11-4-2025

## Γενικεύσεις κ' Επεκτάσεις Πολυωνυμικής Παλινδρόμησης

$Y = f(X) + \varepsilon$, $X$ ανεξ. μεταβλητή

$$f(x) = b_0 + b_1 \underline{x} + b_2 \underline{x^2} + \cdots + b_n \underline{x^n}$$

Γενικότερα

$$f(x) = \sum_{j=1}^{p} b_j h_j(x)$$

$h_j(x)$ : συναρτήσεις βάσης (basis functions)

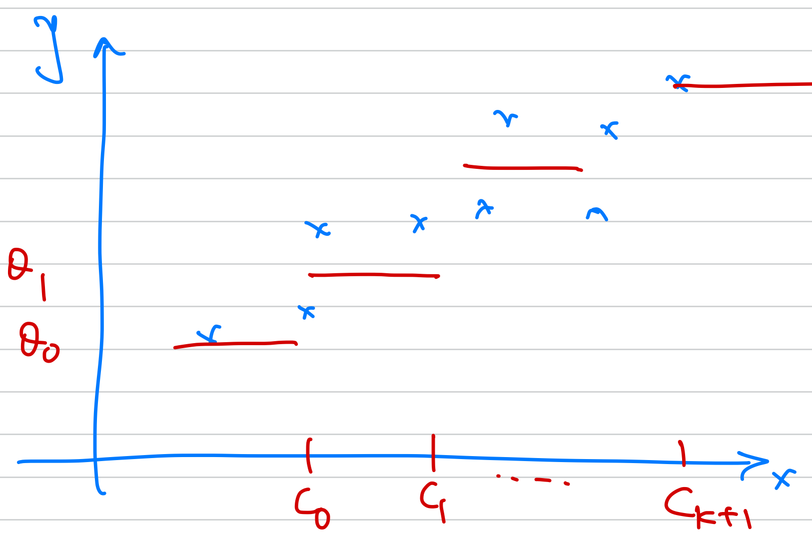
## Splines (piecewise polynomials – τμηματικά πολυωνυμικές)

ⓐ $X$ συνεχής τ.μ.

$$f(x) = \begin{cases} \theta_0, & x \le c_0 \\ \theta_1, & c_0 < x \le c_1 \\ \vdots & \\ \theta_K, & c_{K-1} < x \le c_K \\ \theta_{K+1}, & x > c_K \end{cases}$$

$c_0, \ldots, c_K$ σταθερά

$\theta_0, \theta_1, \ldots \theta_{K+1}$ άγνωστες παράμετροι.



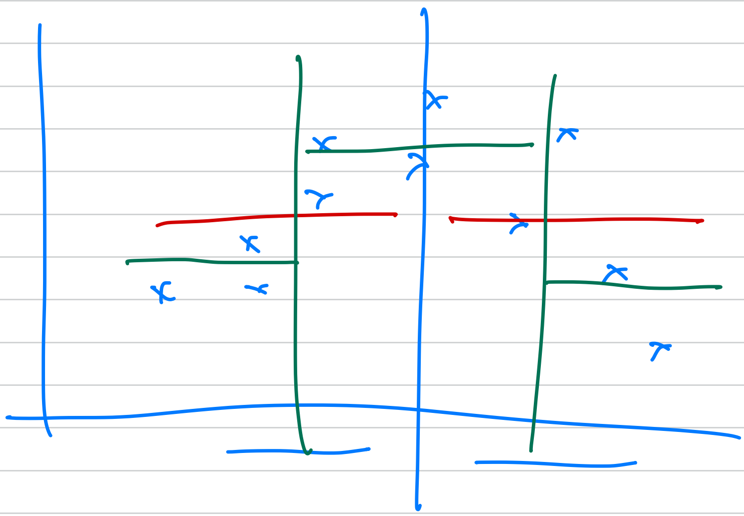
Ερώτημα: η $f(x)$ μπορεί να εκφραστεί ως

$$f(x) = \sum b_m h_m(x) \; ?$$

ποιές είναι οι συν. βάσης ?

$$f(x) = \underbrace{\theta_0 \cdot I(X \le c_0)}_{h_1(x)} + \underbrace{\theta_1 I(c_0 < X \le c_1)}_{h_2} + \dots + \underbrace{\theta_{k+1} I(X > c_k)}_{h_{k+2}}$$

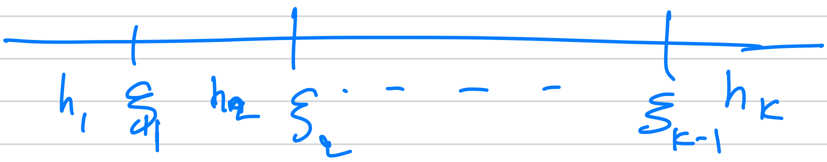
π.χ.



