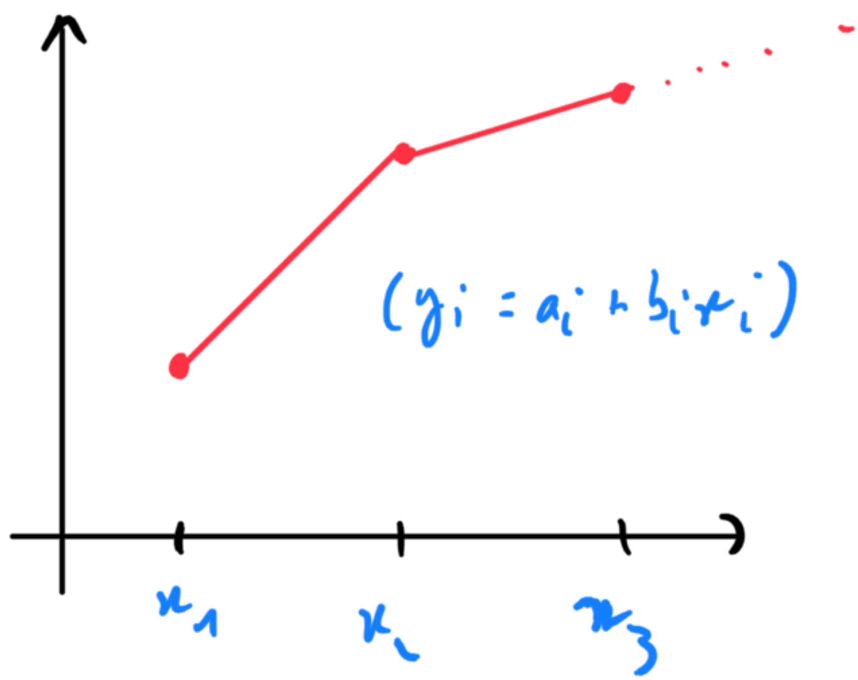
STATS 305 - HW4

Thibaud Bruyelle

① Locally weighted Regression (LWR) assumes that there is more connection between close observations ($|x_i - x_0| < \alpha$).
As LWR estimates a set of parameters for each data point x_i, $i \in [\![1, n]\!]$, the main idea of this method is to give more importance to observations that are close enough. This is why the LWR introduces weights in order to take into account the proximity of regressors

into the model."

Finally, we obtain a set of parameters $(a_0^{(i)}, b_0^{(i)})_{1 \leq i \leq n}$ which defines a locally linear function which is a <u>smooth piecewise linear function</u> :



② The window size w is a parameter that enables the control of the number of extra data observations x_i included in the model for a given pair x_0.

In other words, w can help to have $w(x) = 0$ or $w(x) \neq 0$ without changing the available data. It can be seen as a sensitivity parameter.

Let us consider an arbitrary data pair x_0.

As $w \to 0$,

$$\frac{|x_i - x_0|}{w} = \begin{cases} \gg 1 & \text{if } i \neq 0 \\ 0 & \text{else} \end{cases}$$

$$\Rightarrow \quad W\left(\frac{|x_i - x_0|}{w}\right) = \begin{cases} 0 & \text{if } i \neq 0 \\ 1 & \text{else} \end{cases}$$

so then

$$(\hat{a}_0, \hat{b}_0) \in \underset{a_0, b_0}{\operatorname{argmin}} \left\{ (y_0 - a_0 - b_0 x_0)^2 \right\}$$

<u>the solution is not unique.</u>

As $w \to +\infty$,

$$\frac{|x_i - x_0|}{w} \longrightarrow 0 \quad \text{for all } i$$

Finally $\forall i \in [\![1, n]\!]$,

$$W\left(\frac{|x_i - x_0|}{w}\right) \longrightarrow 1$$

So
all observations are included for

the estimation of each parameter.

Then $(\hat{a}_1, \hat{b}_1) = \dots = (\hat{a}_m, \hat{b}_m)$

The model is equivalent to a <u>simple linear regression</u>.

