Nonparametric Methods

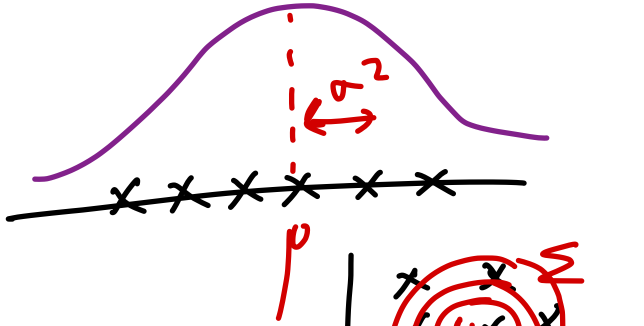
$w^T(x+\epsilon)+w_0$

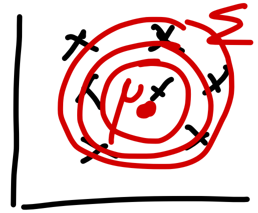
Linear regression $\Rightarrow f(x) = \boxed{w^T . x + w_0}$

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

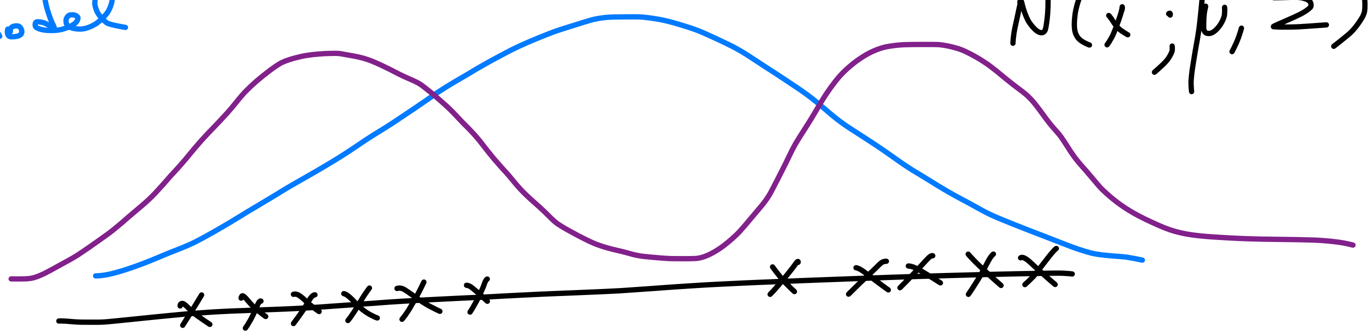
Density estimation $\Rightarrow N(x;\mu,\sigma^2)$

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





bi-model

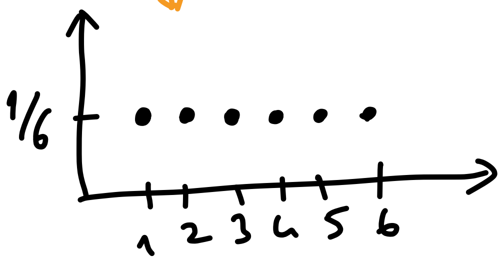


SIMILAR INPUTS $\Rightarrow$ SIMILAR OUTPUTS

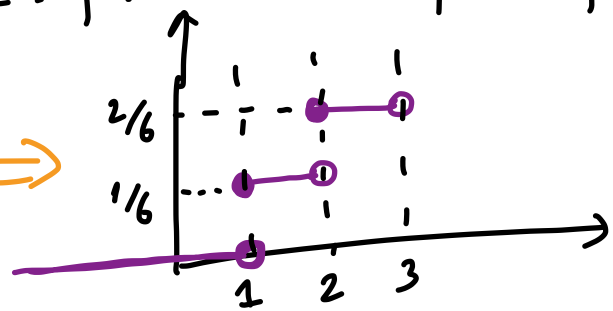
How do we measure similarity?

