Foundations of Machine Learning
AI2000 and AI5000

FoML-06
Linear Regression

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

DiL
Data-driven Intelligence
& Learning Lab

So far in FoML

- What is ML and the learning paradigms
- Probability refresher
- MLE, MAP, and fully Bayesian treatment

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

DiL
Data-driven Intelligence
& Learning Lab

Linear Regression

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

Data-driven Intelligence
& Learning Lab

Linear Regression

- Dataset D $= \{ (\underline{x}_1, t_1) (\underline{x}_2, t_2) \dots (\underline{x}_N, t_N) \}$ ✓

$\underline{w} \in \mathbb{R}^D$

- Input variable $x_i \in \mathbb{R}^D$
- Output variable $t_i \in \mathbb{R}$
- Simplest linear model

$$\hat{t}_i = y(\underline{x}_i, \underline{w}) = w_0 + w_1 \underline{x_i}_1 + w_2 \underline{x_i}_2 + \dots + w_D \underline{x_i}_D$$

$$= w_0 + \underline{w}^T \cdot x_i$$

Linear Regression

$w_0 = 0$

$x \in \mathbb{R}$

$t \in \mathbb{R}$

$\underline{w} = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}$

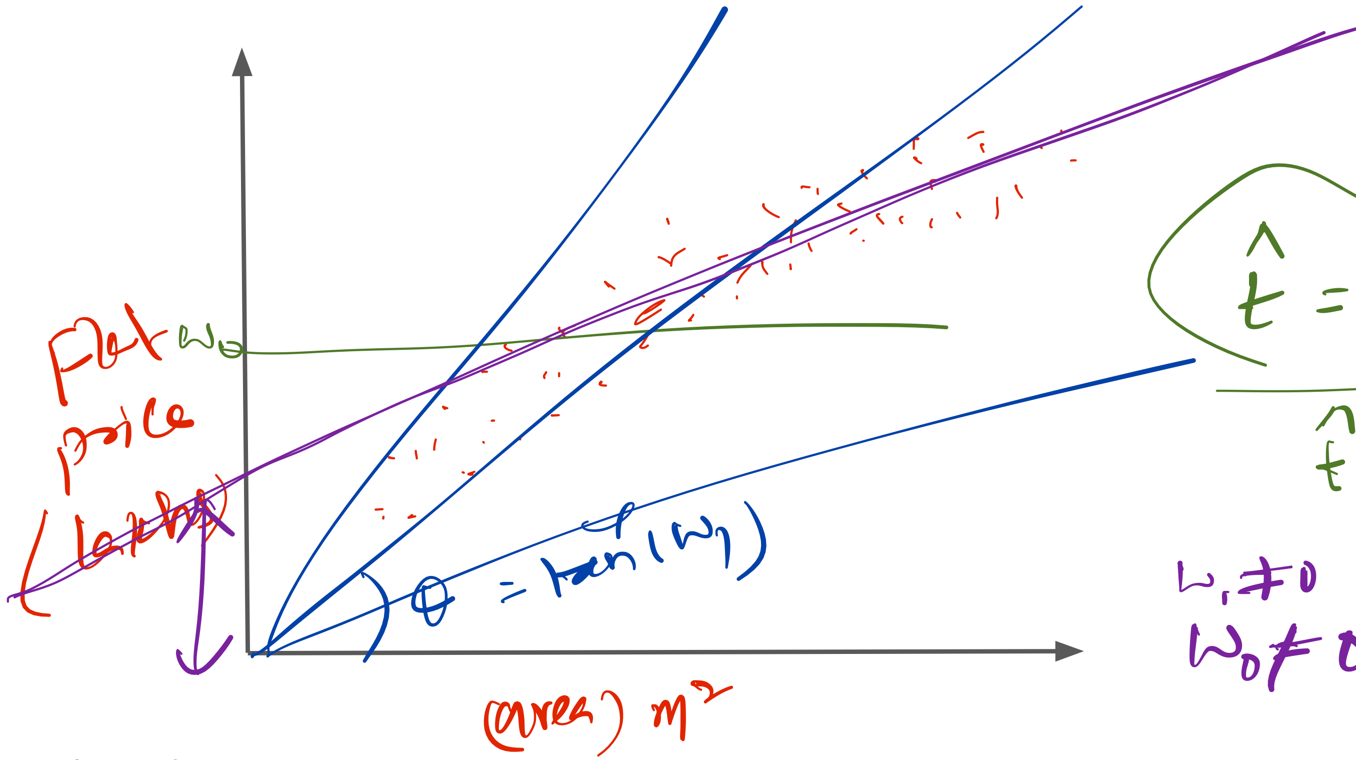
$w_1 = 0$

$\boxed{\hat{t} = w_0 + w_1 x}$

$\hat{t} = w_0$

$w_1 \neq 0$

$w_0 \neq 0$



Linear Basis function Models

- Fix the number of parameters M s.t.
- Choose M-1 basis functions x:
- Mapping/Approximation:

$w_0 \in \mathbb{R}, \ \underline{w} = \begin{bmatrix} w_1 \\ \vdots \\ w_{M-1} \end{bmatrix} \in \mathbb{R}^{M-1}$

$\phi_i(\underline{x}) \in \mathbb{R} \quad i = 1, 2, \ldots M-1$

$\phi_i(\underline{x}) : \mathbb{R}^D \longrightarrow \mathbb{R}$

