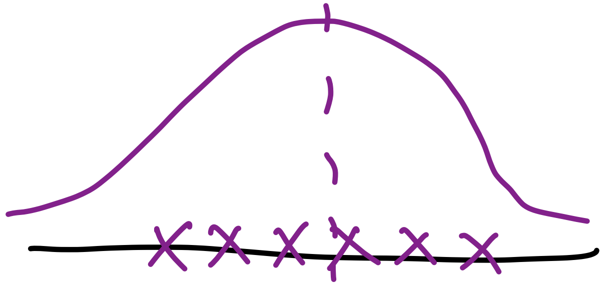
Nonparametric Methods

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

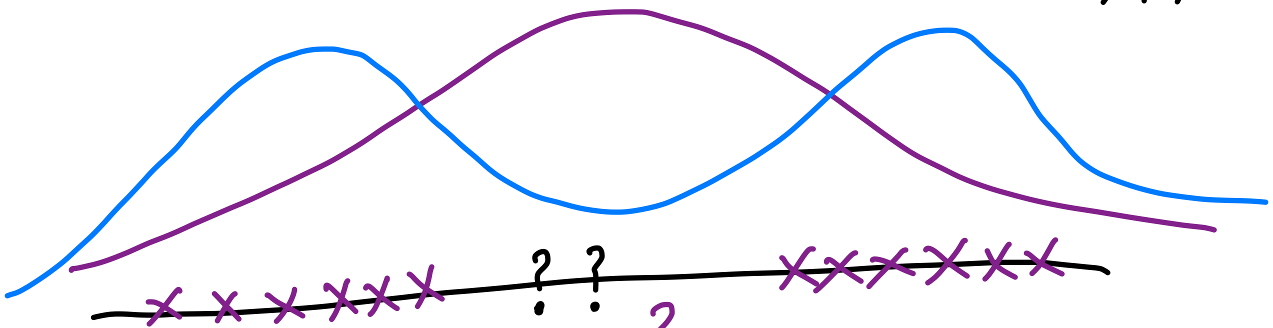
Logistic regression $\Rightarrow \delta(w^T \cdot x + w_0) = \begin{cases} 1 & \text{if } w^T \cdot x + w_0 > 0 \\ 0 & \text{otherwise} \end{cases}$

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

$\mathcal{N}(x; \mu, \Sigma)$



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

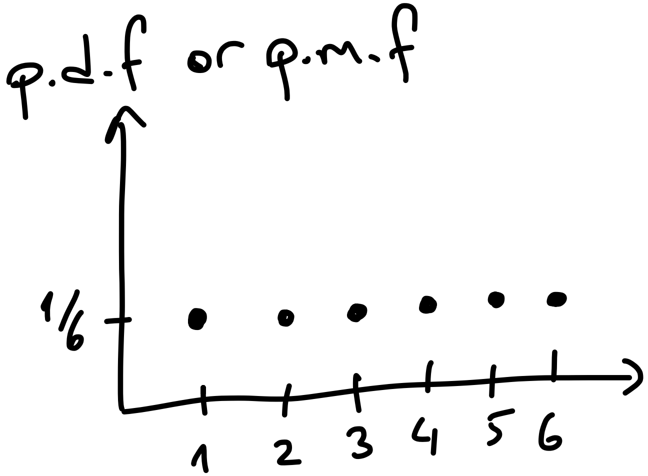


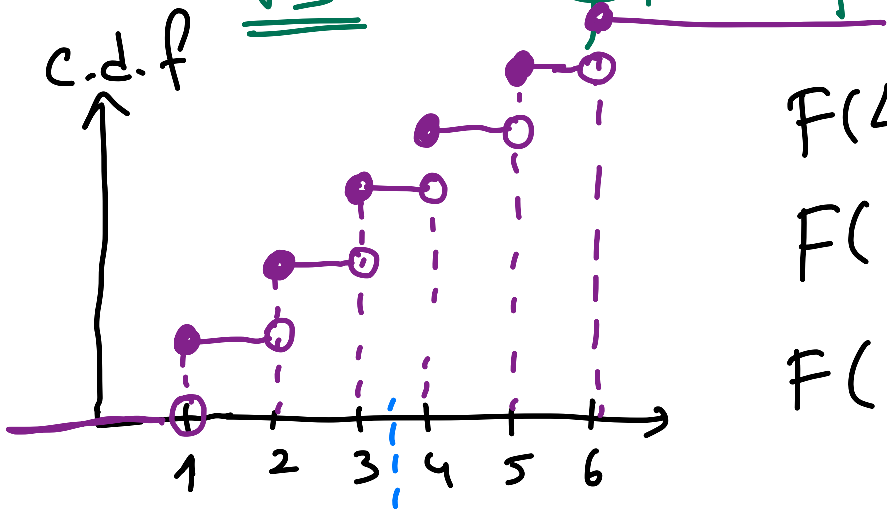
SIMILAR INPUTS $\Rightarrow$ SIMILAR OUTPUTS

How do we measure similarity?