NONPARAMETRIC $\Rightarrow$ "data-dependent" or "local" models $\Rightarrow$ no parametric form c.d.f. vs p.d.f (p.m.f)

p.m.f



$\Longleftrightarrow$



$F(x=0)=0$
$F(x=0.99)=0$
$F(x=1)=1/6$

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

$$F(x=a)=\int_{-\infty}^{a}p(x)\,dx \qquad F(x=a)=\Pr(x\leq a)$$
$$=\sum_{-\infty}^{a}\Pr(x=a)$$

counting function

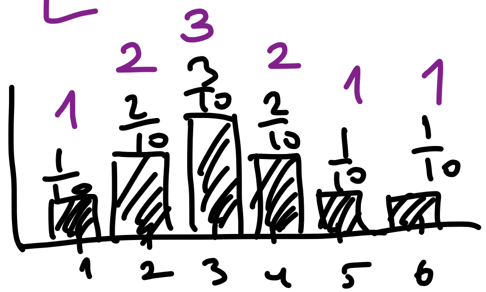
one function

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

$$\lim_{h\to 0}\frac{f(x+\frac{h}{2})-f(x-\frac{h}{2})}{(x+\frac{h}{2})-(x-\frac{h}{2})}$$

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

4, 3, 2, 1, 3, 2, 3, 4, 5, 6



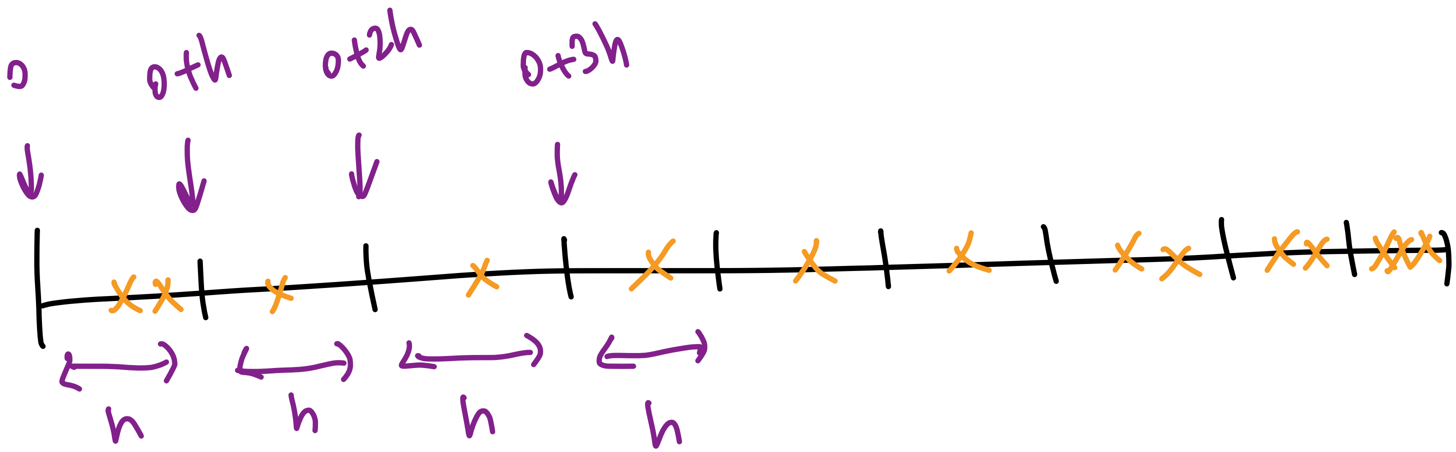
$p(x) \Rightarrow$ I would like to check whether $p(x)$ is a valid density function or not.