# Τμηματικά πολυωνυμικές συναρτήσεις

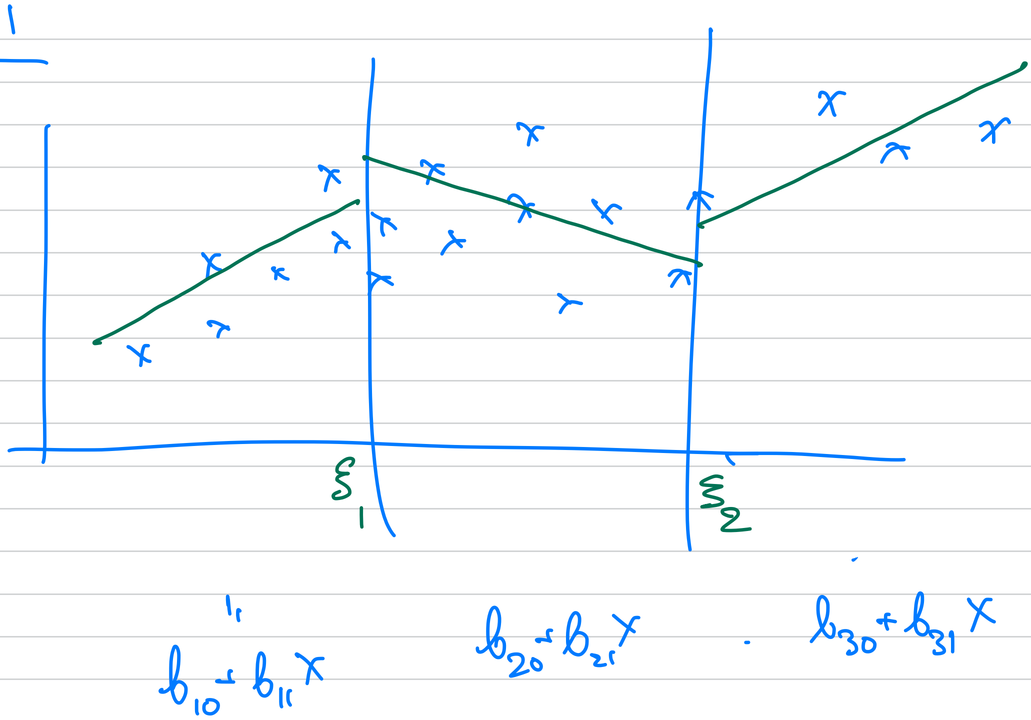
π.χ. $h_m(x) = b_{m0} + b_{m1}x + \dots + b_{mn}x^n$ , $m=1,2,\dots,K$

$$f(x) = \sum \theta_m h_m(x)$$




$$df_{model} = K \cdot (n+1)$$

π.χ. $n=1$



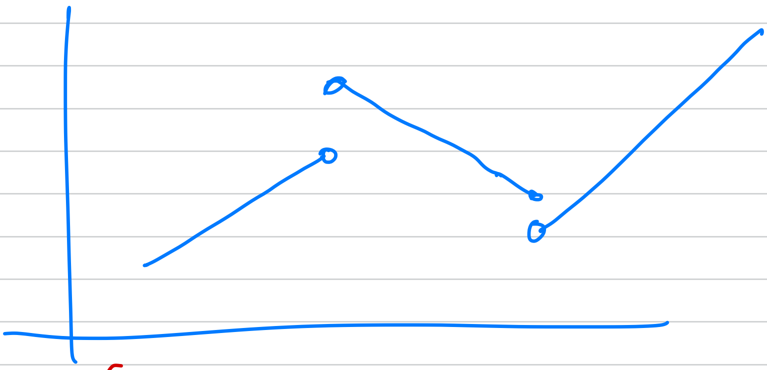

Απαίτηση : Η $f(x)$ συνεχής στα $x=\xi_1,\dots,\xi_K$.

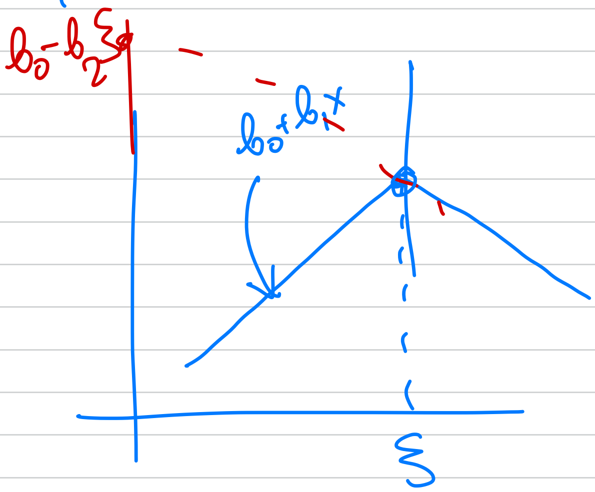
$b_{10}+b_{11}\xi_1 = b_{20}+b_{21}\xi_1$ ← (περιορ. ως προς $b$).

