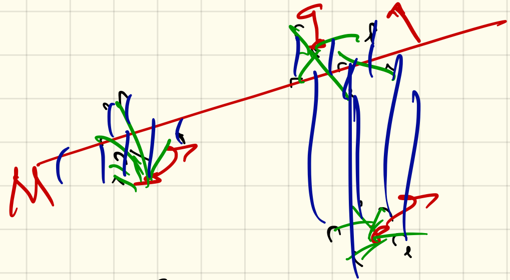
Regression



linear least squares

regression	classification	supervised learning
dim reduction	clustering	unsupervised "

outcome scalar

outcome set

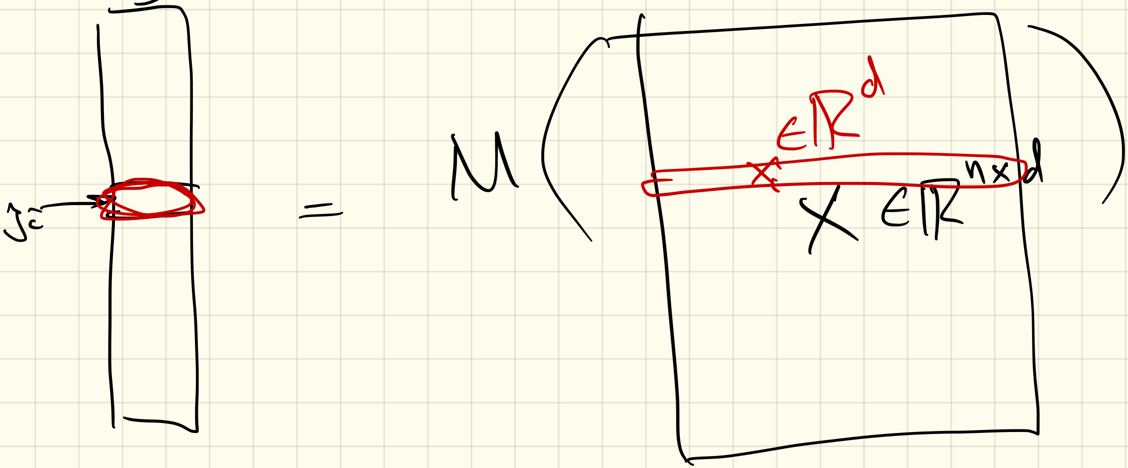
Data input X, y.

one data point $(x_i, y_i) \in (X, y)$

X: explanatory variable

y: dependent var.

