

Estadística III

Regresión paramétrica, suavizamiento

Alejandro López Hernández

FES Acatlán - UNAM

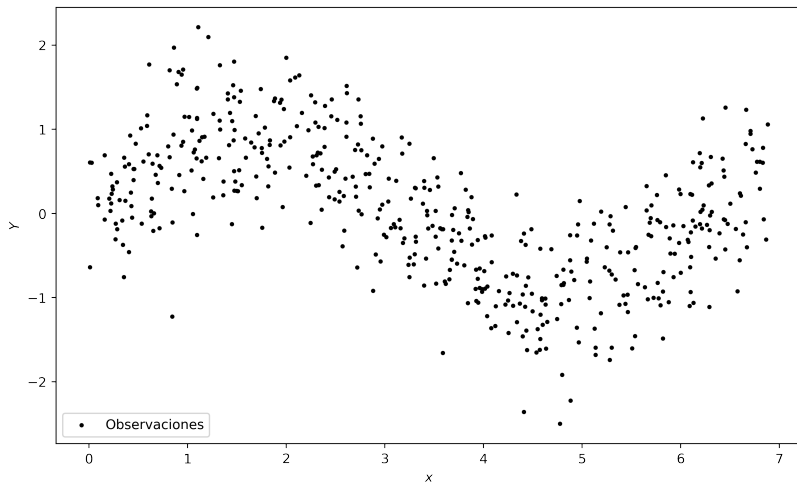
May 21, 2020

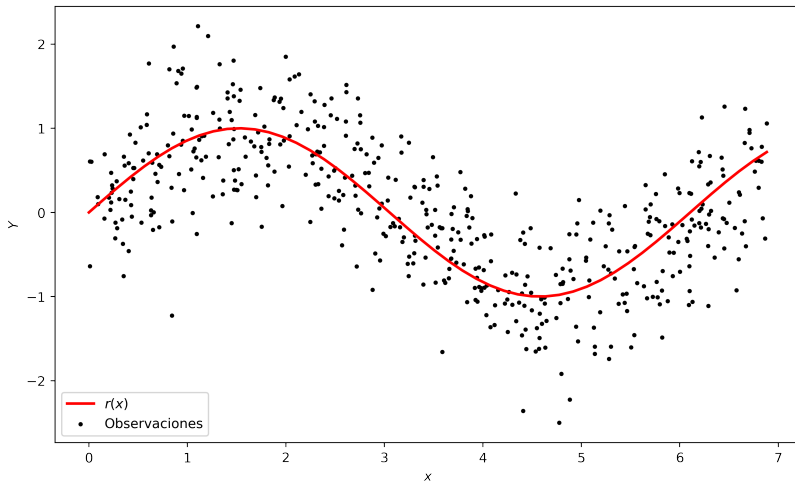
- 1 Introducción
- 2 Regresores Lineales
- 3 Validación cruzada
- 4 Estimación de Kernel Nadaraya-Watson
- 5 Estimación de $\sigma^2(x)$

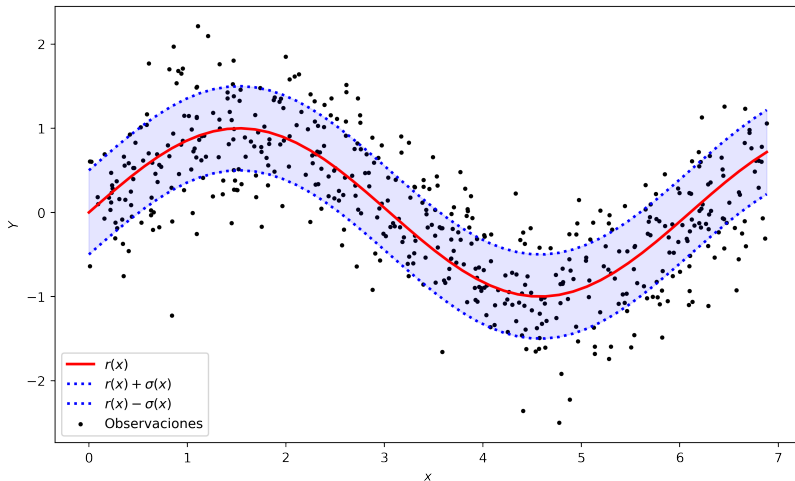
La regresión consiste en encontrar una función que relacione 2 observaciones X, Y , nos es dado n pares de observaciones $(x_1, Y_1), \dots, (x_n, Y_n)$ la forma en la que estos pares están relacionados es con la formula

$$Y_i = r(x_i) + \varepsilon_i, \quad \mathbb{E}(\varepsilon_i) = 0 \text{ para } i = 1, 2, \dots, n$$

Nuestro objetivo es estimar la función $r(x)$.