$$y(\mathbf{x}, \mathbf{w}) = w_0 + w_1 \phi_1(\underline{x}) + w_2 \phi_2(\underline{x}) + \cdots + w_{M-1} \phi_{M-1}(\underline{x})$$

$$y(\underline{x}, \underline{w}) = \hat{t} = w_0 + \underline{\phi}^T \cdot \underline{w} \qquad \underline{\phi} = \begin{bmatrix} \phi_1(\underline{x}) \\ \vdots \\ \phi_{M-1}(\underline{x}) \end{bmatrix}$$

$$= \underline{\tilde{\phi}}^T \cdot \underline{\tilde{w}}$$

$$\underline{\tilde{w}} = \begin{bmatrix} w_0 \\ \underline{w}_1 \\ \vdots \\ \underline{w}_{M-1} \end{bmatrix} \qquad \underline{\tilde{\phi}} = \begin{bmatrix} \phi_0(\underline{x}) = 1 \\ \phi_1(\underline{x}) \\ \vdots \\ \phi_{M-1}(\underline{x}) \end{bmatrix}$$

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

DiL
Data-driven Intelligence & Learning Lab

Example Basis functions

- Components of input

$\underline{x} = (x_1 \; x_2 \cdots x_D)^T$ $\quad M = D$

$\hat{t}(\underline{x}, \underline{w}) = w_0 + \sum_{i=1}^{M-1} w_i \phi_i(\underline{x})$ where $\phi_i(\underline{x}) = \underline{x}_i$

- Powers of input

$\phi_i(x) = x^i$

$x \in \mathbb{R}$

$\underline{w} \in \mathbb{R}^M$

$\hat{t}(\underline{w}, x) = \sum_{i=0}^{M-1} w_i \cdot \phi_i(x)$

$= w_0 + w_1 x + w_2 x^2 + \cdots + w_{M-1} x^{M-1}$

$= \underline{w}^T \cdot \underline{\phi}(x)$

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

DiL
Data-driven Intelligence
& Learning Lab

Example Basis Function

- Gaussian basis functions

$$\phi_i(\underline{x}) = e^{-\frac{1}{2}(\underline{x}-\underline{\mu}_i)^T \Sigma_i^{-1} (\underline{x}-\underline{\mu}_i)}$$

$\underline{x} \in \mathbb{R}^D$

$$\hat{t}(\underline{x}, \underline{w}) = w_0 + \sum_{i=1}^{M-1} w_i \phi_i(\underline{x})$$