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

c.d.f. vs p.d.f.





$F(4.0) = \frac{4}{6}$

$F(3.7) = \frac{3}{6}$

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

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

$\Uparrow$

Continuous R.V.

$$F(x=a) = P(X \leq a)$$

$\Uparrow$

discrete R.V.

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

counting function

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

$$\boxed{\hat{p}(x) = \frac{1}{h}\,\frac{\#\{x_i \leq x+h\} - \#\{x_i \leq x\}}{N}}$$

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

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

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

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

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

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

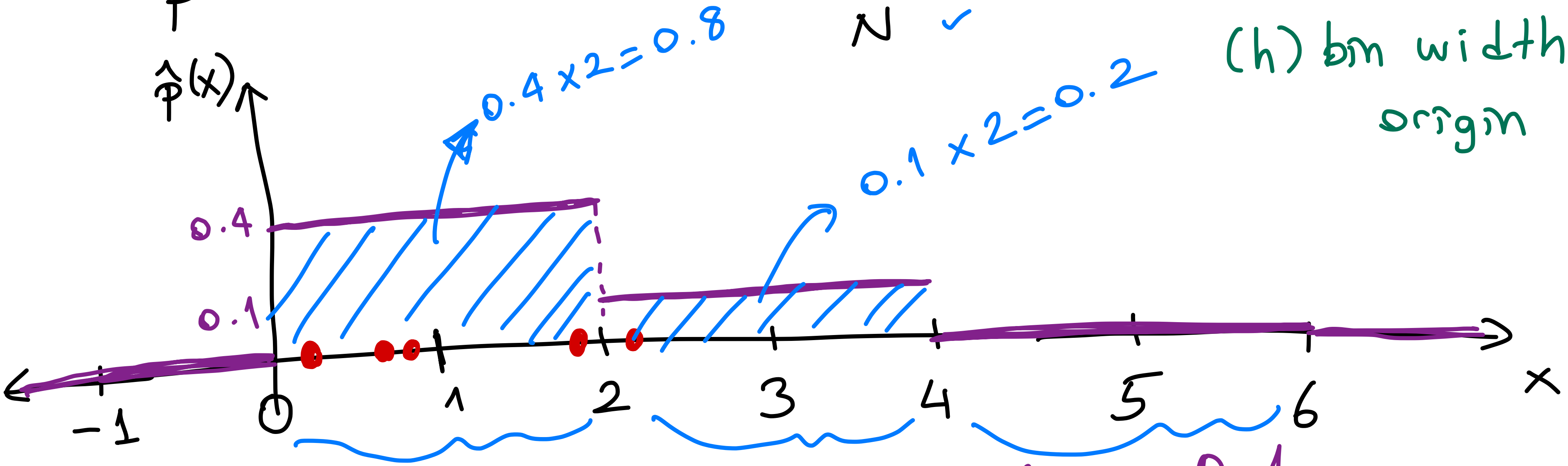
ii) $p(x) \geq 0 \quad \forall x$

Histogram Estimator

$$\hat{p}(x) = \frac{\#\{x_i\text{'s in the same } \overset{\geq 0}{\text{bin}} \text{ as } x\}}{N} \cdot \frac{1}{h}$$