αφαιρούνται $K-1$ βαθμοί ελευθερίας

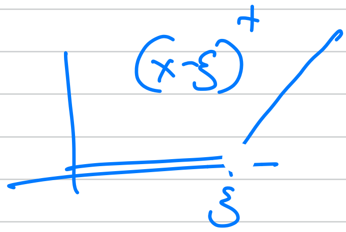
$$f(x) = \begin{cases} b_{00} + b_{10}x, & x < \xi_1 \\ \vdots & \\ \vdots & x > \xi_{K-1} \end{cases}$$

συνεχής στα $\xi_1, \dots, \xi_K$.





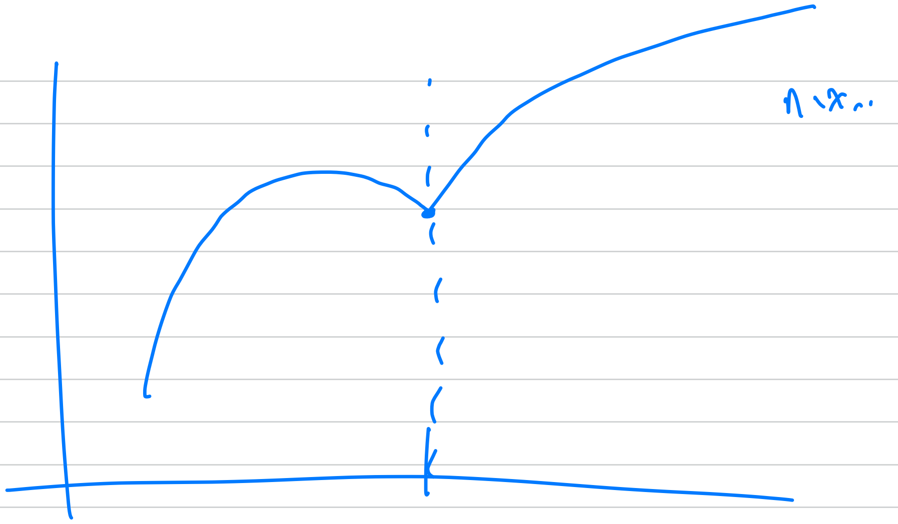

$$f(x) = b_0 + b_1 x + b_2 (x-\xi)^+$$




$$= \begin{cases} b_0 + b_1 x & x \leq \xi \\ (b_0 - b_2\xi) + (b_1 + b_2)x, & x > \xi \end{cases}$$

$$f(x) = b_1 h_1(x) + b_2 h_2(x) + b_3 h_3(x)$$

$$\left.\begin{aligned} h_1(x) &= 1 \\ h_2(x) &= x \\ h_3(x) &= (x-\xi)^+ \end{aligned}\right\}$$ linear spline function (spline of order $M=2$)



$$\xi$$

$$b_{10} + b_{11}x + b_{12}x^2 \qquad b_{20} + b_{21}x + b_{22}x^2$$

Συνθήκες : 1) $f$ συνεχής στα $\xi_1, \dots, \xi_{K-1}$

2) $f'(x)$ " " $\xi_1, \dots, \xi_{K-1}$

$$df = 6 - 2 = 4$$

$$f(x) = \underbrace{b_1 1 + b_2 x + b_3 x^2 +}\; b_4 \left((x-\xi)^+\right)^2$$

Για $x>\xi$

$$f'(x) = \underbrace{b_2 + 2b_3 x} + b_4 \cdot 2(x-\xi)$$

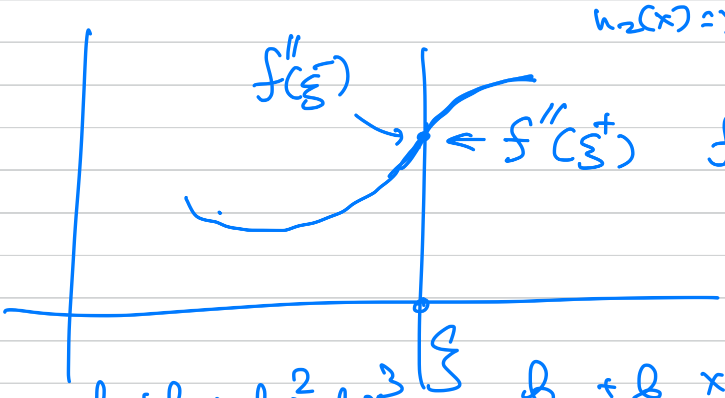
$$f'(\xi^+) = b_2 + 2b_3\xi = f'(\xi^-)$$

splines order $M=3$

---

Παρ. 3ου βαθμού (cubic splines).