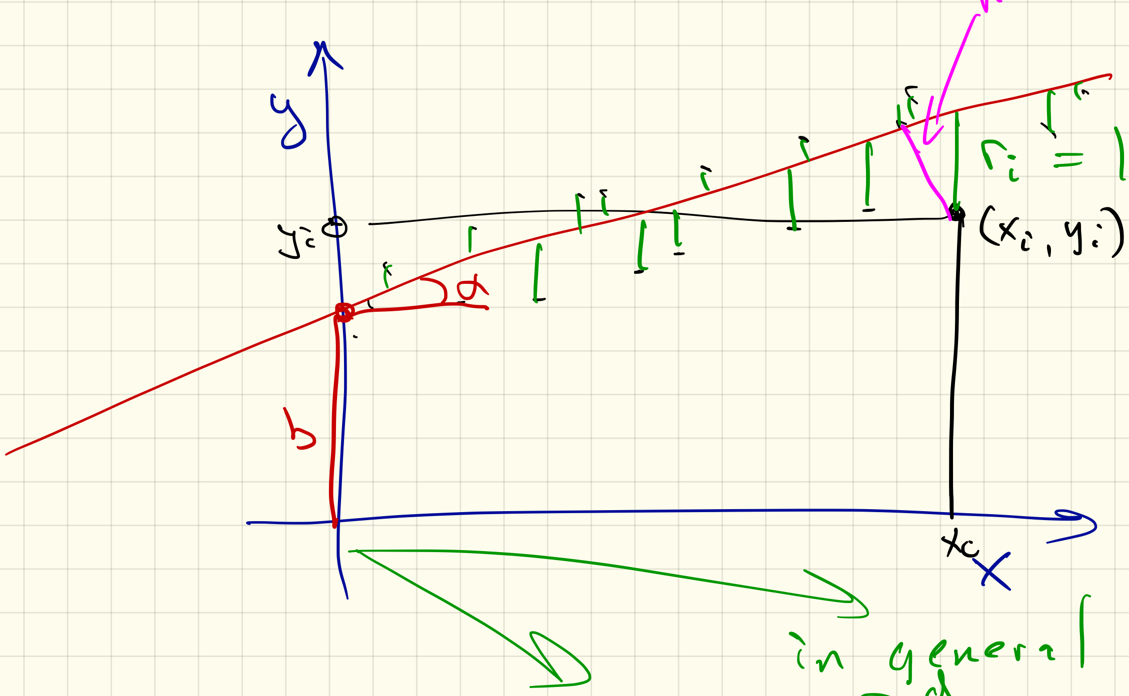
$y \in \mathbb{R}^n$



$$M_\alpha(x_i) = \langle \alpha, x_i \rangle + b$$

$$\alpha = (\alpha_1, \alpha_2 \dots \alpha_d)$$

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



$r_i = |y_i - M_\alpha(x_i)|$

Find α, b

minimize

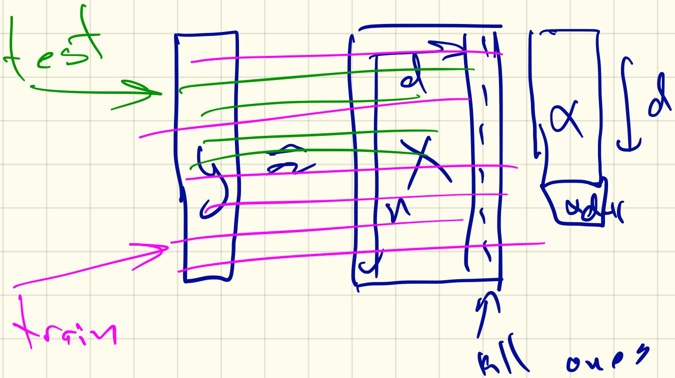
$$\sum_{i=1}^{n} r_i^2$$

in general

$\mathbb{R}^d$

$$M_\alpha(x_i) = \alpha \cdot x_i + b$$

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



$$\alpha^* = \underset{\alpha}{\text{Find}} \quad \text{minimise} \quad \|y - X\alpha\|_2^2$$

$$M_\alpha(x_i) = \langle [x_i, 1], \alpha \rangle$$

$$\alpha = [\alpha_1, \alpha_2 \dots \alpha_{d+1}]$$

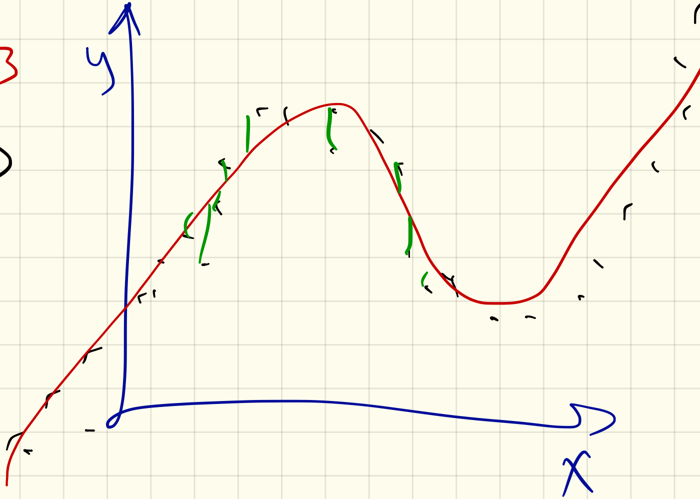
<u>closed form</u>

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

Polynomial Extension

$$M_\alpha(x_i) = \alpha_0 + \alpha_1 x_i + \alpha_2 x_i^2 + \alpha_3 x_i^3$$

$$\langle \alpha , (1, x_i, x_i^2, x_i^3) \rangle$$



$$X_3 = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 \\ 1 & x_2 & x_2^2 & x_2^3 \\ 1 & x_3 & x_3^2 & x_3^3 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 \end{bmatrix}$$