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

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

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

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

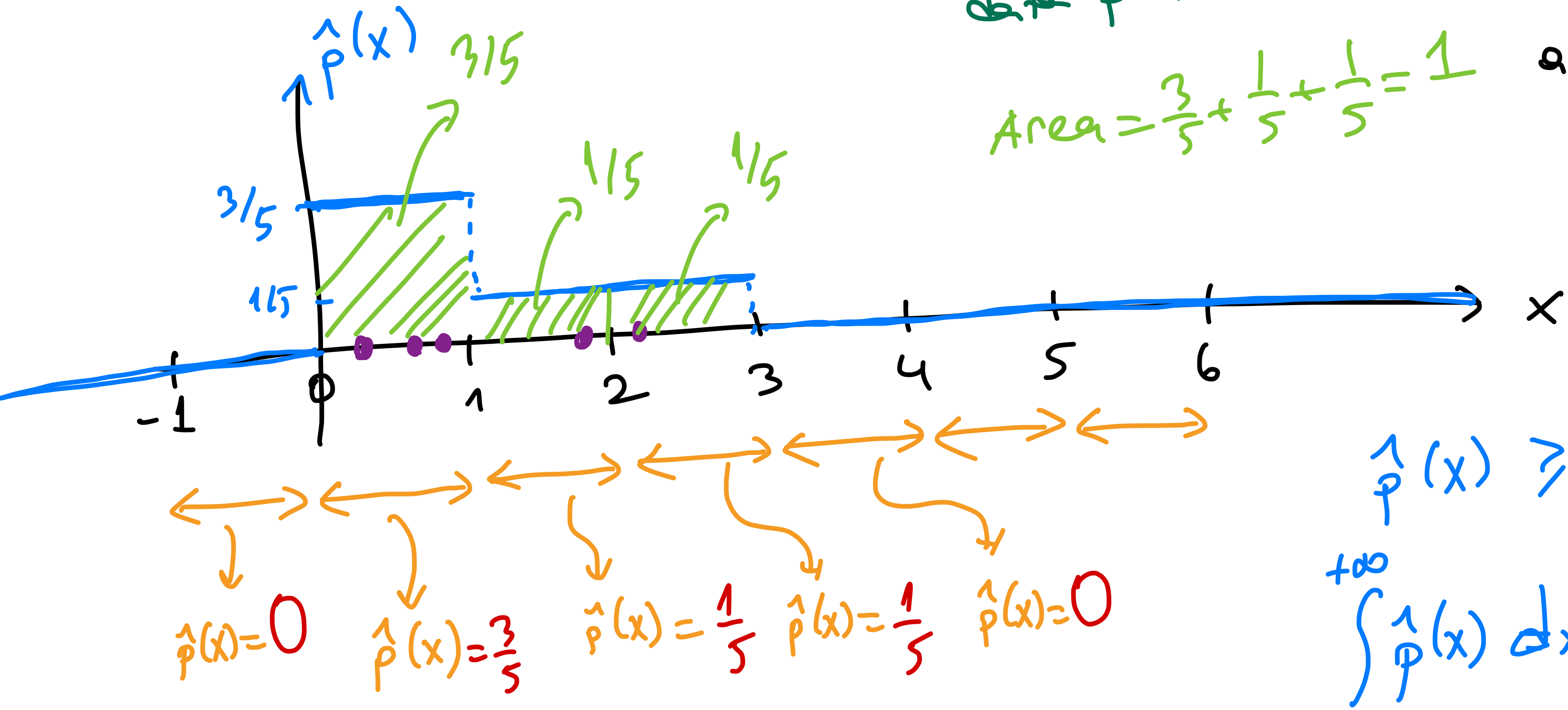
$N \rightarrow$ # of training data points

$h \rightarrow$ bin width, window length $\rightarrow$ model parameter.