Los tipos de funciones de regresores que veremos son **regresores lineales**

Definición

Un estimador \hat{r}_n de r es un **regreso lineal** sí para cada x existe un vector $\ell(x) = (\ell_1(x), \ell_2(x), \dots, \ell_n(x))^T$ tal que

$$\hat{r}_n(x) = \sum_{i=1}^n \ell_i(x) Y_i$$

Se define el vector de los **valores ajustados** como

$$\mathbf{r} = (\hat{r}_n(x_1), \dots, \hat{r}_n(x_n))^T$$

de esta forma si definimos $Y = (Y_1, \dots, Y_n)^T$ podemos escribir \mathbf{r} como

$$\mathbf{r} = LY$$

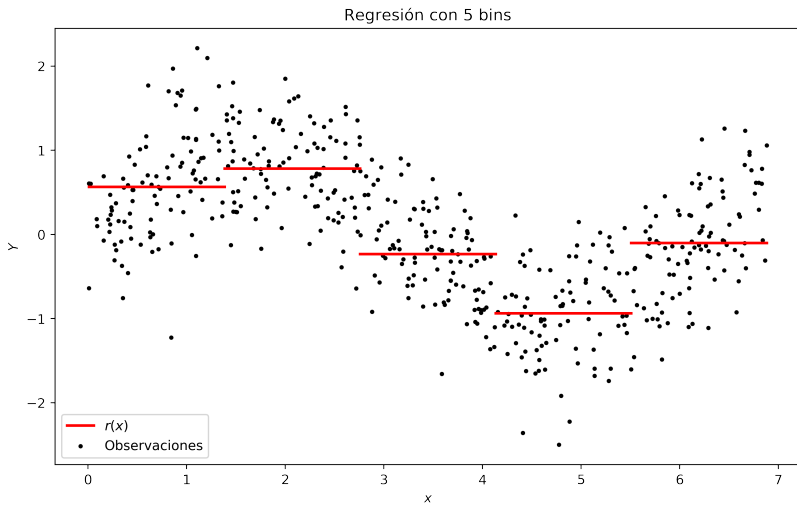
donde L es una matriz de $n \times n$ con valores $L_{ij} = \ell_j(x_i)$.

Ejemplo 1 Dividimos nuestro intervalo (a, b) en m bins del mismo tamaño lo cuales denotamos como B_1, \dots, B_m , definimos $\hat{r}_n(x)$ como

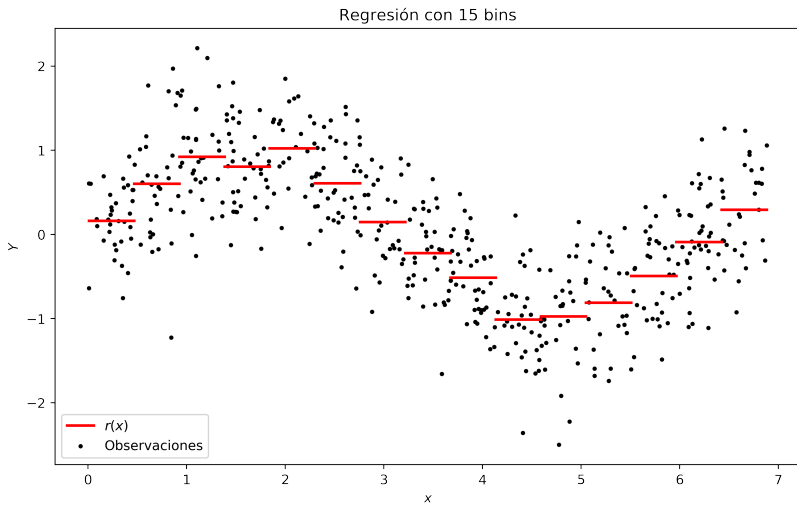
$$\hat{r}_n(x) = \frac{1}{k_j} \sum_{i: x_i \in B_j}^n Y_i$$