③ Let us write:

$$\beta_0 = \begin{pmatrix} a_0 \\ b_0 \end{pmatrix}$$

We have:

$$\|y - X\beta_0\|_2^2 = \sum_{i=1}^{m} (y_i - a_0 - b_0 x_i)^2$$

Let us note the diagonal matrix:

$$\tilde{W}_0 = \left[\left(\{ W_0 \}_{ij} \right)^{1/2} \right]_{1 \leq i, j \leq m}$$

We have:

$$\|\tilde{W}_0(\underline{y} - X\beta_0)\|_2^2 = \sum_{i=1}^{n} w\left(\frac{|x_i - x_0|}{\bar{w}}\right)(y_i - \beta_0 x_i)^2$$

So our loss function is:

$$\|\overbrace{\tilde{W}_0 \underline{y} - \tilde{W}_0 X \underline{\beta_0}}^{:= f(\beta_0)}\|_2^2 =$$

$$\underline{y}^T \tilde{W}_0^T W_0 \underline{y} - 2\beta_0^T \underline{X}^T \tilde{W}_0^T \hat{W}_0 y$$

$$+ \underline{\beta_0}^T X^T \bar{W}_0^T \tilde{W}_0 X \beta_0 .$$

$$\frac{\partial \underline{f}(\beta_0)}{\partial \beta_0} = -2 \underline{X}^T \tilde{W}_0^T \hat{W}_0 y$$

$$+ 2 X' W_0 W_0 X \underline{\beta_0}$$

Let us note that:

$$\tilde{W}_0^T \tilde{W}_0 = \tilde{W}_0^2 = W_0 .$$

So:

$$f(\beta_0) = 0 \Leftrightarrow X^T W_0 X \underline{\beta_0} = X^T W_0 y$$

<u>Is $(X^T W_0 X)$ invertible?</u>

Let us note:

$$r := \min\left(\{ i \,/\, \{W_0\}_{ii} \neq 0 \}\right).$$

Then we can write:

$$W_0 = \begin{array}{c} \begin{array}{cc} \overleftrightarrow{\;r\;} & \overleftrightarrow{\;n-r\;} \end{array} \\ \left(\begin{array}{cc} W_{01} & 0 \end{array}\right) \end{array} \; \updownarrow r$$

$$W_0 = \begin{array}{c} \begin{pmatrix} 0 & 0 \end{pmatrix} \end{array} \begin{array}{l} \updownarrow n-r \end{array}$$

where all diagonal terms in W_{01} are non zero.

In a similar way, let us write:

$$X = \underbrace{\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}}_{r} \begin{array}{l} \updownarrow r \\ \updownarrow n-r \end{array}$$

Then:

$$W_0 X = \begin{pmatrix} W_{01} & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

$$= \begin{array}{c} \\ n\times 2 \end{array}\begin{pmatrix} w_{01}x_1 \\ 0 \end{pmatrix}$$

So:

$$X^T W_0 X = \overbrace{\begin{pmatrix} X_1^T & X_2^T \end{pmatrix}}^{n}\begin{pmatrix} W_{01}X_1 \\ 0 \end{pmatrix}$$

$$= \underbrace{X_1^T \overbrace{W_{01} X_1}^{n\times 2}}_{2\times 2}$$

$$W_{01} \times X_1 = \begin{pmatrix} w_{01} & w_{01}x_1 \\ | & | \\ w_{0n} & w_{0n}x_n \end{pmatrix}$$

where $w_{0i} = W\left(\frac{|x_i - x_0|}{w}\right)$

So

$$\begin{pmatrix} w_{01} & w_{01}x_1 \\ | & | \\ w_{0n} & w_{0n}x_n \end{pmatrix}$$

$$X_1^T W_{01} X_1 = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{i=1}^{n} w_{0i} & \sum_{i=1}^{n} w_{0i} x_i \\ \sum_{i=1}^{n} w_{0i} x_i & \sum_{i=1}^{n} w_{0i} x_i^2 \end{pmatrix}$$

And

$$\left(\sum w_{0i}\right)\left(\sum w_{0i} x_i^2\right) - \left(\sum w_{0i} x_i\right)^2 \neq 0$$

so the matrix is invertible and we get

$$\boxed{\begin{pmatrix}\hat a_0\\ \hat b_0\end{pmatrix} = (X^T W_0 X)^{-1}(X^T W_0 y)}$$

4

$$\hat f(x_0) = 1\updownarrow\underbrace{(1 \quad x_0)}_{2}(X^T W_0 X)^{-1}(X^T W_0)\, y$$

$$= h(x_0)^T y$$

where

$$\boxed{\underline{h}(x_0) = W_0 \, X (X^T W_0 X)^{-1}\begin{pmatrix}1\\ x_0\end{pmatrix}}$$