$h_1(x)=1$, $h_2(x)=x$, $h_3(x)=x^2$, $h_4(x)=x^3$, $h_5(x)=\left((x-\xi)^+\right)^3$




$$\underbrace{f''(\xi^-) = f''(\xi^+)}$$

$$b_{10} + b_{11}x + b_{12}x^2 + b_{13}x^3 \qquad b_{20} + b_{21}x + b_{22}x^2 + b_{23}x^3$$

# συνέχεια έως παραγώγου $2^{ης}$ τάξης

## Γενικά

Spline τάξης $M+1$ με $K-1$ κόμβους $\xi_1, \dots, \xi_{K-1}$

σε κάθε διάστημα $f(x)$ πολυώνυμο βαθμού $M$

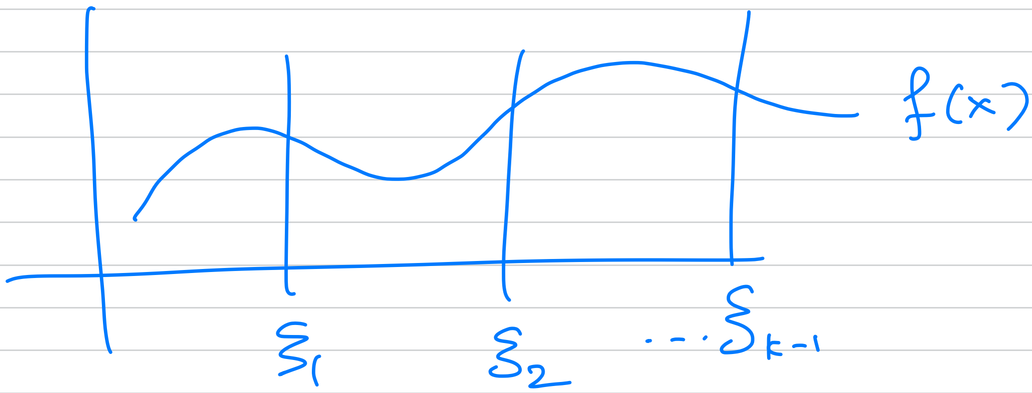
με περιορισμούς ότι είναι συνεχής

κ' έχει συνεχείς παραγώγους έως και $M-1$ τάξης

Cubic splines, τάξης 4

κυβικά πολυώνυμα

με συνεχείς 1' κ' 2 παραγώγους




Συναρτήσεις βάσης

$$h_1(x) = 1$$
$$h_2(x) = x$$
$$h_3(x) = x^2$$
$$h_4(x) = x^3$$
$$h_5(x) = (x-\xi_1)_+^3$$
$$h_6(x) = (x-\xi_2)_+^3$$
$$\vdots$$
$$h_{4+k-1}(x) = (x-\xi_{k-1})_+^3$$

$df_{mod} = k+3$ συναρτήσεις βάσης

regression splines

Υλοποίηση (σε R π.χ.)

| $Y$ | $X$ | $h_1(x)$ | $\dots$ | | $h_{K+3}(x)$ |
|---|---|---|---|---|---|
| | | $1$ | $x_1$ | $x^2$ | $(x-\xi_{K})_+^3$ |
| | | $\vdots$ | | $\vdots$ | $\vdots$ |
| | | $1$ | $x$ | | |

$$X$$

$$Y = Xb + \varepsilon$$

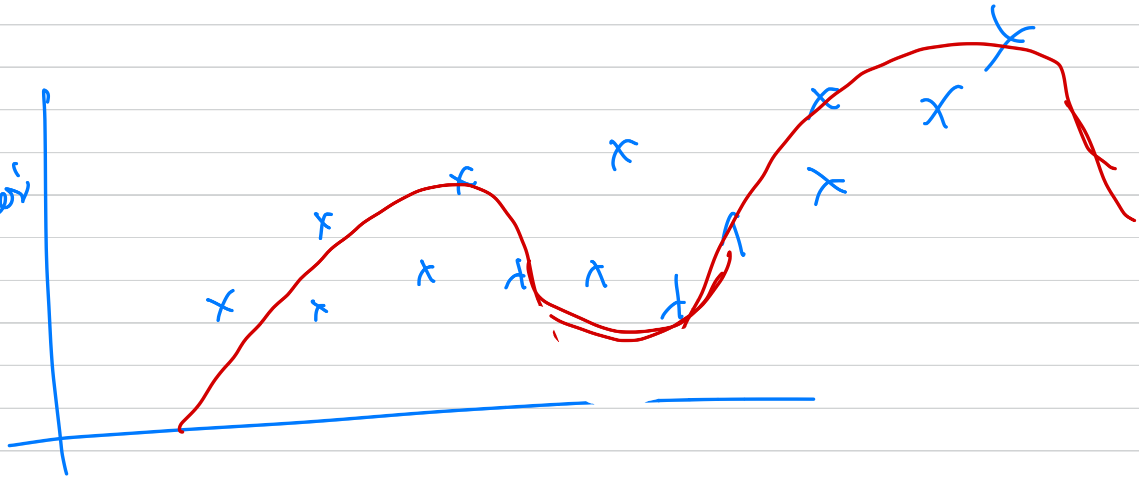
Splines library     bs ← (regression splines)