(h) bin width = ? 2

origin = ? 0



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

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

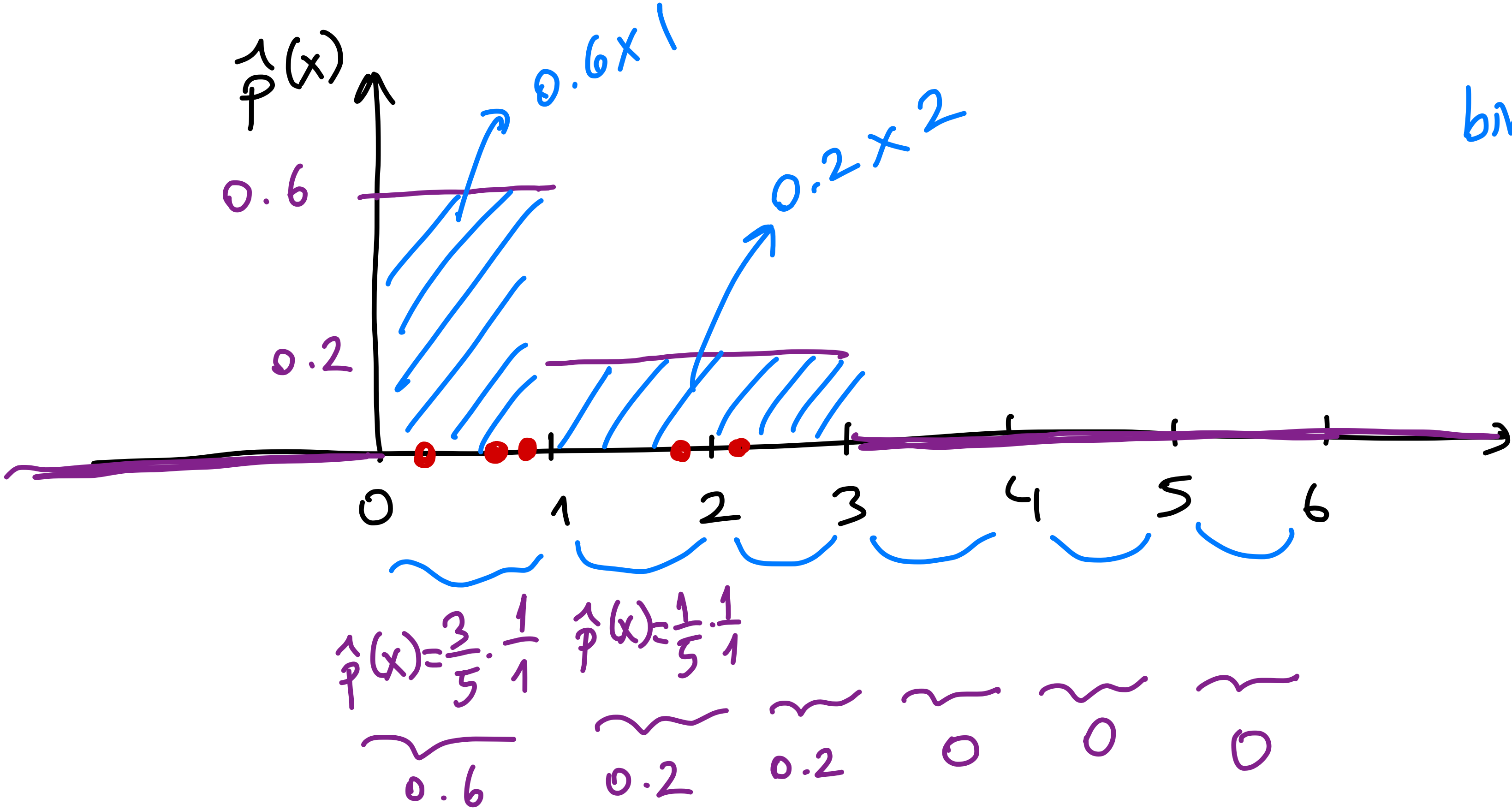
$\hat{p}(x) = \frac{0}{5} \cdot \frac{1}{2} = 0$

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

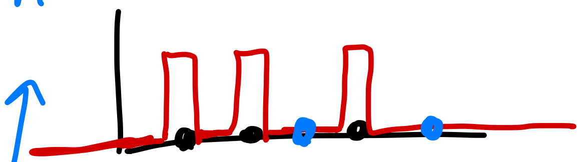
is $\int_{-\infty}^{+\infty} \hat{p}(x)\, dx = 1$? ✓

bin width $(h) = 1$

origin $= 0$

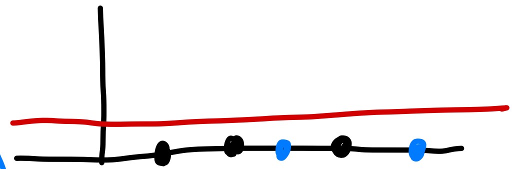


$h = \text{small}$



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

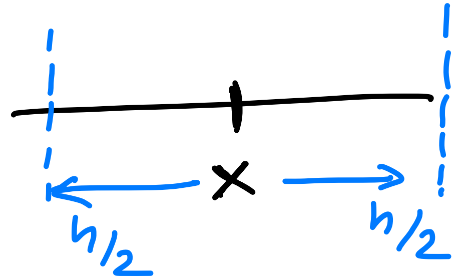
if "h" is too large $\Rightarrow$ underfitting