Let us compute $(X^T W_0 X)$

$$\det\left[\left(X^T W_0 X \right)^{-1} \right] :$$

$$\left(\sum_{i=1}^{n} w_{0i}\right)\left(\sum_{i=1}^{n} w_{0i} x_i^2\right) - \left(\sum_{i=1}^{n} w_{0i} x_i\right)^2$$

$$:= \Delta_0$$

and let us write:

$$m_0 := \sum_{i=1}^{n} w_{0i} \, ,$$

$$\mu_0 := \sum_{i=1}^{n} w_{0i} x_i$$

$$p_0 := \sum_{i=1}^{n} w_{0i} x_i^2$$

So

$$(X^T W_0 X)^{-1} = \frac{1}{\Delta_0} \begin{bmatrix} p_0 & -n_0 \\ -n_0 & m_0 \end{bmatrix}$$

then:

$$h(x_0) = w_0^T X \begin{bmatrix} p_0/\Delta_0 & -n_0/\Delta_0 \\ -n_0/\Delta_0 & m_0/\Delta_0 \end{bmatrix} \begin{pmatrix} 1 \\ x_0 \end{pmatrix}$$

$$= w_0 X \begin{bmatrix} \frac{1}{\Delta_0}(p_0 - n_0 x_0) \\ \\ \frac{1}{\Delta_0}(-n_0 + m_0 x_0) \end{bmatrix}$$

So finally, $\forall i \in [\![1, n]\!]$:

$$h_i(x_0) = w_{0i} \frac{1}{\Delta_0} \left[p_0 - x_0 n_0 + x_i(m_0 x_0 - n_0) \right]$$

$$
\boxed{= W\left(\frac{|x_0 - x_i|}{\bar{w}}\right) \cdot \frac{1}{\Delta_0}\left[p_0 - x_0 r_0 + x_i\left(m_0 x_0 - r_0\right)\right]}
$$

where

- $w_{ai} = W\left(\dfrac{|x_i - x_0|}{\bar{w}}\right)$

- $\Delta_0 = \sum_i' w_{ai} \cdot \sum_i' w_{ai} x_i^2 - \left[\sum_i' w_{ai} x_i\right]^2$

- $p_0 = \sum_i' w_{ai} x_i^2$

- $m_0 = \sum_i' w_{ai}$

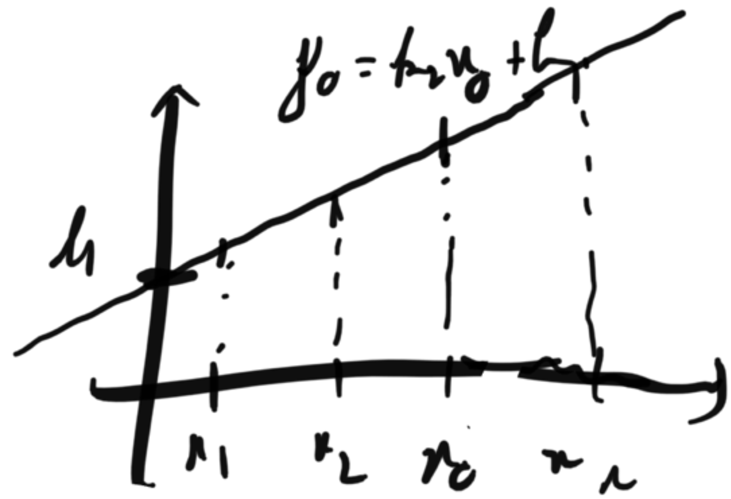
- $r_0 = \sum_i' w_{ai} x_i$

(3) Let us assume that

$\exists (b_1, b_2) \in \mathbb{R}^2, \quad Y = X \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$

which is equivalent to say that

$(x_i, y_i)_i$ lie on the same line.



In this case we can always set $\hat{a}_0 = b_2$ and $\hat{b}_0 = b_1$ so we will get an RSS equals to 0. This is independent from the weights.

Finally : $\hat{f}(x_0) = b_1 + b_2 x_0$

so <u>the fit lies on the same</u>
<u>line</u>.

