FoDA L12

High-dimensional and Polynomial (Linear) Regression

Sep 29, 2022

Input

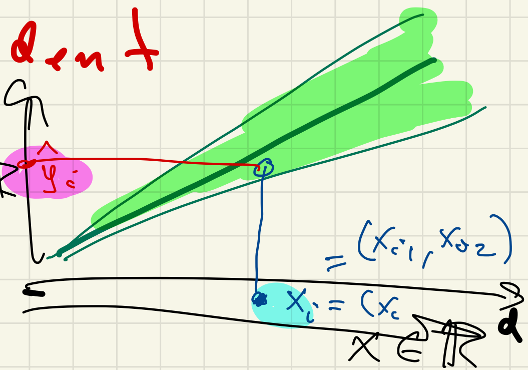
$$(X, y) \subset \mathbb{R}^d \times \mathbb{R}$$

$$(X, y) = \{(x_1, y_1), (x_2, y_2), \dots (x_n, y_n)\}$$

explanatory dependent

$x_i \in \mathbb{R}^d$ $y_i \in \mathbb{R}$

$$x_i = (x_{i1}, x_{i2}, \dots x_{id})$$



Model

$$\hat{y}_i = M_\alpha(x_i) = \alpha_0 + \sum_{j=1}^{d} \alpha_j x_{ij}$$

$$= \alpha_0 1 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \dots + \alpha_d x_{id}$$

$$\alpha = (\alpha_0, \alpha_1 \dots \alpha_d) \in \mathbb{R}^{d+1}$$

$$= \langle \alpha, \underbrace{(1, x_i)}_{\in \mathbb{R}^{d+1}} \rangle = \langle \alpha, (1, x_{i1}, x_{i2}, \dots x_{id}) \rangle$$

	time: X_1	jiggle: X_2	scroll: X_3	sales: y
1	232	33	402	2201
1	10	22	160	0
1	6437	343	231	7650
1	512	101	17	5599
1	441	212	55	8900
1	453	53	99	1742
1	2	2	10	0
1	332	79	154	1215
1	182	20	89	699
1	123	223	12	2101
1	424	32	15	8789

x_1 y_1

X_1 X_2 X_3 y

$$\tilde{X} = [1 ; X_1 ; X_2 ; X_3] \in \mathbb{R}^{n \times d+1} = \mathbb{R}^{10 \times 4}$$

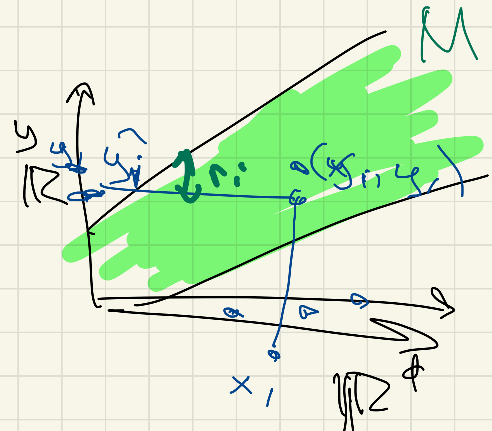
Input $X \in \mathbb{R}^{n \times d}$ $\quad y \in \mathbb{R}^n$

Goal $SSE((X,y), M_\alpha) = S(\alpha)$

$$= \sum_{i=1}^{n} (y_i - M_\alpha(x_i))^2 = \sum_{i=1}^{n} (r_i)^2$$

$$S(\alpha) = \sum_{i=1}^{n} (y_i - \langle \alpha, (1, x_i) \rangle)^2$$

$$\alpha^* = \underset{\alpha \in \mathbb{R}^{d+1}}{\operatorname{argmin}} \; SSE((X,y), M_\alpha)$$



$$\alpha^* = \underset{\alpha \in \mathbb{R}^{d+1}}{\operatorname{argmin}} \; S(\alpha)$$

$$\hookrightarrow \alpha^* = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T y$$

$$\alpha^* = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T y$$

$$a = \frac{\langle x, y \rangle}{\langle x, x \rangle}$$

$\tilde{X} \in \mathbb{R}^{n \times (d+1)}$

$\tilde{X}^T \tilde{X} \in \mathbb{R}^{(d+1) \times (d+1)}$ $\Leftarrow$ square

full rank?