Donde k_j es la cantidad de observaciones en B_j

Regresores Lineales

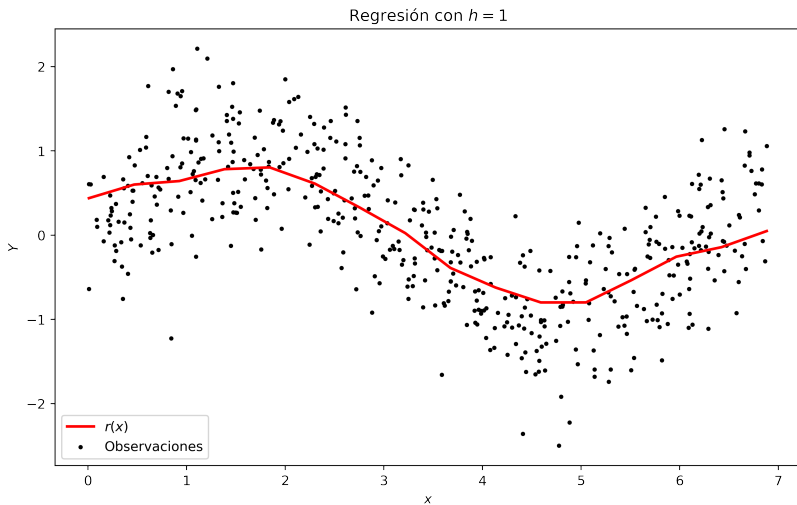


Regresores Lineales

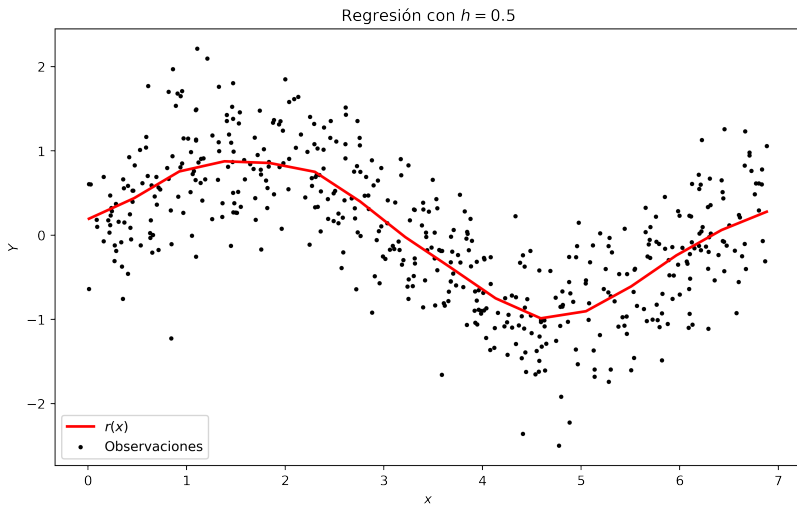


Ejemplo 2 Sea $h > 0$ y $B_x = \{i : |x_i - x| \leq h\}$, sea n_x la cantidad de puntos que pertenecen a B_x , para cada x tal que $n_x > 0$ se define

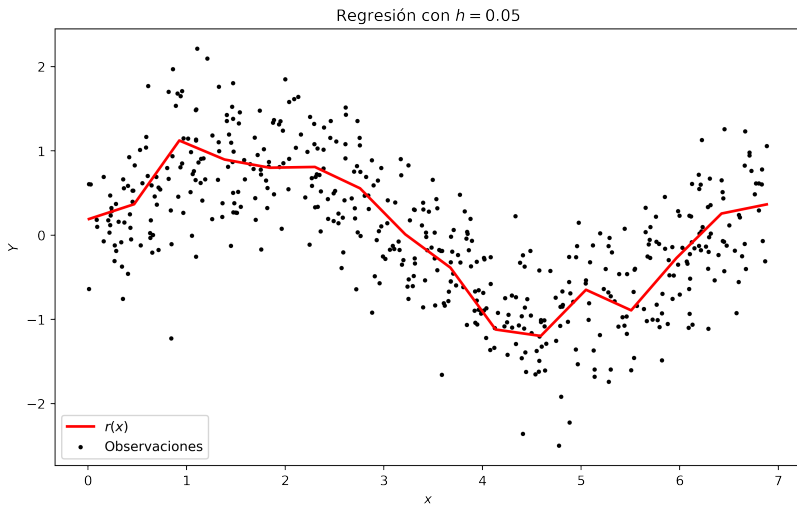
$$\hat{r}_n(x) = \frac{1}{n_x} \sum_{i \in B_x}^n Y_i$$