(6) let us show that the

bias of $\hat{f}(x_0)$ is

$$f(x_0) - h(x_0)' f$$

where $f = (f(x_1), \dots, f(x_n))^T$.

$$\mathbb{E}\left(f(x_0) - \hat{f}(x_0)\right)$$

$$= \mathbb{E}\left[f(x_0) - h(x_0)^T y\right]$$

$$= f(x_0) - \mathbb{E}\left(h(x_0)^T y\right)$$

$$= f(x_0) - h(x_0)^T \mathbb{E}(y)$$

Since we assumed the model

$$y = f(x) + \varepsilon$$

Finally the bias writes:

$$f(x_0) - h(x_0)^\top f$$

because $\mathbb{E}(y) = \mathbb{E}(f(x) + \varepsilon)$

$$= \mathbb{E}(f(x)) + \underbrace{\mathbb{E}(\varepsilon)}_{=0}$$

$$= f(x)$$

$$= \begin{bmatrix} f(x_1) \\ | \\ f(x_n) \end{bmatrix}$$

Finally:

$$\boxed{\mathbb{E}\left(f(x_0) - \hat{f}(x_0)\right) = f(x_0) - h(x_0)^T \underline{f}\ .}$$

Let $i \in [\![1, n]\!]$.

$$\underline{f(x_i) = f(x_0) + f'(x_0)(x_i - x_0)}$$
$$\underline{+\, o(x_i - x_0)}$$

is the Taylor development of $f(x_i)$ to the first order.

So $h(x_0)^T \underline{f}$ writes:

$$\sum_{i=1}^{n} h_i(x_0)^T f(x_i) =$$

$$\sum_{i=1}' h_i(x_0)\times f(x_0) + h_i(x_0) f'(x_0)(x_i - x_0)$$

$$= \boxed{f(x_0)\sum_{i=1}^{n}{}' h_i(x_0) + f'(x_0)\sum_{i}^{n}{}' h_i(x_0)(x_i - x_0)}$$

First, let us show that:

$$\sum_{i=1}^{n}{}' h_i(x_0) = 1 .$$

Actually:

$$\sum_{i=1}^{n}{}' h_i(x_0) =$$

$$\frac{1}{\Delta_0}\sum{}' w_{ci}\left[p_0 - r_0 x_0 + x_i(-r_0 + m_0 x_0)\right]$$

Then:

- $\sum_i w_{oi} p_o = p_o \underbrace{\sum_i w_{oi}}_{= m_o}$

$$= p_o \times m_o$$

- $r_o x_o \sum_i w_{oi} = r_o \times m_o \times x_o$

- $r_o \sum_i w_{oi} x_i = r_o \times r_o$

- $m_o x_o \sum_i w_{oi} x_i = r_o m_o x_o$

So $\sum_{i=1}^{n} h_i(x_o)$ equals:

$$\frac{1}{\Delta_o} \times \left(-r_o m_o x_o + r_o m_o x_o + \underbrace{p_o m_o - r_o^2}_{= \Delta_o} \right)$$

$= 1$.

Finally : $\boxed{\sum_{i=1}^{n} h_i(x_0) = 1}$.

<u>Secondly</u>, let us show that

$$\sum_{i=1}^{n} (x_i - x_0) h_i(x_0) = 0 .$$

From what we wrote above,

$$-x_0 \sum_{i=1}^{n} h_i(x_0) = -x_0 .$$

Now let us show that :

$$\sum_{i=1}^{n} x_i h_i(x_0) = x_0 .$$

- $\sum x_i \, p_0 = x_0 p_0$

$\cdot\ \sum_i' x_i\, n_0 x_0 = n_0^2 x_0$

$\cdot\ \sum_i' x_i^2 w_{0i}\, n_0 = p_0 n_0$

$\cdot\ \sum_i' x_i^2 w_{0i}\, m_0 x_0 = p_0 m_0 x_0$

So $\sum_i' h_i(x_0)\, x_i$ writes:

$$\frac{1}{\Delta_0}\Big(n_0 p_0 - n_0^2 x_0 - p_0 n_0 + p_0 m_0 x_0 \Big)$$

$$= \frac{1}{\Delta_0}\Big(x_0 \underbrace{\big(p_0 m_0 - n_0^2 \big)}_{= \Delta_0} \Big)\Big]$$

$$= x_0 .$$

Finally: $f'(x_0) \sum_i' (x_i - x_0)\, h_i(x_0) = 0$

and to the first order:

$$\boxed{\mathbb{E}\left(f(x_0) - \hat{f}(x_0)\right) \doteq 0.}$$