usually $n >> d$

$(\tilde{X}^T \tilde{X})^{-1} \in \mathbb{R}^{(d+1) \times (d+1)}$

$\tilde{X}^T \in \mathbb{R}^{(d+1) \times n}$

$y \in \mathbb{R}^n$

easier as input

$$X = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \\ 1 & x_n \end{pmatrix} \in \mathbb{R}^{n \times d}$$

$x_i \in \mathbb{R}^{d+1}$

Why $\alpha^* = (X^T X)^{-1} X^T y$?

why $\alpha^* = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T y$

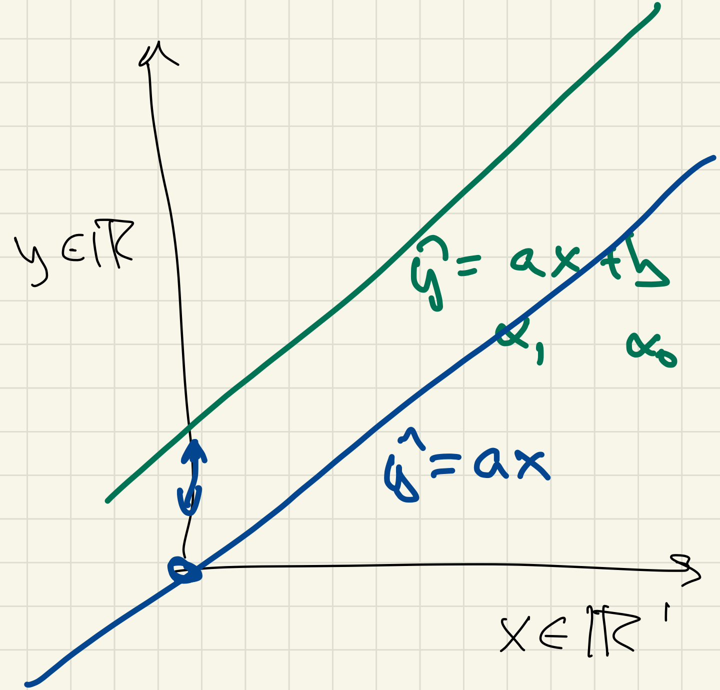
if $\{x_1, x_2, \ldots, x_n\} \in \mathbb{R}$

$\tilde{x}_1 = (1, x_1) \in \mathbb{R}^2$

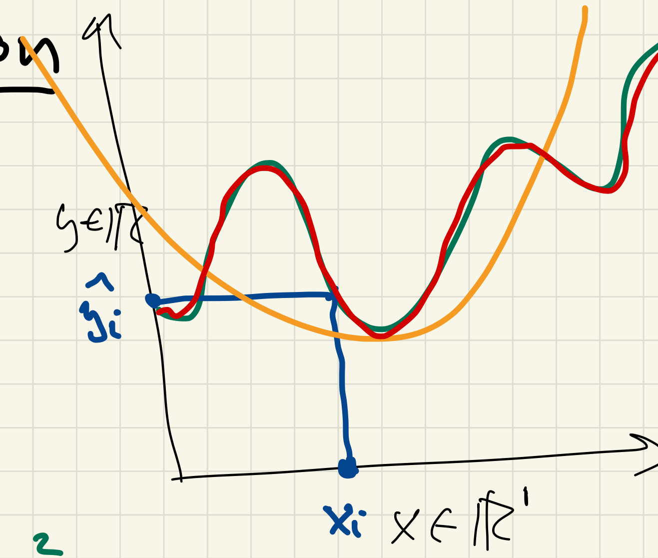
$$\tilde{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \\ 1 & x_n \end{bmatrix} \qquad y \in \mathbb{R}^n$$