Regresores Lineales



Regresores Lineales



Para poder seleccionar nuestro parámetro de suavizamiento necesitamos de un criterio con el que podamos medir el error que estamos cometiendo al estimar r , nuestro error lo podemos medir como

$$R(h) = \mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n (\hat{r}_n(x_i) - r(x_i))^2 \right)$$

Debido a que R depende de r no podemos minimizar directamente R , en lugar de eso, buscaremos minimizar

$$\hat{R}(h) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{r}_n(x_i))^2$$

Sin embargo esta estimación resulta no ser muy efectiva, de alguna forma sobreajusta al momento de comparar $\hat{r}_n(x_i)$ con Y_i estamos usando la información original para comparar. Para medir el riesgo utilizaremos validación cruzada, excluyendo un dato a la vez:

Definición

El score de validación cruzada esta definido como

$$cv = \hat{R}(h) = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{r}_{(-i)}(x_i))^2$$

Donde $\hat{r}_{(-i)}$ es la estimación de r sin utilizar la i -ésima observación (x_i, Y_i)

Con esta función de h podemos buscar el valor h que minimiza el error.

En el caso particular de los regresores lineales $\hat{R}(h)$ se puede simplificar

Teorema

Si \hat{r}_n es un estimador lineal entonces la validación cruzada se puede reescribir como

$$\hat{R}(h) = \frac{1}{n} \sum_{i=1}^n \left(\frac{Y_i - \hat{r}_n(x_i)}{1 - L_{ii}} \right)^2$$

Estimación de Kernel Nadaraya-Watson

Nos enfocaremos en el estimador de **Kernel Nadaraya-Watson**, esta estimación consiste en utilizar un Kernel K y un ancho de banda $h > 0$, promedio de forma ponderada cada uno de los puntos, la ponderación esta dada por la función K .

Definición

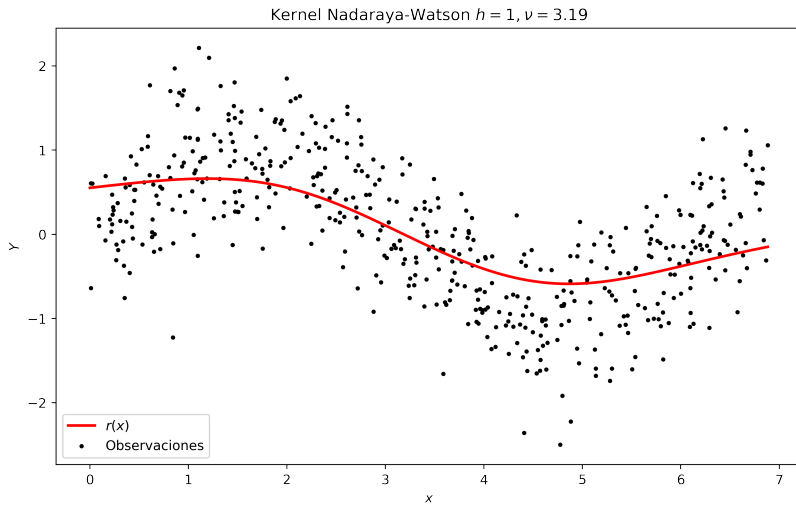
Sea $h > 0$, se define el estimador de Kernel Nadaraya-Watson como,