bs( x, df= , order=(4), degree=(3),
knots = c( . . . ) )

default για knots : χωρίζει το range(x)
σε ισαπέχοντα quantiles

## 1η επέκταση   Natural Splines

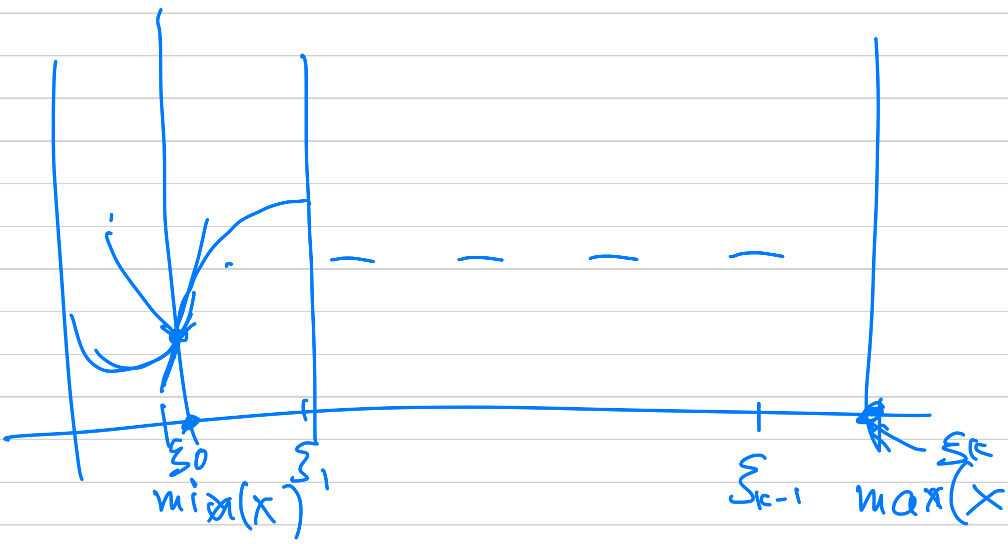
πχ.
πολ. 4ου βαθμού



**μεγάλο variance στα άκρα**

## Natural Splines

Επιπλέον περιορισμοί : Πέραν από τα δύο άκρα του range(x) η f(x) πρέπει να επεκτείνεται γραμμικά




$f''(\xi_0) = 0 \quad f''(\xi^k) = 0.$

## Συναρτήσεις βάσης

$\xi_1, \dots, \xi_K$.

$N_1(x) = 1, \quad N_2(x) = x,$

$$N_{2+k}(x) = d_k(x) - d_{K-1}(x), \quad k = 1, 2, \dots K-2$$

$$d_k(x) = \frac{(x-\xi_k)_+^3 - (x-\xi_K)_+^3}{\xi_K - \xi_k}, \quad k = 1, 2, \dots, K-1$$

R : συνάρτηση ns(        )

---

## Smoothing Splines

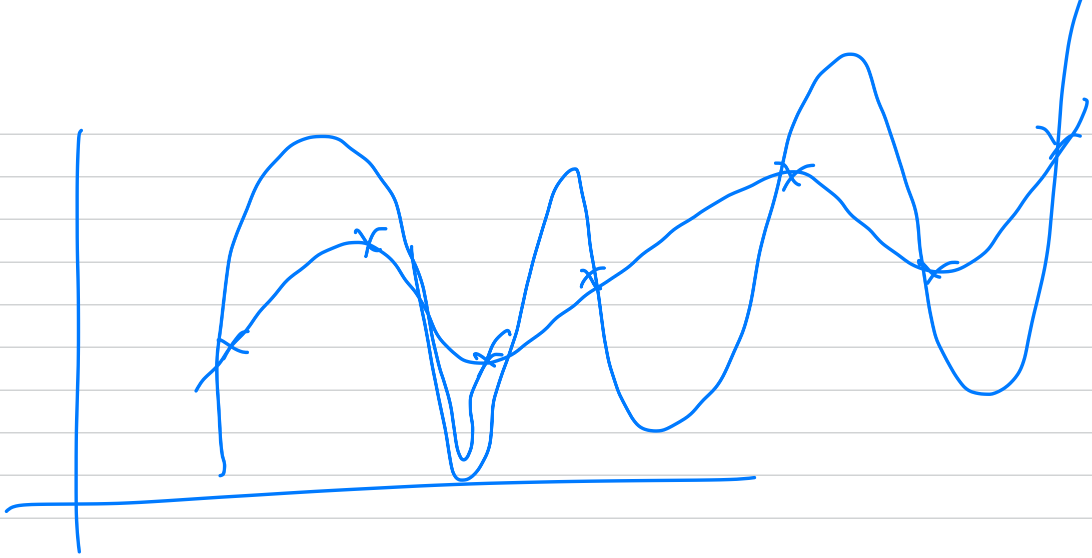
$y, x \qquad y = \underbrace{f(x)}_{} + \varepsilon$