$h = \text{large}$

Naive Estimator

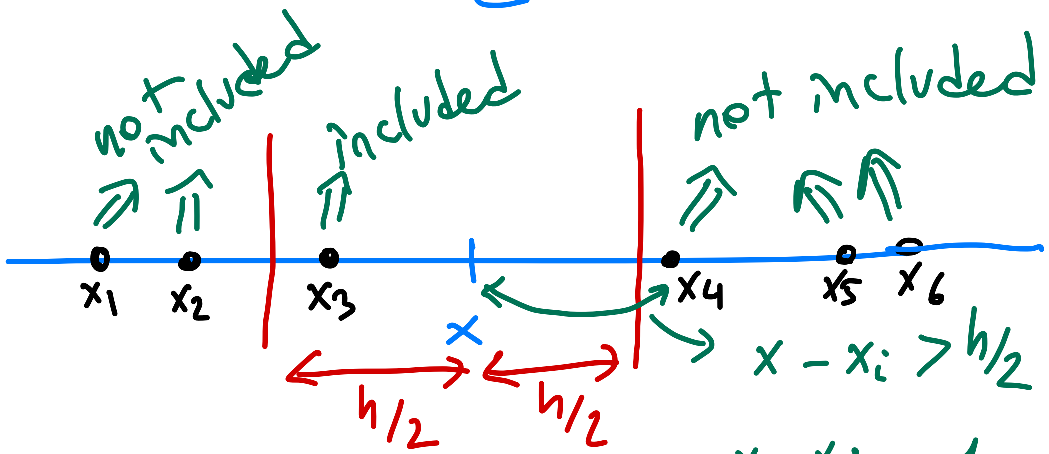
$$\hat{p}(x) = \frac{\#\{x - h/2 < x_i \le x + h/2\}}{N} \cdot \frac{1}{h}$$



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

weight function

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



$x - x_i > h/2$

$$u = \frac{x - x_i}{h} > 1/2$$

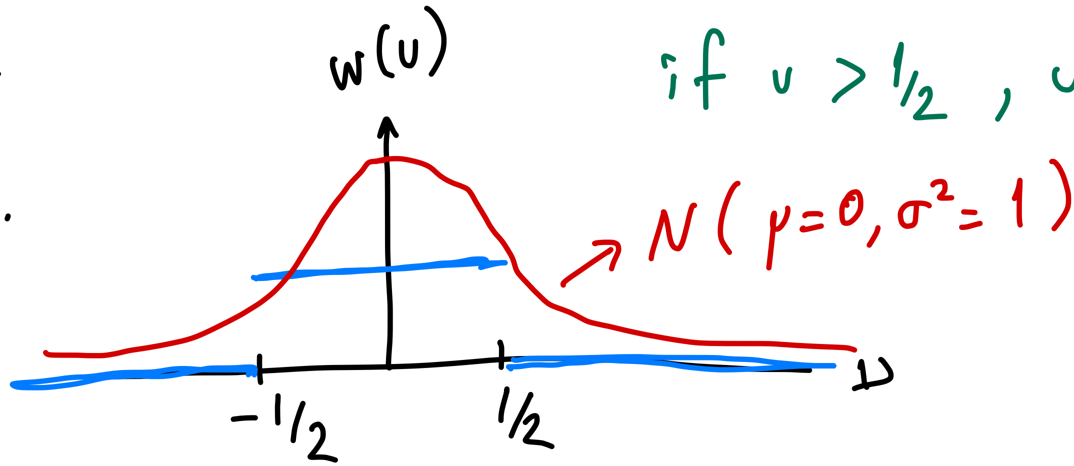
if $u > 1/2$, $w(u) = 0$

Exercise #8:

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

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

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



Kernel Estimator (PARZEN WINDOWS) (KDE)

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

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

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

$\mu = 0$

$\sigma^2 = 1$

Exercise #9: Show that Parzen windows produces a valid density estimator

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

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

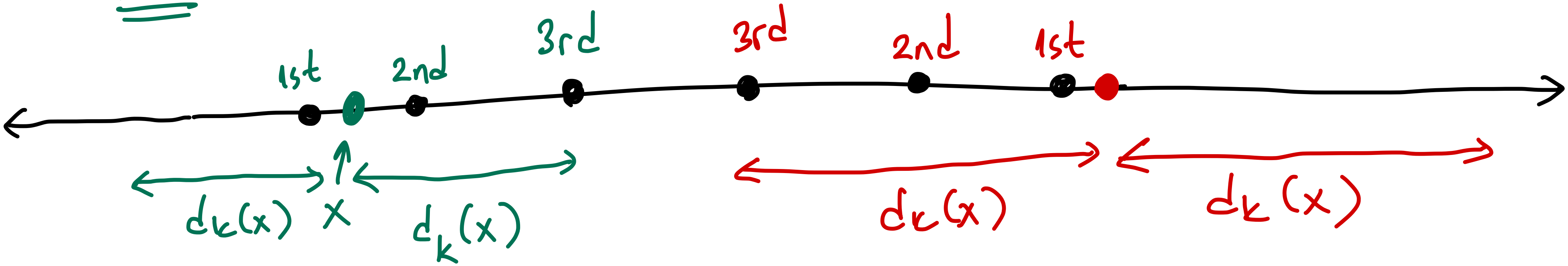
k-Nearest Neighbor Estimator

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

# of data points that fall into the bin.

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

$k=3$



Exercise #10: Show that k-nearest neighbor estimator is NOT a valid density estimator.

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

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