assume $h = 1$

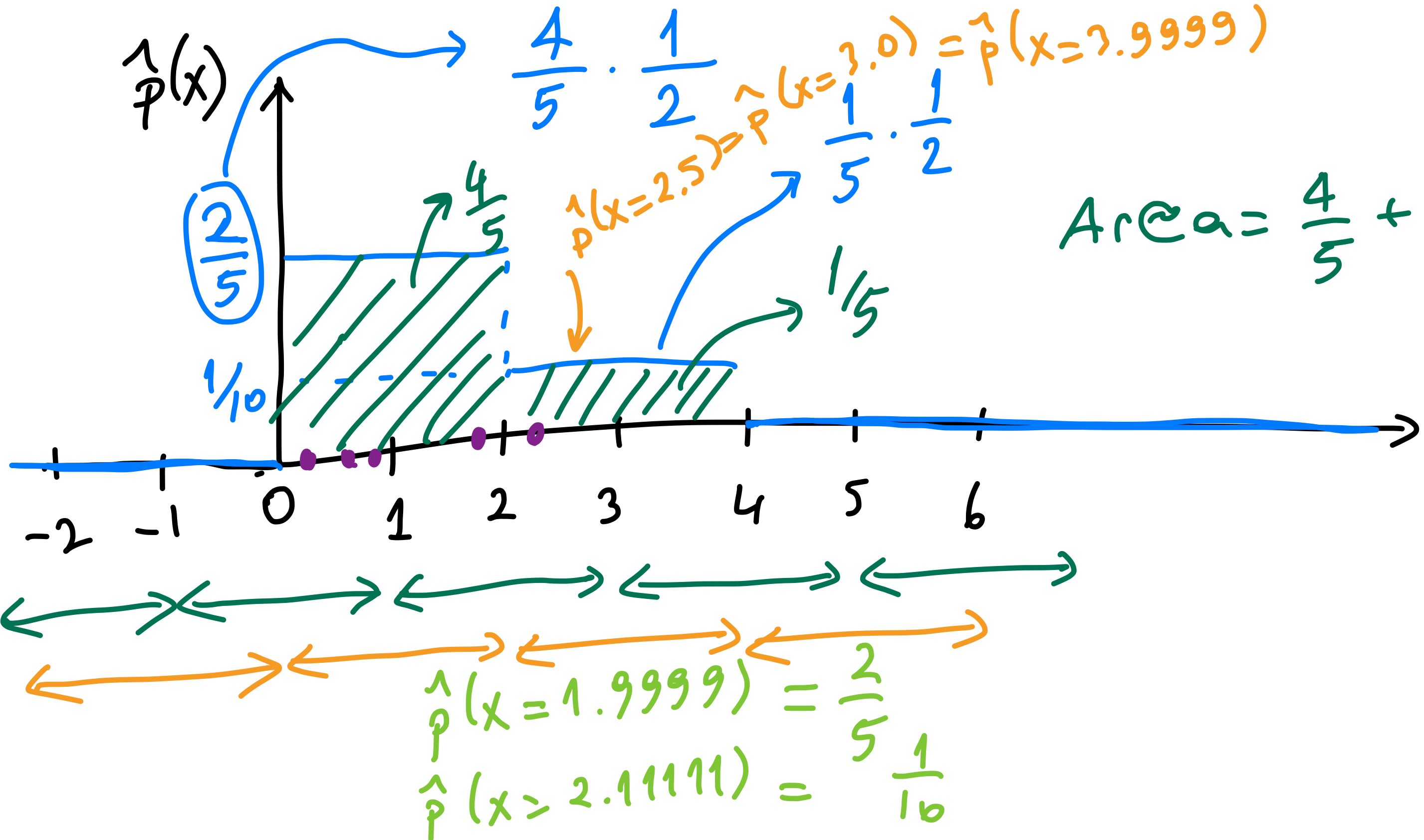
$N = 5$

$$\text{Area} = \frac{3}{5} + \frac{1}{5} + \frac{1}{5} = 1$$



$\hat{p}(x) \geq 0 \quad \forall x$ ✓

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



$$\text{Area} = \frac{4}{5} + \frac{1}{5} = 1$$

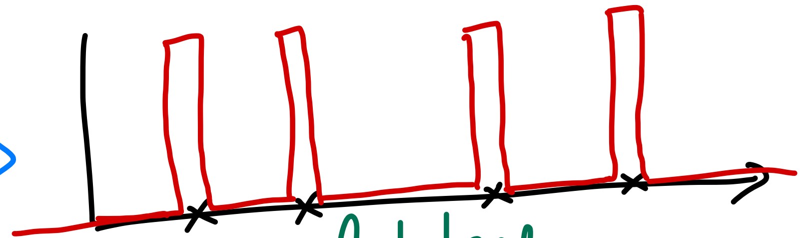
$h = 2$

origin $= 0$

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

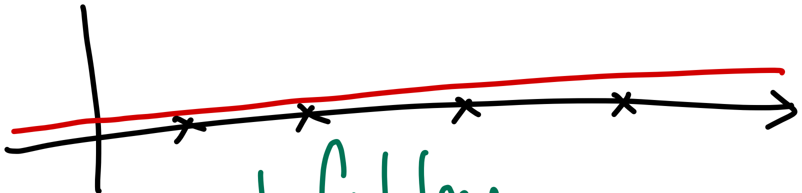
$x_{N+1} \Rightarrow \hat{p}(x) = 0$

if "h" is too small $\Rightarrow$



overfitting

if "h" is too large $\Rightarrow$



underfitting

Naive Estimator

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

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

$$-\frac{1}{2} < \frac{x - x_i}{h} < \frac{1}{2}$$

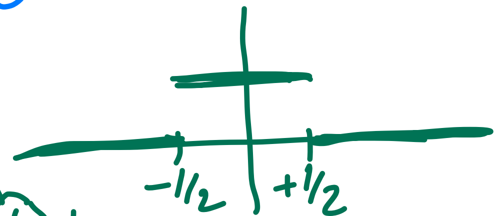
$h \to 2$

if the distance between x & x_i is less then $h/2$

$x - \frac{h}{2} < x_i$	$x_i \le x + \frac{h}{2}$
$x - x_i < \frac{h}{2}$	$-\frac{h}{2} \le x - x_i$
$\vdots$	$\vdots$
$\frac{x - x_i}{h} < \frac{1}{2}$	$-\frac{1}{2} \le \frac{x - x_i}{h}$

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

$$\int_{-\infty}^{+\infty} w(u)\,du = 1$$



Exercise

show that $\hat{p}(x)$ is a valid density estmator.