$f(x) \in C^2$ (συνεχείς 2ες παράγωγοι)

$$\underbrace{RSS(f, \lambda)}_{\min(f)} = \sum_{i=1}^{n} \left(y_i - f(x_i)\right)^2 + \lambda \underbrace{\int \left(f''(t)\right)^2 dt}_{\text{διακύμανση της } f''(x)}$$

Για $\lambda = 0$ :

οποιαδήποτε $f(x) \in C^2 : f(x_i) = y_i$



Για $\lambda \to \infty$ $\quad f''(t) \approx 0 \quad \forall t$

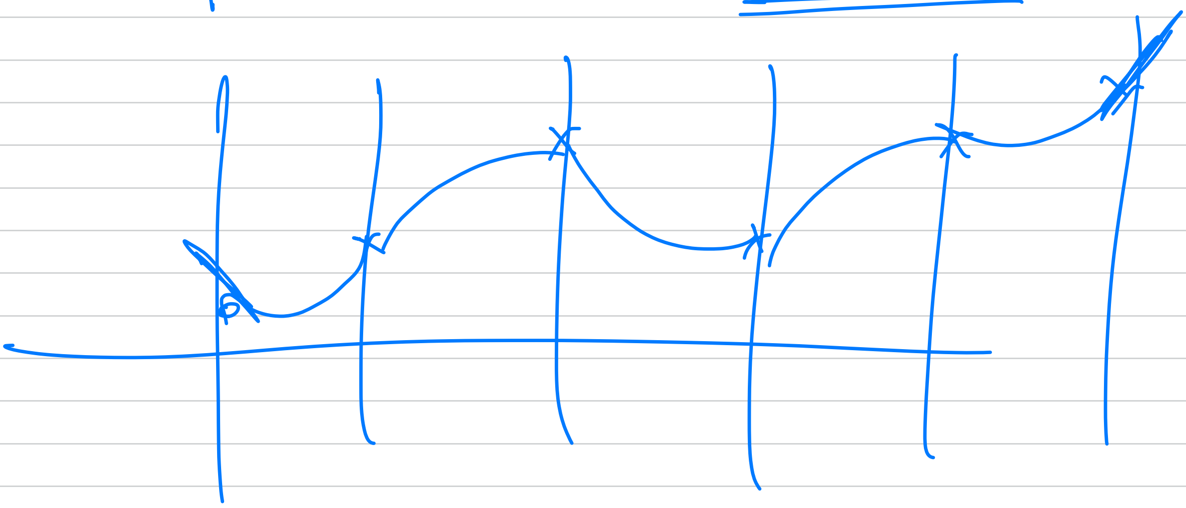
$\Longrightarrow$ $f(t)$ "γραμμική"

$$\min_{f \in C^2} \{ RSS(f, \lambda) \} \quad ?$$

<u>Λύση</u> Ένα natural spline

$$f(x) = \sum_{j=1}^{N} \theta_j N_j(x)$$

$N_j$ : natural cubic splines με κόμβους στα σημεία <u>$x_1, \ldots, x_n$</u>



## Περισσότερες ανεξ. μεταβλητές $X_1, \dots, X_p$.

### 1) Generalized Additive Models (GAM)

$$f(x_1, \dots, x_p) = \sum_{j=1}^{p} \theta_j \, f_j(x_j)$$

$f_j(x_j)$ : μοντελοποίηση της $x_j$
(π.χ. με splines, etc).

$$f_j(x_j) = \sum_{i=1}^{m_j} b_{ji} \underbrace{h_{ji}(x_j)}_{\text{συν. βάσης της } f_j(x_j)}$$

### Παρένθεση (Αλληλεπιδράσεις - Interactions)

$X_1 = \text{age}$

$X_2 = \text{smoking} \begin{cases} \nearrow \text{yes} \\ \searrow \text{no} \end{cases} \Rightarrow S = I(X_2 = \text{yes})$

$Y$ = αναπν. ικανότητα

$$y = b_0 + b_1 X_1 + b_2 S$$

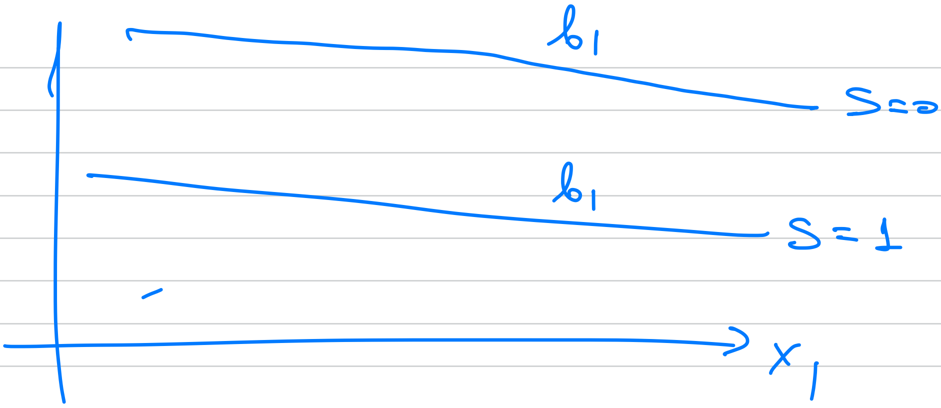
main effects model
(μοντέλο κύριων επιδράσεων)

