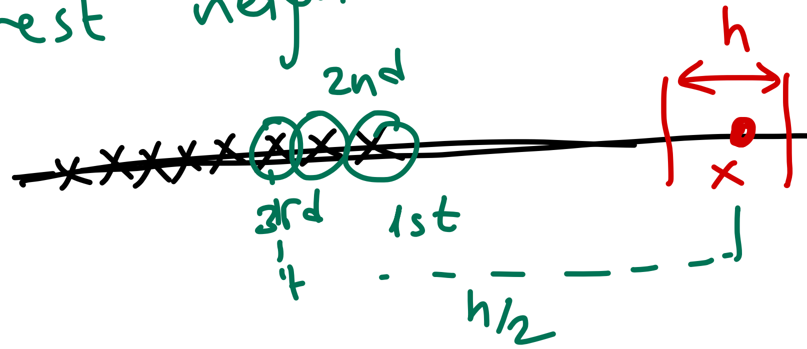
ii) $\displaystyle\int_{-\infty}^{+\infty} \hat{p}(x)\, dx = 1$

k-Nearest Neighbor Estimator

$$\hat{p}(x) = \frac{k}{N \underbrace{2\, d_k(x)}_{h}}$$

k = # of data points that fall into the bin

$d_k(x)$ = the distance to the k^{th} nearest neighbor.

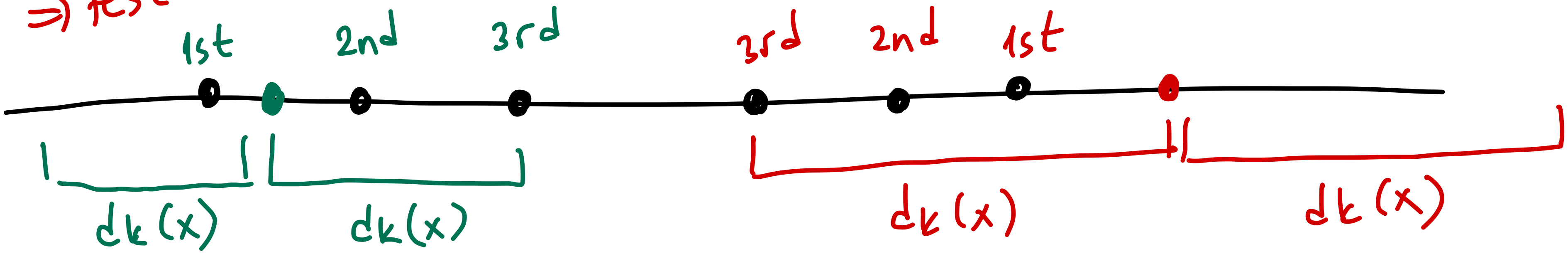


● ⇒ training

● ⇒ test

● ⇒ test

$\underline{k=3}$



$$\hat{p}(\bullet) > \hat{p}(\bullet)$$

Exercise: Show that k-nearest neighbor estimator is **NOT** a valid density estimator.

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

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