Hyper parameters

μ_i, Σ_i, M

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

DiL
Data-driven Intelligence
& Learning Lab

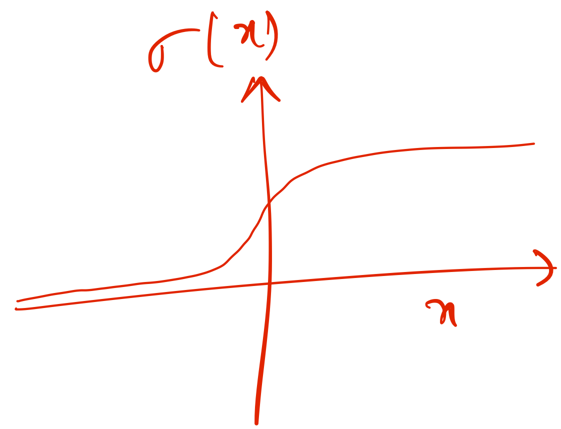
Example Basis Function

- Logistic sigmoid basis functions

$$\hat{t}_i = w_0 + \sum_{i=1}^{m} w_i \cdot \phi_i(x)$$

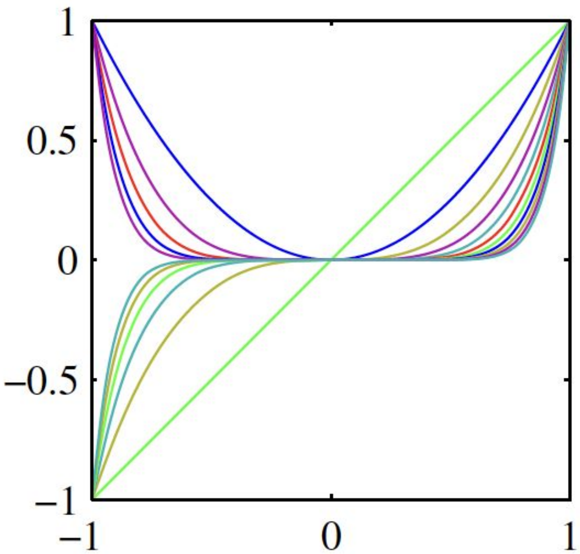
$$\phi_i(x) = \sigma\left(\frac{x - \mu_i}{s_i}\right)$$

$$x, \mu_i, s_i \in R$$

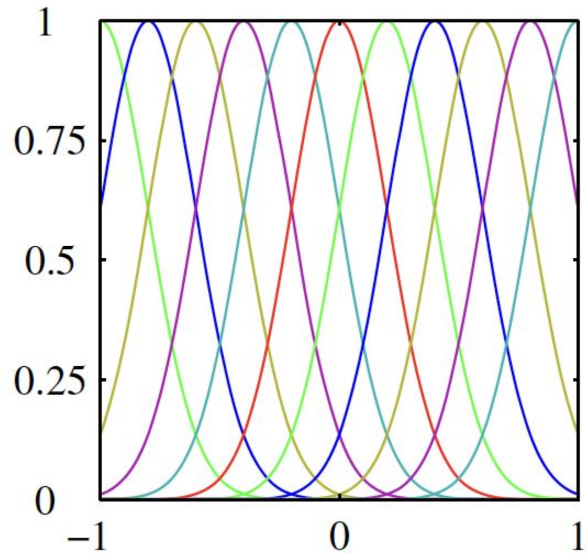


Example Basis Function

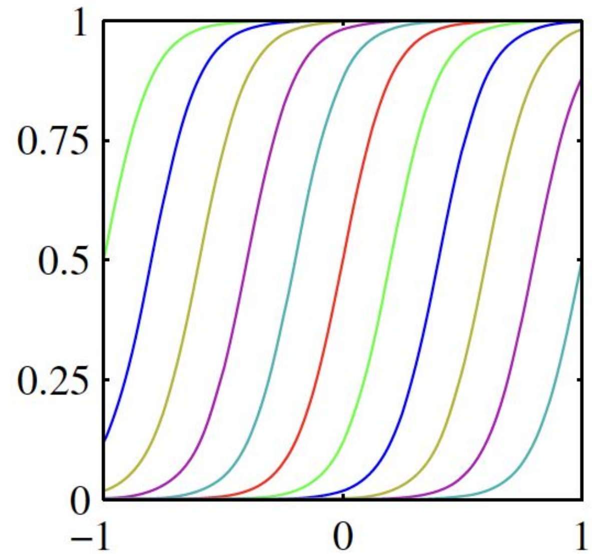
$\phi_i(x) = \sigma\left(\frac{x-\mu_i}{s_i}\right)$