$$\hat{r}_n(x) = \sum_{i=1}^n \ell_i(x) Y_i$$

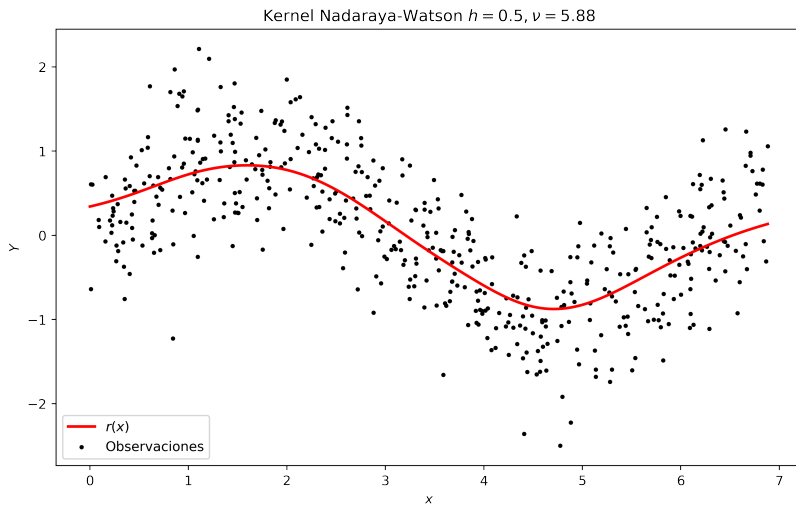
donde K es un kernel y,

$$\ell_i(x) = \frac{K\left(\frac{x-x_i}{h}\right)}{\sum_{j=1}^n K\left(\frac{x-x_j}{h}\right)}$$