$$\alpha_3^* = (X_3^T X_3)^{-1} X_3^T y$$

Gauss-Markov Theorem

if goal $\propto$ $\boxed{\underset{y,x\in\mathbb{R}^n}{\text{minimize}} \ \|y - x\alpha\|_2^2}$

- expected residual $r_i = y_i - x_i\alpha = 0$

hard to change

- assume r_i uncorrelated w/ r_j $i \neq j$

then $\alpha^* = (X^T X)^{-1} X^T y$ is optimal

variance $\{r_i\}$ as small as possible.

↳ add bias $E[r_i] \neq 0$

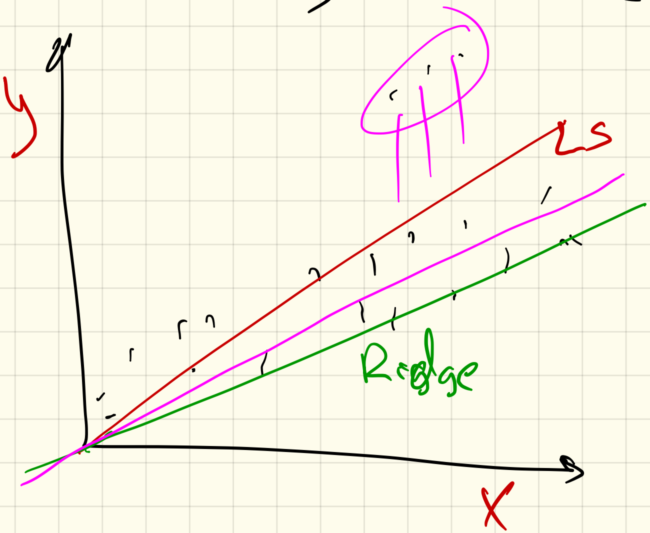
↳ reduce variance $\{r_i\}$:

↳ induce sparsity $\Longrightarrow$ some $\alpha_j = 0$

Tikhonov Regularization (Ridge Regression)

