

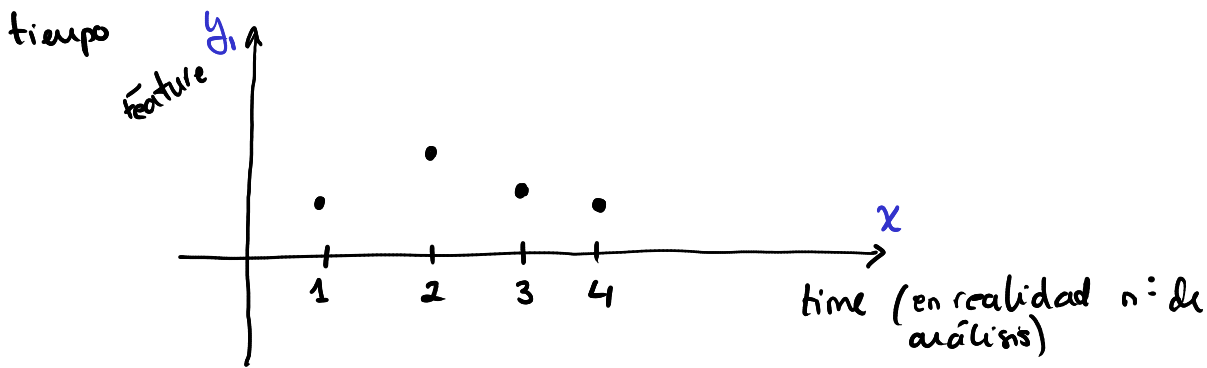
Vamos a capturar dinâmicas temporales com regressão linear.

- La idea es:

1) Para cada paciente vamos a utilizar como features:

a) Las que vienen en la base de datos por el primer análisis

b) Para cada feature: tenemos que es una función del



Vamos a capturar la evolución ajustando un polinomio de la siguiente forma

nota: el subíndice  $y(i)$  - hace referencia al feature.

$$\hat{y}_i = \hat{\alpha}_1 x_i^2 + \hat{\beta}_1 x_i + \hat{\delta}_i$$

- La idea es usar least-squares (regresión lineal) para estimar  $\{\hat{\alpha}_1, \hat{\beta}_1, \hat{\delta}_1\}$   
 ↪ ojo esto es d.f. de  $X$

De esta forma cada paciente tendría como vector  $\underline{x}$  de features

$X = \begin{bmatrix} \vdots \\ \vdots \\ \vdots \end{bmatrix}$  variables del primer hemograma.

Variables temporales: tantas como  $3 \cdot N$  siendo  $N$  el número de variables del hemograma.

NOTA: Como estimar  $\alpha, \beta, \delta$ . Como tenemos pocas muestras (4 en el mejor caso) podríamos probar con bootstrap.

Lo que te decía es: vamos a comenzar con least-squares (quizá con regularización) a palo seco y revisemos los pacientes.

En cuanto a comparación de modelos; yo tenía las siguientes comparaciones (para que no tengamos miles de cosas para comparar)

$M_1 \rightarrow$  Boosting con datos de entrada  $X = \begin{bmatrix} \bullet_1 \\ \bullet_2 \\ \bullet_3 \\ \vdots \\ \bullet_N \end{bmatrix}$  Variables del primer hemograma

$\hookrightarrow$  Features selection para modelo sparse:  $M_{1,a}$  que nos quedamos con  $\alpha_a = \begin{bmatrix} \bullet_2 \\ \bullet_3 \\ \vdots \\ \bullet_{N_a} \end{bmatrix}$  con  $N_a < N$ .

De este modelo  $M_{1,a}$  sacamos performance  $\begin{cases} \text{acc} \\ \text{Sen} \\ \text{Spe} \end{cases}$

$M_2 \rightarrow$  Boosting con datos de entrada  $X = \begin{bmatrix} \bullet_1 \\ \bullet_2 \\ \vdots \\ \bullet_N \\ \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \vdots \\ \gamma_N \end{bmatrix}$  Variables primer hemograma

Dudas: quizá podríamos cortar todas las variables y así eliminar todos los  $\gamma_i$ .

Variables de tendencia.

Dudas: i) No sé si es mejor hacer  $M_2$  a partir de  $M_{1,a} \rightarrow$  es decir de una features selection previa o hacer features selection desde  $M_2$  con todos las variables?;

$\hookrightarrow$  Haces feature selection  $M_{2,a}$  de tal forma que

$$X_a = \begin{bmatrix} \bullet_3 \\ \bullet_N \\ \beta_3 \\ \gamma_4 \\ \alpha_5 \\ \alpha_6 \end{bmatrix}$$