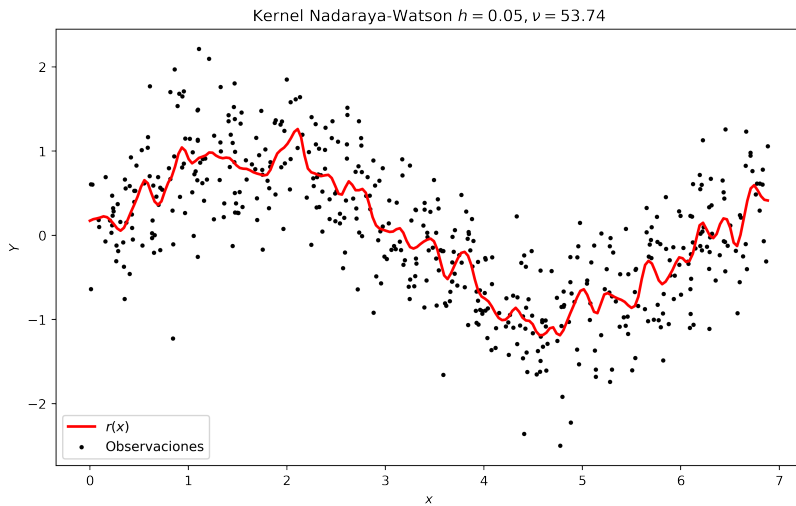
Estimación de Kernel Nadaraya-Watson



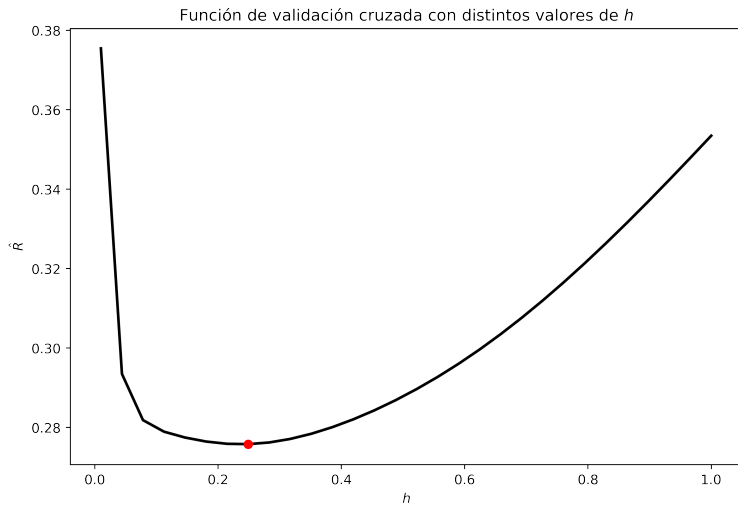
Estimación de Kernel Nadaraya-Watson



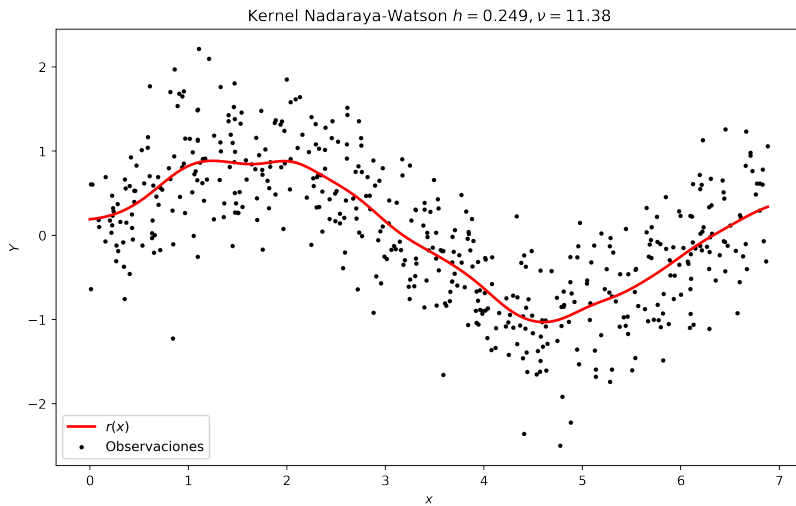
Estimación de Kernel Nadaraya-Watson



Estimación de Kernel Nadaraya-Watson



Estimación de Kernel Nadaraya-Watson



En el ejemplo que hemos considerado la varianza tiene un comportamiento homocedastico, si suponemos que la varianza de nuestro modelo es constante, es decir $\sigma^2 = \text{Var}(\varepsilon_i)$. En esta situación podemos utilizar el siguiente estimador para regresores lineales.

Teorema

Sea \hat{r}_n un regresor lineal y sea

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{r}_n(x_i))^2}{n - 2\nu + \tilde{\nu}}$$

Donde $\nu = \text{tr}(L)$ y $\tilde{\nu} = \text{tr}(L^T L)$ Si $\nu = o(n)$ y $\tilde{\nu} = o(n)$ entonces $\hat{\sigma}^2$ es un estimador consistente de σ^2

En el ejemplo que hemos considerado la varianza tiene un comportamiento homocedastico, si suponemos que la varianza de nuestro modelo es constante, es decir $\sigma^2 = \text{Var}(\varepsilon_i)$. En esta situación podemos utilizar el siguiente estimador para regresores lineales.

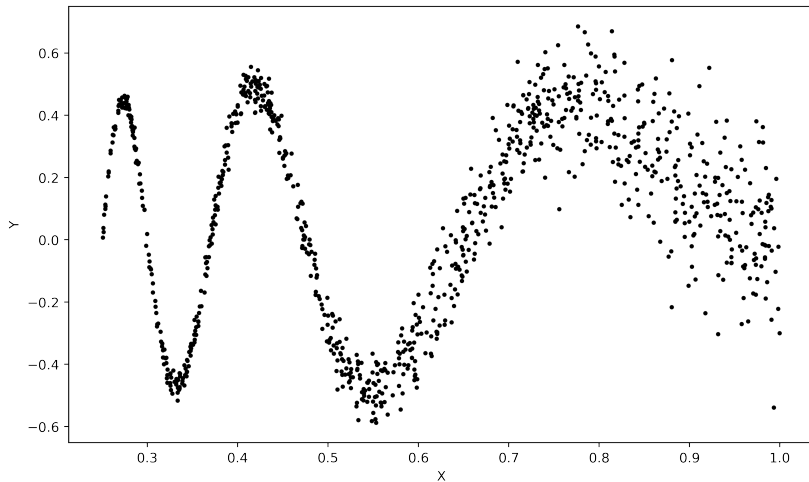
Teorema

Sea \hat{r}_n un regresor lineal y sea

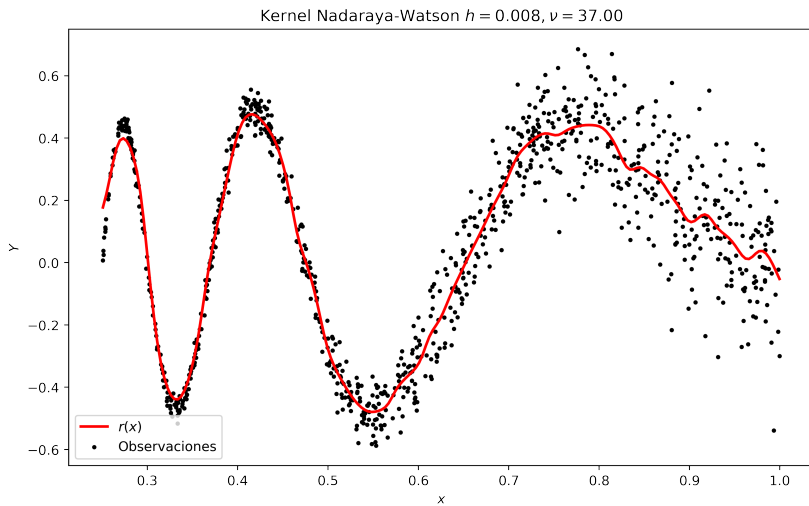
$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{r}_n(x_i))^2}{n - 2\nu + \tilde{\nu}}$$

