FoDA L11

Linear Regression

explanatory & dependent variables

Sep 27, 2022

Input

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

$X = \{x_1, x_2 \dots, x_n\} \subset \mathbb{R}^{d=1}$

explanatory variable

$y = \{y_1, y_2, \dots y_n\} \subset \mathbb{R}$

dependent variable

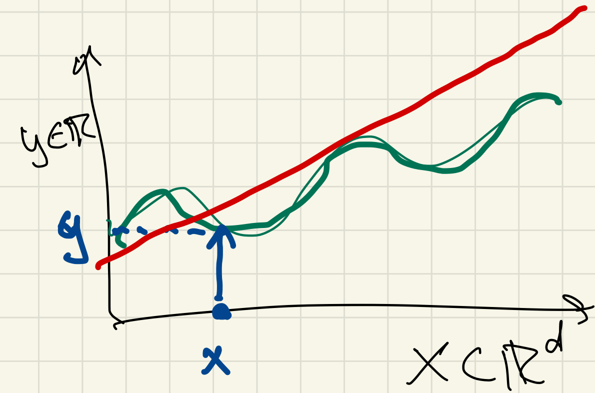
Goal

learn a function

$$f(x) \to y$$

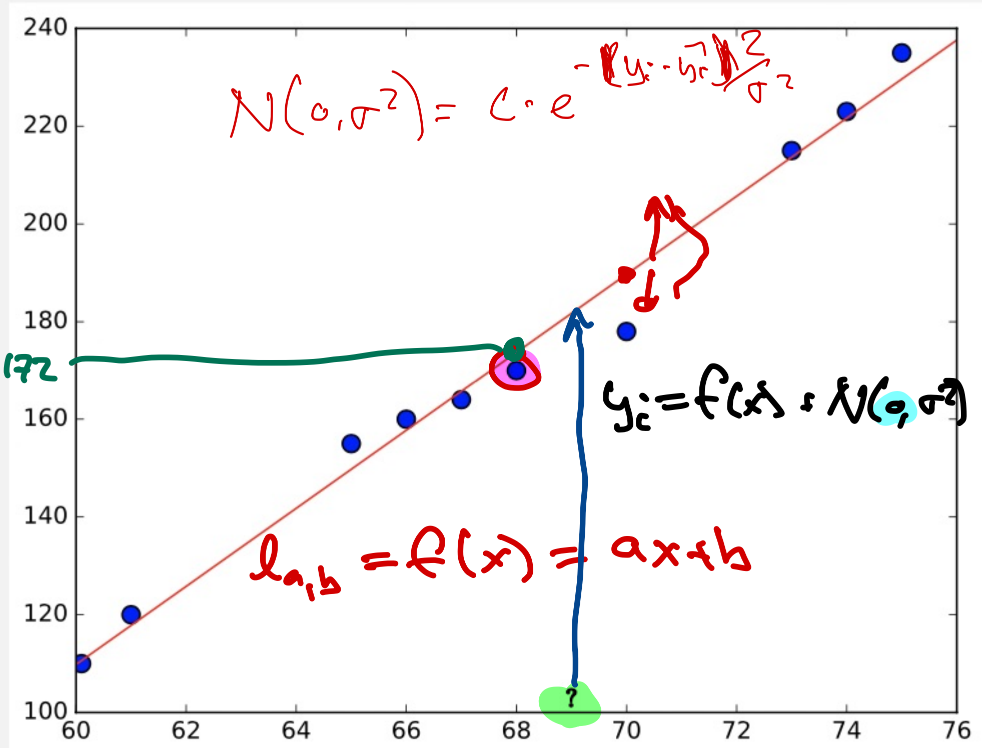
linear function

$$f(x) = ax + b \to y$$



X y

height (in)	weight (lbs)
66	160
68 x_2	170 y_2
60	110
70	178
65	155
61	120
74	223
73	215
75	235
67	164
69	?