$\alpha = (b, a) = (\tilde{X}^T \tilde{X})^{-1} \tilde{X}^T y$

$$y = \cancel{\alpha_0} + \sum_{j=1}^{d} \alpha_j \, x_{i,j}$$



Polynomial Regression



Input $(x, y) \in \mathbb{R} \times \mathbb{R}$

$= \{(x_i, y_i)\}_i \qquad x_i \in \mathbb{R}$

$y_i \in \mathbb{R}$

$$\hat{y} = M_{p=2}(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2$$

$$\hat{y} = M_p(x) = \alpha_0 + \alpha_1 x + \dots + \alpha_p x^p$$

$$= \alpha_0 + \sum_{j=1}^{p} \alpha_j x^j = \sum_{j=0}^{p} \alpha_j x^p$$

$$= \langle \alpha, (1, x, x^2, \dots, x^p) \rangle$$

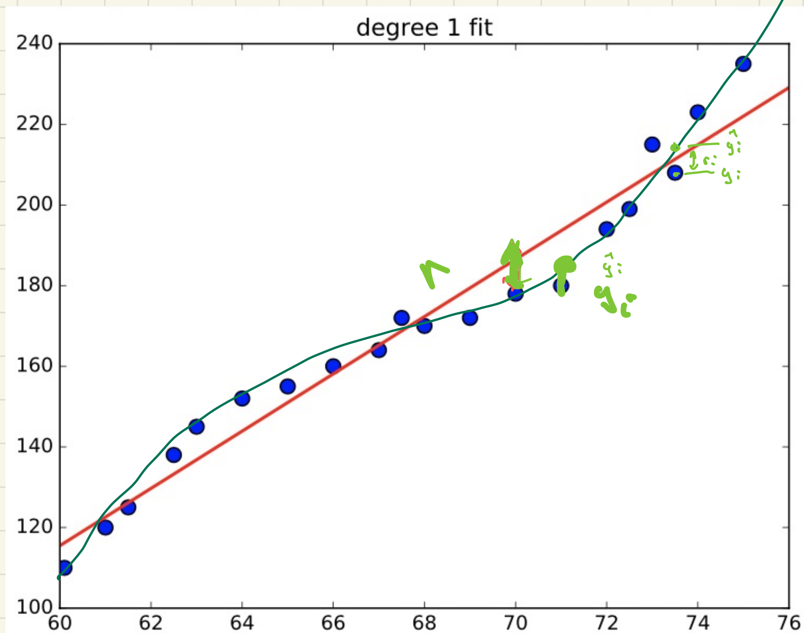
Goal

$\alpha_p^* \in \mathbb{R}^{p+1}$

$= \operatorname{argmin}_{\alpha} SSE((x, y), M_{\alpha})$

height (in)	weight (lbs)
66	160
68	170
60	110
70	178
65	155
61	120
74	223
73	215
75	235
67	164
61.5	125
73.5	208
62.5	138
63	145
64	152
71	180
69	172
72.5	199
72	194
67.5	172

$$SSE((X, b), M_{p,\alpha}) = \sum_{i=1}^{n} \left(\underbrace{y_i - M_{p,\alpha}(x_i)}_{r_i} \right)^2$$



Input $(x, y) \in \mathbb{R}^n \times \mathbb{R}^n$

Model $M_{p,\alpha}(x) = \sum_{j=0}^{p} \alpha_j x^j$ $\quad \alpha = (\alpha_0, \alpha_1, \ldots, \alpha_p)$

$\hookrightarrow \mathbb{R}$

Goal $\alpha^* = \underset{\alpha \in \mathbb{R}^{p+1}}{\operatorname{argmin}} \sum_{i=1}^{n} \left(y_i - M_{p,\alpha}(x_i)\right)^2$

$$\tilde{X}_p = \begin{bmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^p \\ 1 & x_2 & x_2^2 & \cdots & x_2^p \\ & & & & \\ 1 & x_n & x_n^2 & \cdots & x_n^p \end{bmatrix}$$

$\in \mathbb{R}^{n \times p+1}$

$v_1 \in \mathbb{R}^{p+1}$

$v_2 \in \mathbb{R}^{p+1}$

$M_{p,\alpha}(x_i) = \langle \alpha, v_i \rangle$

$$\alpha^* = (\tilde{X}_p^T \tilde{X}_p)^{-1} \tilde{X}_p^T y$$

$$X = \begin{bmatrix} 2 \\ 4 \\ 3 \end{bmatrix} \qquad y = \begin{bmatrix} 1 \\ 6 \\ 5 \end{bmatrix}$$

$p = 5$

$$\tilde{X}_p = \begin{bmatrix} 1 & 2 & 4 & 8 & 16 & 32 \\ 1 & 4 & 16 & 64 & 256 & 1024 \\ 1 & 3 & 9 & 27 & 81 & 243 \end{bmatrix}$$

$$\in \mathbb{R}^{3 \times 6}$$