Donde $\nu = \text{tr}(L)$ y $\tilde{\nu} = \text{tr}(L^T L)$ Si $\nu = o(n)$ y $\tilde{\nu} = o(n)$ entonces $\hat{\sigma}^2$ es un estimador consistente de σ^2

Sin embargo no siempre podemos asumir homocedasticidad...



Estimación de $\sigma^2(x)$



En este caso podemos asumir que la varianza esta en función de x , podemos suponer que

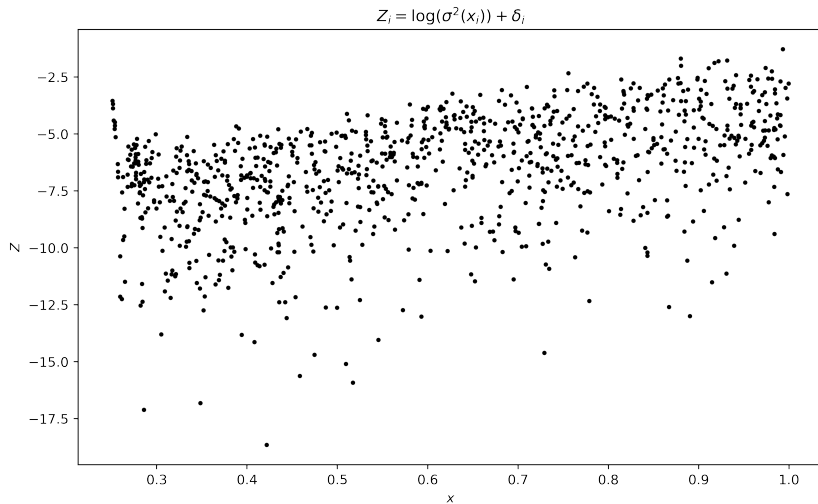
$$Y_i = r(x_i) + \sigma(x_i)\varepsilon_i$$

Con este modelo y teniendo una estimación de r , podemos hacer regresión para $\sigma(x)$. Si definimos $Z_i = \log(Y_i - r(x_i))^2$ y $\delta_i = \log \varepsilon_i$ entonces

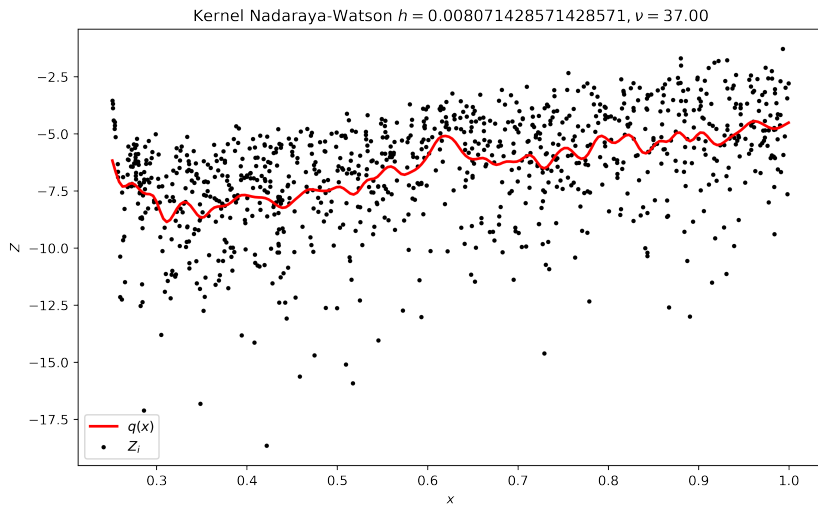
$$Z_i = \log(\sigma^2(x_i)) + \delta_i$$

Ahora podemos hacer regresión de Z_i con las x_i

Estimación de $\sigma^2(x)$



Estimación de $\sigma^2(x)$



Sin embargo al hacer la regresión y obtener $\hat{q}_n(x)$ solo tenemos que exponenciar ya que

$$\sigma^2(x) = e^{q(x)}$$

Estimación de $\sigma^2(x)$