$\phi_i(x) = x^i$



$\phi_i(x) = e^{-\frac{1}{2\sigma_i^2}(x-\mu_i)^2}$



Linear Regression via MLE

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

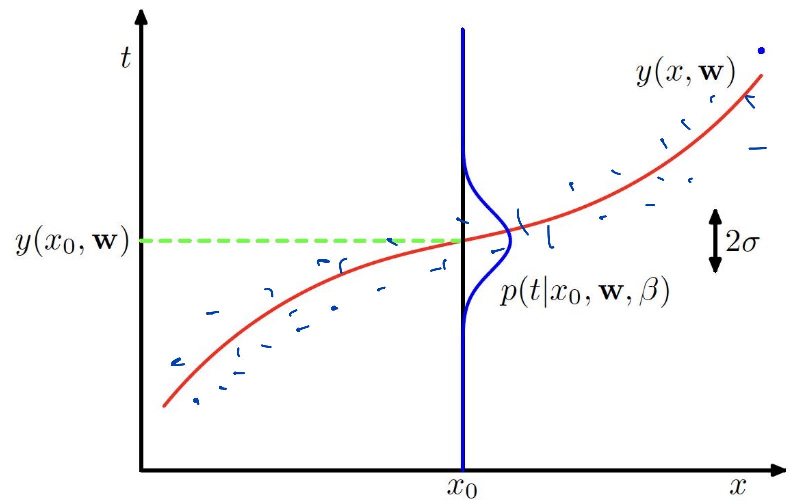
DiL
Data-driven Intelligence
& Learning Lab

Linear Regression

- Given data D $\quad D = \{(x_1, t_1), (x_2, t_2), \ldots (x_N, t_N)\} = \{\mathbf{x}, \mathbf{t}\}$

Input variables $\quad \underline{x}$

Target variables $\quad \underline{t}$



Linear Model with basis functions

$$y(\mathbf{x}, \mathbf{w}) = w_0 + \sum_{j=1}^{M-1} \phi_j(x)\, w_j$$

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

DiL
Data-driven Intelligence & Learning Lab

Maximum Likelihood

- Assume Gaussian noise around the target

$$t = \underline{y(x, \mathbf{w}) + \sigma \cdot \epsilon}, \quad \epsilon \in \mathcal{N}(0, 1)$$

$\beta^{-1} = \sigma^2$

Maximum Likelihood

- Assume Gaussian noise around the target

$$t = y(x, \mathbf{w}) + \sigma \cdot \epsilon, \quad \epsilon \in \mathcal{N}(0, 1)$$

$$p(t|x, \mathbf{w}, \beta) = \mathcal{N}\left(t \,\middle|\, \underline{w}^T \underline{\phi}(\underline{x}), \beta^{-1}\right)$$

Data matrix $\begin{bmatrix} | & | & & | \\ x_1 & x_2 & \cdots & x_N \\ | & | & & | \end{bmatrix}$

