¿Cómo elegir h?

$$MSE(x) = E\left(\hat{f}(x) - f(x)\right)^2$$

$$= E\,\hat{f}(x)^2 - 2E\,\hat{f}(x)\,f(x) + E f^2(x) - \left(E\hat{f}(x)\right)^2 + \left(E\hat{f}(x)\right)^2$$

$$= \underbrace{E\left(\hat{f}^2(x) - (E\hat{f}(x)\right)^2}_{Var} - 2E\,\hat{f}(x)\,f(x) + \left(E\hat{f}(x)\right)^2 + f^2(x)$$

$$= Var(\hat{f}(x)) + \left(E\hat{f}(x) - f(x)\right)^2$$

$$= Var(\hat{f}(x)) + \left[Sesgo(\hat{f}(x))\right]^2$$

- Si $\underset{n\to\infty}{Sesgo} = 0$ $\quad \hat{f}(x)$ es insesgado

- Si $\underset{n\to\infty}{MSE(x)} = 0$ $\quad \hat{f}(x)$ es consistente

Puede mostrarse que KDE con un kernel Gaussiano es consistente

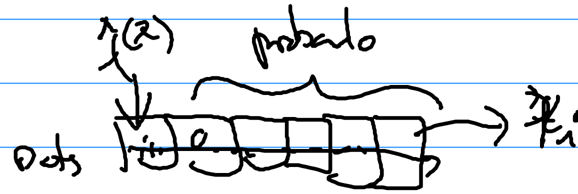
OTRO CRITERIO: ISE

$$ISE = \int \left(\hat{f}(x) - f(x)\right)^2 dx$$

$$MISE = \int E\left(\hat{f}(x) - f(x)\right)^2 dx$$

Observa que: (obtener h^*)

$$ISE = \underbrace{\int \hat{f}^2(x)\,dx}_{\text{fácil: método numérico}} - 2\int \underline{\hat{f}(x) f(x)}\,dx + \cancel{\int f^2(x)\,dx}$$



La otra integral

$$\int \hat{f}(x) f(x)\,dx = E\,\hat{f}(x)$$

Usando sólo los datos de la muestra la calculamos puedes usar CV-leave one-out, es decir, quitamos un dato cada vez y estimas la densidad.

<u>EL OBJETIVO</u> es que se utilicen todos los datos para estimar $\hat{f}(x)$ y para validarla

$$E\,\hat{f}(x) \approx \frac{1}{n}\sum_i \hat{f}_{-i}(x_i)$$

En el caso de KDE

$$\hat{f}_{-i}(x_i) = \frac{1}{(n-1)h}\sum_{j\neq i} K\left(\frac{x_i - x_j}{h}\right)$$

Entonces escogemos h^* que minimice

$$\boxed{\underbrace{\int \hat{f}^2(x)\,dx}_{\text{numéricamente}} - \frac{2}{n}\sum_{i=1}^{n} \hat{f}_{-i}(x_i)}$$