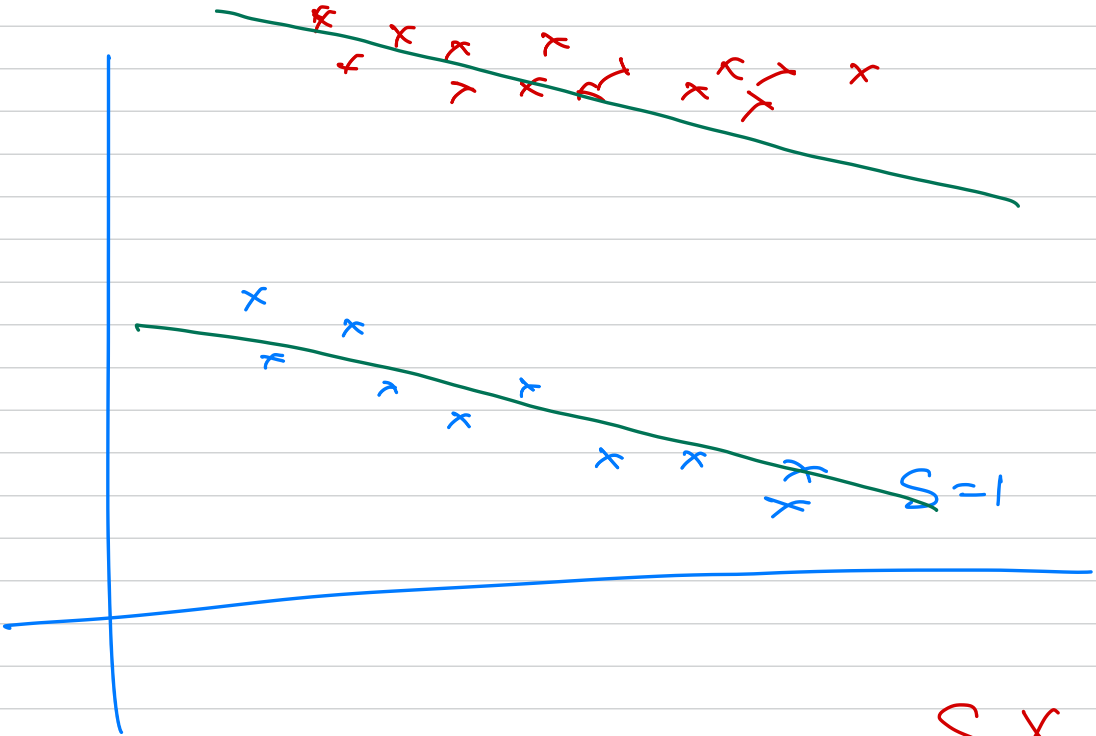
$$\left.\begin{aligned} S=0 &: \quad y = b_0 + b_1 X_1 \\ S=1 &: \quad y = (b_0 + b_2) + b_1 X_1 \end{aligned}\right\}$$