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

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

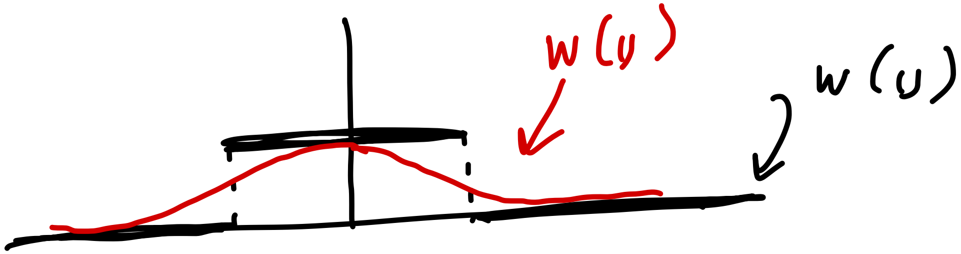
Hint:
$$\int_{-\infty}^{+\infty} \hat{p}(x)\,dx = \int_{-\infty}^{+\infty} \left[\frac{1}{Nh} \sum_{i=1}^{N} w\left(\frac{x - x_i}{h}\right)\right] dx = \frac{1}{Nh} \sum_{i=1}^{N} \int_{-\infty}^{+\infty} w\underbrace{\left(\frac{x - x_i}{h}\right)}_{u} dx = \frac{1}{Nh} \sum_{i=1}^{N} h = 1 \checkmark$$

$h\,du$

$$\frac{x - x_i}{h} = u$$

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

$$dx = h \cdot du$$



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

KERNEL DENSITY ESTIMATOR (KDE) [PARZEN WINDOWS]

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

$$\frac{1}{\sqrt{2\pi\underset{1}{\sigma}^2}} \cdot \exp\left[-\frac{(\overset{u}{x} - \overset{0}{\mu})^2}{2\underset{1}{\sigma}^2}\right]$$

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

Exercise: Show that Parzen windows is a valid density estimator.

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

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

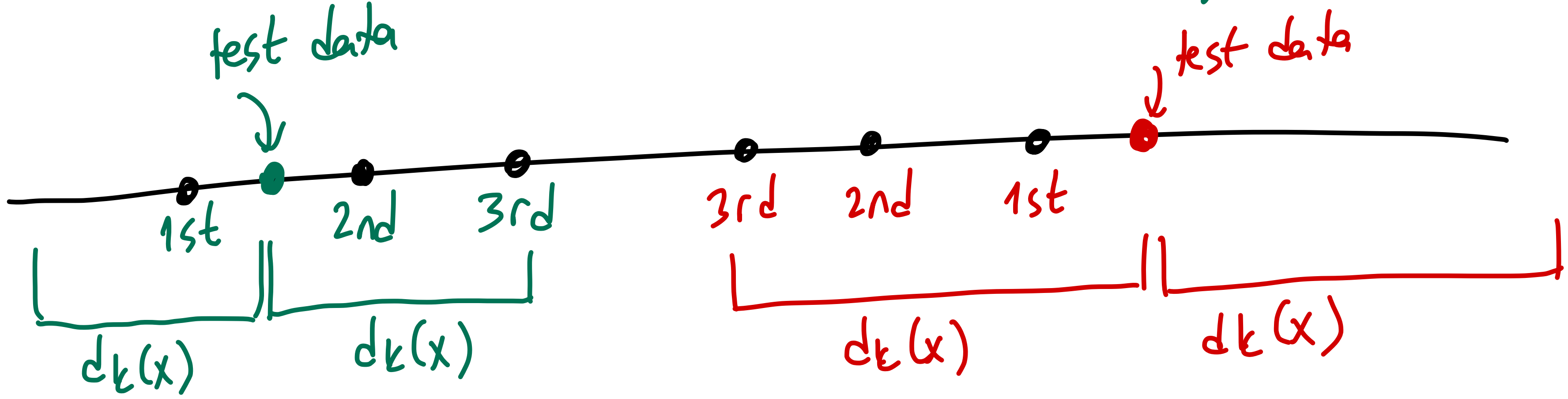
k-Nearest Neighbor Estimator (kNN Estimator)

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

k = # of data points that fall into the bin

$d_k(x)$ = the distance to the kth nearest neighbor

k=3



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

Exercise: show that k-nearest neighbor estimator is **NOT** a valid density estimator. **Hint:** proof by contradiction.