Foundations of Machine Learning AI2000 and AI5000

FoML-07
Geometrical Interpretation of Linear Regression (least squares)

Dr. Konda Reddy Mopuri
Department of AI, IIT Hyderabad
July-Nov 2025

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Data-driven Intelligence & Learning Lab

So far in FoML

- What is ML and the learning paradigms
- Probability refresher
- MLE, MAP, and fully Bayesian treatment
- Linear Regression with basis functions

$$\sum_{i=1}^{N} \left(t_i - w^T \phi(x_i)\right)^2$$

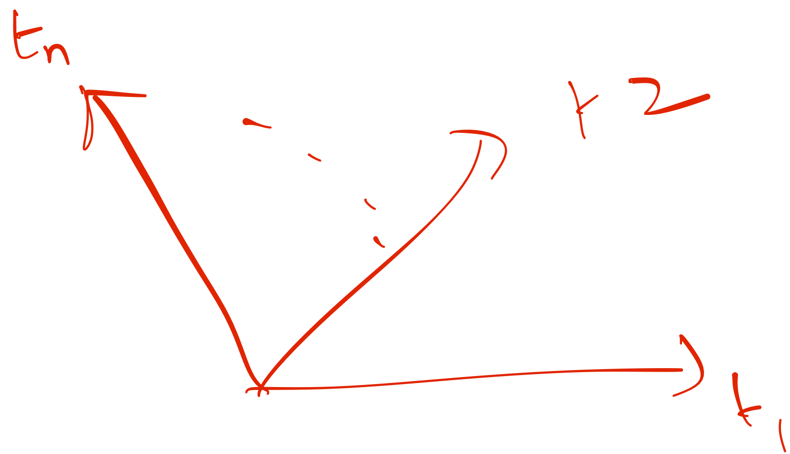
Geometrical Interpretation of Least Squares

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Data-driven Intelligence & Learning Lab

Geometry of Least Squares

- Consider an N-dim space
- Axes are given by t_n (n = 1, 2, ….. N)



Geometry of Least Squares

- Consider an N-dim space
- Axes are given by t_n ($n = 1, 2, \ldots\ldots N$)
- $\underline{\mathbf{t}} = (t_1, t_2, \ldots., t_N)$ becomes a vector in that space

$$\bar{t} = \begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix}_{N \times 1}$$

vector of stacked targets/Labels

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence & Learning Lab

Geometry of Least Squares

- Values of each basis function ϕ_j is a vector
  - Evaluated at all the training data

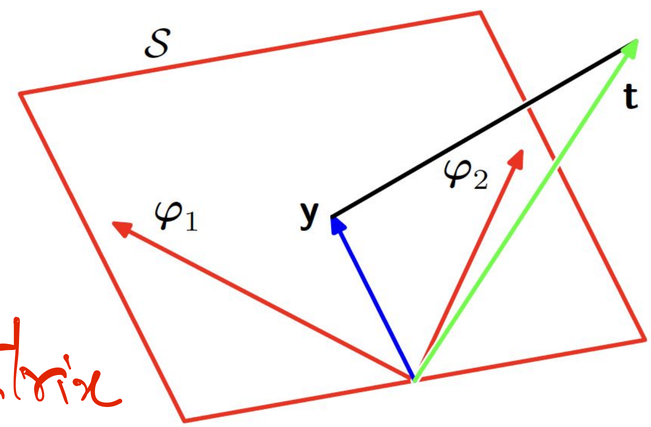
$$\Phi = \begin{bmatrix} | & | & & | \\ \phi_0 & \phi_1 & \cdots & \phi_{M-1} \\ | & | & & | \end{bmatrix}$$

$N \times M$

$$\phi_i = \begin{bmatrix} \phi_i(x_1) \\ \phi_i(x_2) \\ \vdots \\ \phi_i(x_N) \end{bmatrix}_{N \times 1}$$

$i = 0, 1, \ldots M-1$

i th column of Φ the design matrix



N-dim vectors, M in total

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

DiL
Data-driven Intelligence & Learning Lab

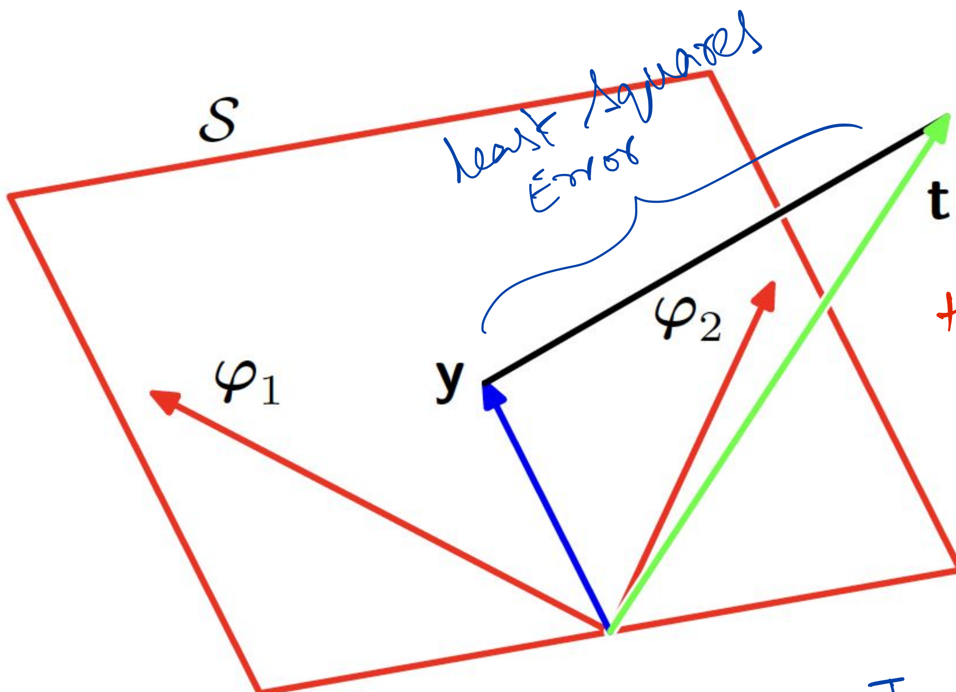
Geometry of Least Squares

ϕ_i vectors span a space 'S'

$\sum_{i=0}^{M-1} \alpha_i \phi_i$

if $M < N$

the $\dim(S) = M$



$$\underline{y} = \begin{bmatrix} w^T \phi(x_1) \\ \underline{w}^T \underline{\phi}(x_2) \\ \vdots \\ w^T \phi(x_N) \end{bmatrix}_{N \times 1}$$

vector of stacked responses
predictions

$\Rightarrow \underline{y} = \Phi \underline{w}$

$\underline{y}$ is linear comb. of $\{\underline{\phi}_i\}$

Geometry of Least Squares

- Solution for w corresponds to the choice of prediction (y) that is the orthogonal projection of **t** (vector of targets) onto the subspace spanned by the basis functions

Nearest vector to $\bar{t}$ in the space spanned by $\{\phi_i\}$

Rough work

భారతీయ సాంకేతిక విజ్ఞాన సంస్థ హైదరాబాద్
भारतीय प्रौद्योगिकी संस्थान हैदराबाद
Indian Institute of Technology Hyderabad

Data-driven Intelligence & Learning Lab