Error

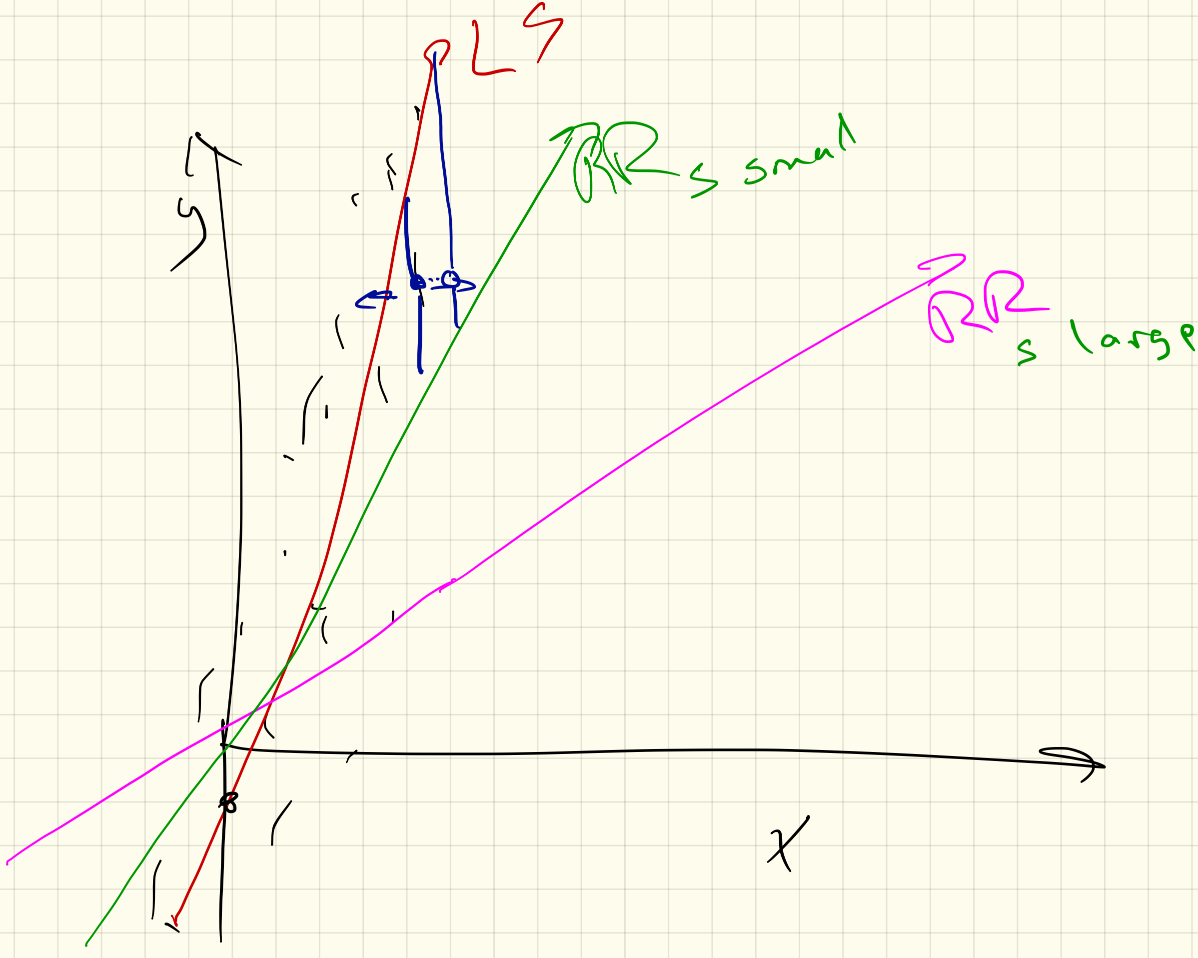
$$L_{z,s}(x,y,\alpha) = \overbrace{\sum_{x_i \in X} \left(y_i - \langle \alpha, x_i \rangle\right)^2}^{\text{least squares}} + \overbrace{s\|\alpha\|_2^2}^{\text{bias}}$$



closed form

$$\alpha_s^* = (X^T X + s^2 I)^{-1} X^T y$$

Choose s by cross-validation

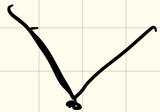


LS
RR s small
RR s large
y
x

Lasso (Basis Pursuit)

↳ induces sparsity

some $\alpha_i = 0$



$$L_{1,s}(X,y,\alpha) = \sum_{x_i \in X} \left(y_i - \langle x_i, \alpha \rangle\right)^2 + s\|\alpha\|_1$$

$$L_1^t(X,y,\alpha) = \sum_{x_i \in X} \left(y_i - \langle x_i, \alpha \rangle\right)^2 \quad \text{s.t. } \|\alpha\|_1 \le t$$

$\forall \alpha_s^* \quad \exists t \quad$ s.t. $\quad \alpha_s^* = \alpha_t^*$

Solve $\alpha_s^* = \arg\min_{\alpha} L_{1,s}(X,y,\alpha)$ set $t = \|\alpha_s^*\|_1$

then $\alpha_s^* = \alpha_t^* = \arg\min_{\alpha} L_{1,t}(X,y,\alpha)$

Ways Around Gauss-Markov

- **Robust Estimators**

Data $(x_i, y_i) \underset{iid}{\sim}$ Model $\begin{cases} P(\text{noise}) = 0.05 : \text{cause weird} \\ p(\text{data}) = 0.95 = y_i = \alpha x + \varepsilon \end{cases}$

$\varepsilon \sim N(0,1)$