$N(0,\sigma^2) = c \cdot e^{-(y_i - \hat{y}_i)^2/\sigma^2}$

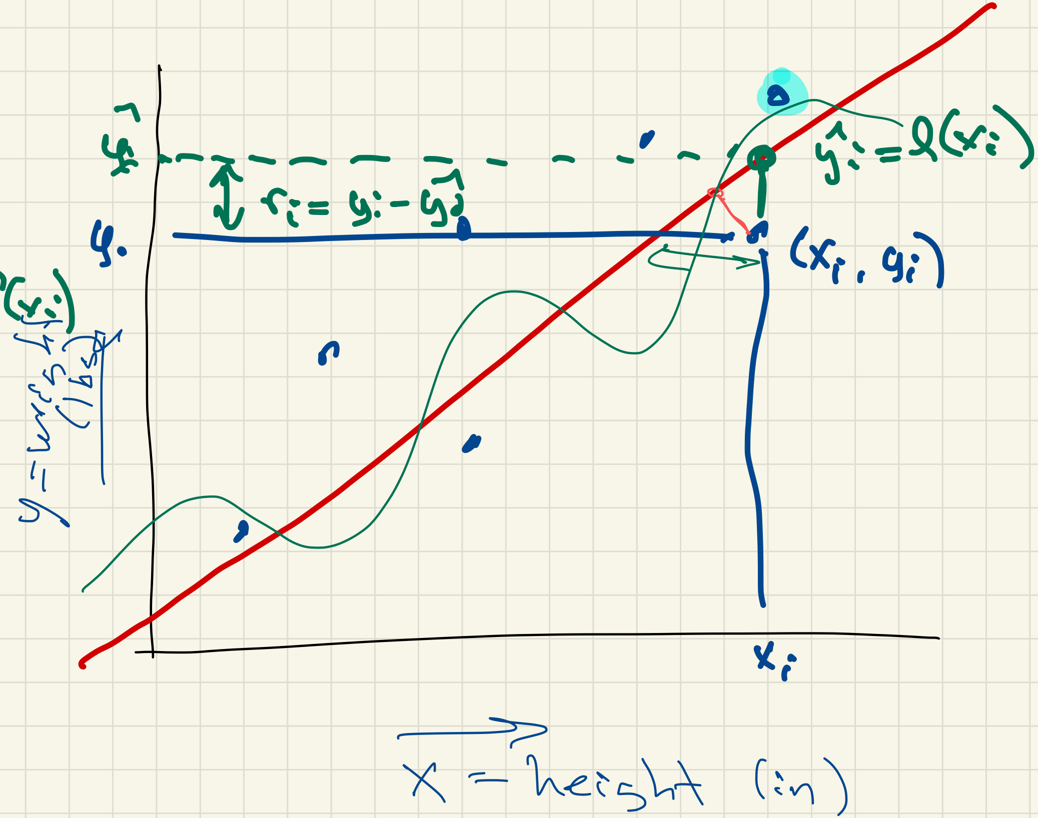
$y_i = f(x) + N(0,\sigma^2)$

$l_{a,b} = f(x) = ax+b$

residual

$$r_i = y_i - \hat{y}_i$$

$$= y_i - l(x_i) = y_i - f(x_i)$$



Cost of a function f (or ℓ) on (X, y)

Sum of Squared Errors

$$SSE((x, y), \ell) = \sum_{i=1}^{n} (y_i - \ell(x_i))^2 = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} r_i^2$$

- looks like variance
- MLE of all lines $\rightarrow$ SSE
- $SSE((x, y), \ell) = \sum_{i=1}^{n} r_i^2 = \|r\|^2$

  $r = (r_1, r_2, \ldots r_n)$
- simple "closed form" soln.

Root Mean Squared Error

RMSE

$$RMSE((x_i, y_i), \ell) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \ell(x_i))^2}$$

root

mean

fixes units

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

Goal line $\underbrace{l : ax+b}_{\text{model}}$ $l_{a,b}(x) \to y$

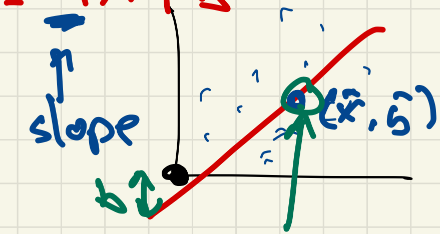
minimize $SSE((x,y), l_{a,b})$

$$= \underbrace{\sum_{i=1}^{n} (y_i - (ax_i + b))^2}_{\text{cost function}}$$

parameters

$$a^*, b^* = \underset{(a,b)\in\mathbb{R}^2}{\operatorname{argmin}} \; SSE((x,y), l_{a,b})$$

$l_{a,b}(x) = ax + b$



Solve a^*, b^*

1. $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ $\qquad$ $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

2. $\tilde{X} = \{(x_1 - \bar{x}), (x_2 - \bar{x}), \dots, (x_n - \bar{x})\}$

   $\tilde{y} = \{(y_1 - \bar{y}), \dots (y_n - \bar{y})\}$

3. $a^* = \frac{\langle \tilde{y}, \tilde{x} \rangle}{\|\tilde{x}\|^2} = \frac{\|\tilde{y}\| \cdot \cancel{\|\tilde{x}\|} \cdot \cos \theta_{\tilde{x},\tilde{y}}}{\|\tilde{x}\|^{\cancel{2}}}$

4. $b^* = \bar{y} - a^* \bar{x}$ $\qquad$ $(\bar{y} = a^* \bar{x} + b^*)$