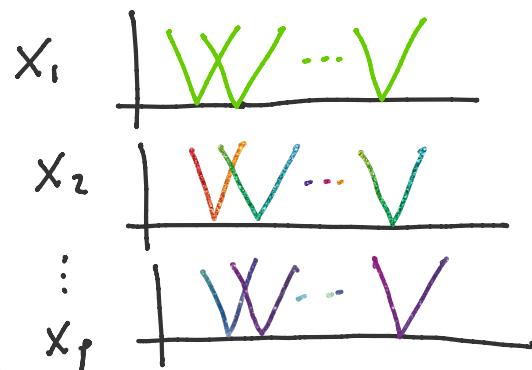


Reducir sesgo
over fit

I. Forward Stepwise Linear Regression

$$1. y_1 \sim 1 \rightarrow h_0 = 1$$

$$\hookrightarrow \hat{y}_1 = \beta_0 \cdot h_0$$



$$2. y_2 \sim 1 + h_0(x_i - t_i)_+ + h_0(t_i - x_i)_+$$

$$\hookrightarrow \hat{y}_2 = \beta_0 h_0 + \beta_1 h_1 + \beta_2 h_2$$

$$3. 3.1. y_3 \sim 1 + h_0(x_i - t_i)_+ + h_0(t_i - x_i)_+$$

$$+ h_0(x_j - t_j)_+ + h_0(t_j - x_j)_+$$

$$\hat{y}_3 = \beta_0 h_0 + \beta_1 h_1 + \beta_2 h_2 + \beta_3 h_3 + \beta_4 h_4$$

$(x_i - t_i)_+, (t_i - x_i)_+$
son las que MÁS
minimizan el error
de entrenamiento.

$$3.2. y_3 \sim 1 + h_0(x_i - t_i)_+ + h_1(x_j - t_j)_+$$

$$+ h_0(x_i - t_i)_+ + h_2(t_j - x_j)_+$$

$$+ h_0(x_i - t_i)_+ + h_0(t_i - x_i)_+$$

$$\hat{y}_3 = \beta_0 h_0 + \beta_1 h_1 h_3 + \beta_2 h_1 h_4 \} \quad \begin{matrix} \beta_0 \hat{h}_0 + \beta_1 \hat{h}_1 + \beta_2 \hat{h}_2 \\ + \beta_3 \hat{h}_3 + \beta_4 \hat{h}_4 \end{matrix}$$

$$+ \beta_3 h_1 + \beta_4 h_2 \}$$

$$3.3. \hat{y}_3 = \beta_0 h_0 + \beta_1 h_2 h_3 + \beta_2 h_2 h_4 \} \quad \begin{matrix} \beta_0 \hat{h}_0 + \beta_1 \hat{h}_1 + \beta_2 \hat{h}_2 \\ + \beta_3 \hat{h}_3 + \beta_4 \hat{h}_4 \end{matrix}$$

$$+ \beta_3 h_1 + \beta_4 h_2 \}$$

$$\vdots$$

$$M. \hat{y} = \beta_0 + \sum_{m=1}^M \beta_m \hat{h}_m \rightarrow \text{Criterio de paro}$$

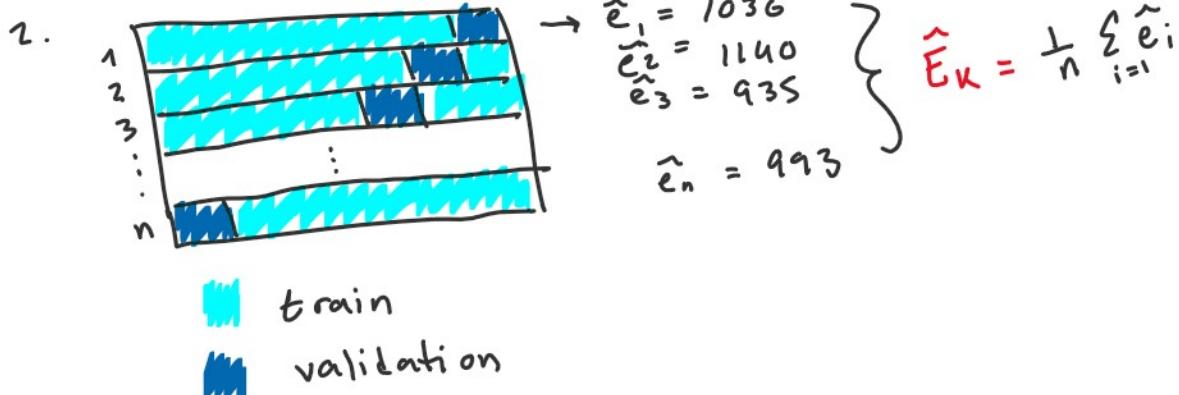
- Que \hat{h}_{M+1} ya no reduzca el error de entrenamiento considerablemente
- Tener a lo más $M < M'$ variables \hat{h}_m

II. Cross Validation \rightarrow Reducir variante

Vamos quitando \hat{h}_m por medio de cross validation, tal que la variable

que elijamos aumente MENOS el error de validación.

$$1. \hat{y} = \beta_0 + \beta_1 \hat{h}_1 + \dots + \beta_k \hat{h}_k + \dots + \beta_m \hat{h}_m$$



$$3. \hat{E}_1 = 940$$

$$\hat{E}_2 = 1002$$

$$\vdots$$

$$\hat{E}_l = 905 \rightarrow \hat{h}_l \text{ al omitirla } \hat{h}_{l+1} \text{ subió}$$

menos el error de validación \therefore la quitamos

$$\vdots$$

$$\hat{E}_M = 983$$

4. Repetimos desde 1. pero ahora con:

$$\hat{y} = \beta_0 + \beta_1 \hat{h}_1 + \dots + \beta_{l-1} \hat{h}_{l-1} + \beta_{l+1} \hat{h}_{l+1} + \dots + \beta_m \hat{h}_m$$

Criterio de paro {

- Que el error ya no cambie
- Significativamente
- Criterio del usuario
p.ej. $M < M''$

III Ejemplo de modelo final

$$\begin{aligned} \hat{y} &= \beta_0 + \beta_1 \hat{h}_1 + \underbrace{\beta_2 \hat{h}_2}_{\hat{h}_1} + \beta_3 \hat{h}_3 \\ &= 10 + 2 \cdot (\underbrace{x_1 - 3}_{\hat{h}_1}) + \underbrace{\beta_3 \hat{h}_3}_{\hat{h}_2} \\ &\quad + 5 \cdot (\underbrace{x_3 - 2}_{\hat{h}_1}) + (3 - x_4) + \\ &\quad - 2 \cdot (-1 - x_1) + (30 - x_{10}) + (x_{15} - 9) + \end{aligned}$$

$$+ 5 \cdot x^3 - \dots$$
$$- 3 \cdot (-1 - x_1) + (30 - x_{10}) + (x_{15} - 9) +$$
$$+ 4(x_{20} - 2) +$$
$$-$$