Targets vector $\begin{bmatrix} t_1 \\ t_2 \\ \vdots \\ t_N \end{bmatrix}$

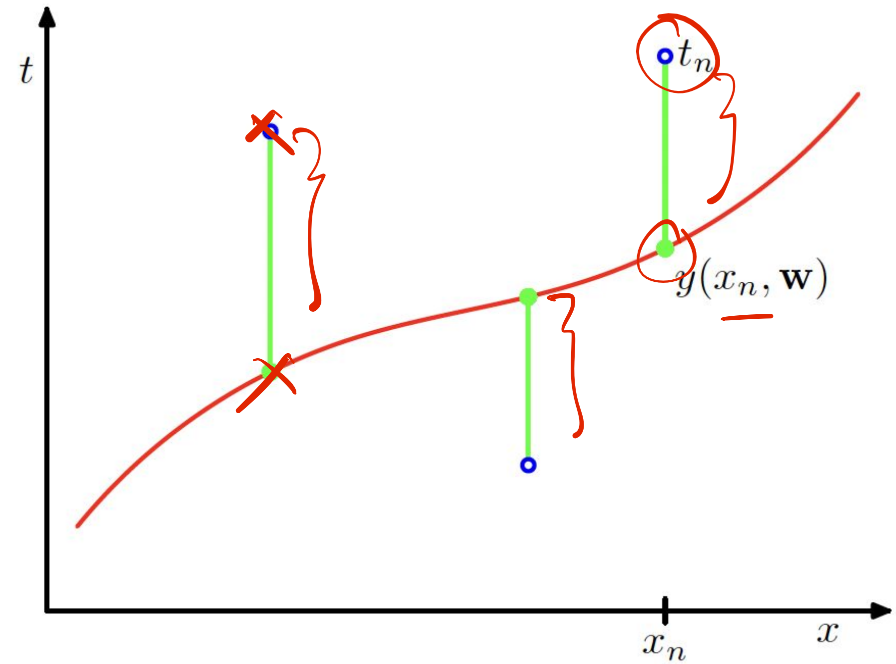
ML: sum of squares error

- Likelihood $$p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) = \prod_{i=1}^{N} \mathcal{N}(t_i|\mathbf{w}^T \boldsymbol{\phi}(\mathbf{x}_i), \beta^{-1})$$

NLL = $-\frac{N}{2}\log\beta + \frac{N}{2}\log 2\pi + \beta \underbrace{\sum_{i=1}^{N}\left[t_i - \underline{w}^T\underline{\phi}(x_i)\right]^2}$

Sum-of-squared error E_D (w) = $\frac{1}{2}\sum_{i=1}^{N}\left[t_i - \underline{w}^T\underline{\phi}(x_i)\right]^2$

ML: sum of squares error



ML Estimates

- Minimize the NLL (or, the sum of squared errors)

$$\frac{\partial}{\partial \underline{w}}\left(\frac{1}{2}\sum_{i=1}^{N}\left[t_i - \underline{w}^T\underline{\phi}(x_i)\right]^2 \right) = 0$$

$$\sum_{i=1}^{N}\left[t_i - \underline{w}^T\underline{\phi}(x_i)\right]\cdot(-1)\,\underline{\phi}(x_i)^T = 0$$

$$\sum_{i=1}^{N} t_i\,\underline{\phi}^T(x_i) = \sum_{i=1}^{N} \underline{w}^T\cdot\underline{\phi}(x_i)\,\phi^T(x_i)$$

$\longrightarrow$ Apply transpose

$$\sum_{i=1}^{N} t_i\cdot\underline{\phi}(x_i) = \sum_{i=1}^{N} \underline{\phi}(x_i)\,\underline{\phi}(x_i)^T\cdot\underline{w}$$

ML Estimates

$$\underline{\phi}_0 \qquad \Phi = \begin{bmatrix} \phi_0(x_1) & \phi_1(x_1) & \dots & \phi_{M-1}(x_1) \\ \phi_0(x_2) & \phi_1(x_2) & \dots & \phi_{M-1}(x_2) \\ \vdots & \vdots & & \\ \phi_0(x_N) & \phi_1(x_N) & \dots & \phi_{M-1}(x_N) \end{bmatrix}$$

design matrix

- Optimal w* satisfies

$$\left(\sum_{i=1}^{N} \underline{\phi}(x_i)\,\underline{\phi}(x_i)^T\right) \cdot \underline{w} = \sum_{i=1}^{N} \underline{\phi}(x_i) \cdot t_i$$

$$\Phi^T \Phi \cdot \underline{w} = \Phi^T \underline{t}$$

$$\underline{w} = [\Phi^T \Phi]^{-1} \Phi^T \underline{t}$$

pseudo inverse

$$\mathbb{E}[t'|\mathbf{x}', \mathbf{w}_{\mathbf{ML}}] = \underline{w}_{ML}^T\, \underline{\phi}(x')$$

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

DiL
Data-driven Intelligence & Learning Lab

Next
SGD

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

Data-driven Intelligence & Learning Lab