$b_2$ = επίπτωση $S$ ανεξ. από το $X_1$

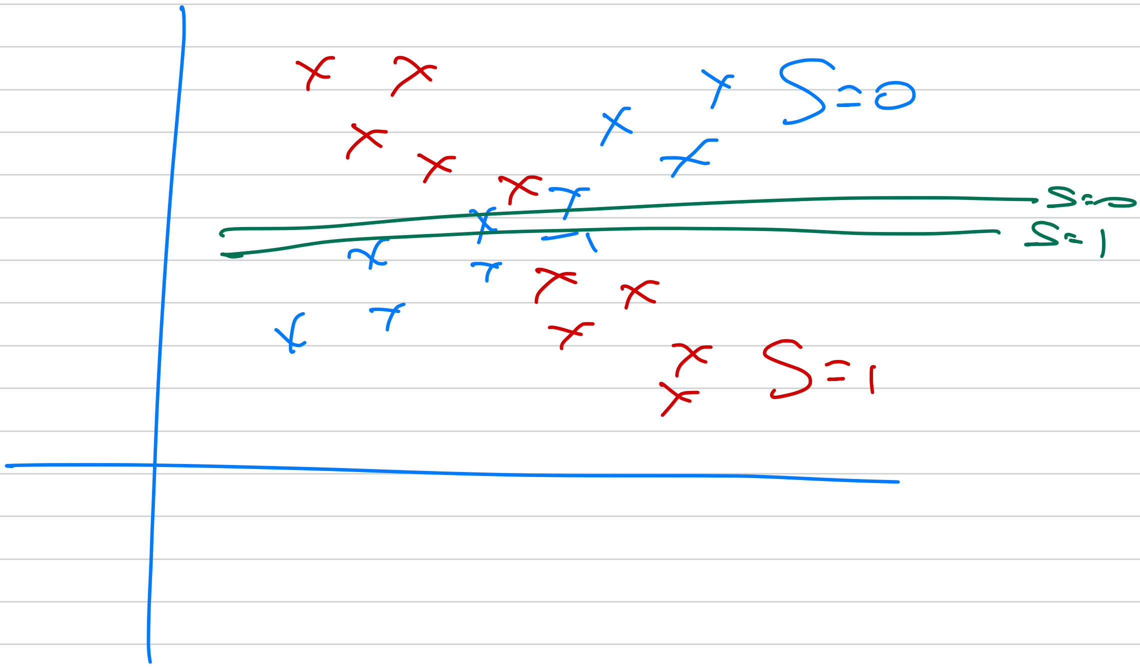
$E(y \mid S=1) - E(y \mid S=0) \quad \forall x_1$

$b_1 = E(y \mid x_1 + 1) - E(y \mid x_1) \quad \forall S = 0, 1$




S, X assoc.

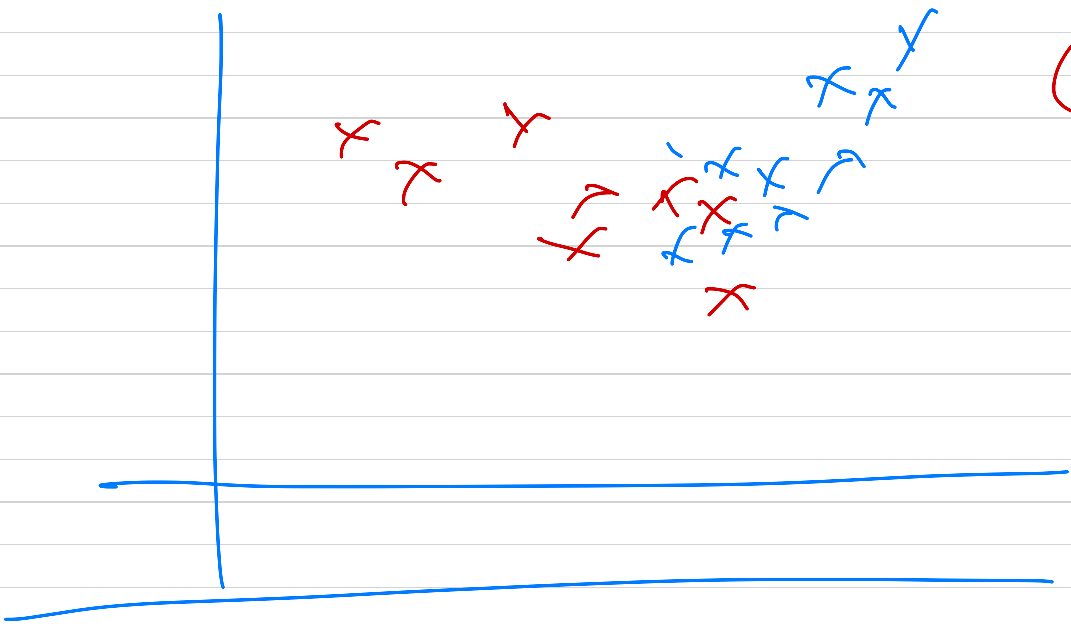


$$f(x) = b_0 + b_1 x_1 + b_2 S + b_3 \cdot \underbrace{x_1 \cdot S}$$

$S=0 \quad f(x) = b_0 + b_1 x_1$

$S=1 \quad f(x) = (b_0 + b_2) + (b_1 + b_3) x_1$

$b_3$ : interaction effect
αλληλεπίδραση



$(S, X)$ συσχετ

$x_1, x_2$ ποσοτικές

$$E y = b_0 + b_1 x_1 + b_2 x_2 = f(x_1, x_2)$$

$$\frac{\partial f}{\partial x_1} = b_1 \quad \forall x_2 \qquad \frac{\partial f}{\partial x_2} = b_2 \quad \forall x_1$$

$$f(x) = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2$$ interaction

$$\frac{\partial f}{\partial x_1} = b_1 + b_3 x_2 \qquad \frac{\partial f}{\partial x_2} = b_2 + b_3 x_1$$

# Γραμμικό Μοντέλο

$H_0$ : ∄ αλληλεπ. $\quad$ $H_1$ : ∃ αλληλεπίδραση

$H_0 : b_3 = 0$ $\quad$ $H_1 : b_3 \neq 0$.

t-test