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

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

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

Data-driven Intelligence & Learning Lab

Linear Regression

- Dataset D

- Input variable
- Output variable
- Simplest linear model

Linear Regression

Linear Basis function Models

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

$$y(\mathbf{x}, \mathbf{w}) =$$

Example Basis functions

- Components of input

- Powers of input

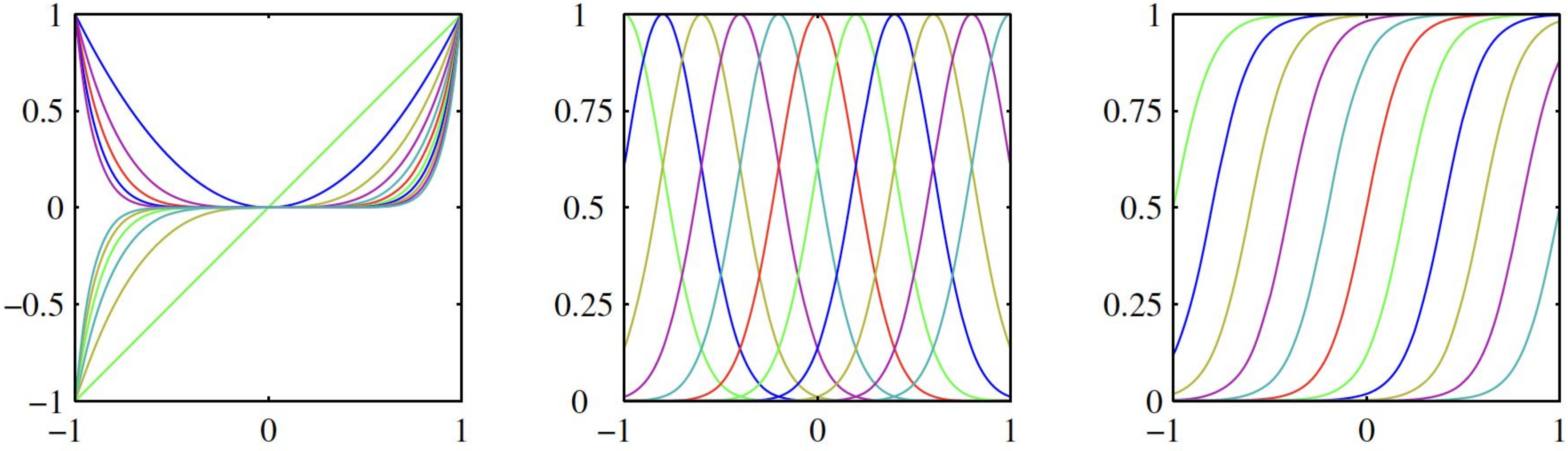
Example Basis Function

- Gaussian basis functions

Example Basis Function

- Logistic sigmoid basis functions

Example Basis Function



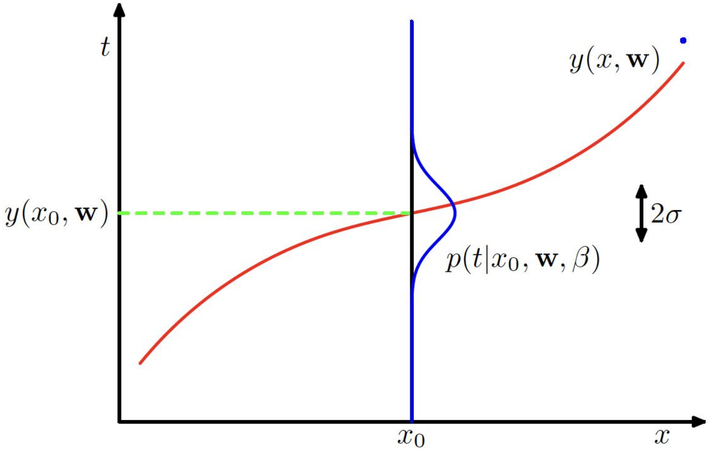
Linear Regression via MLE

Linear Regression

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

Input variables

Target variables



Linear Model with basis functions

$$y(\mathbf{x}, \mathbf{w}) =$$

Maximum Likelihood

- Assume Gaussian noise around the target

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

Maximum Likelihood

- Assume Gaussian noise around the target

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

$$p(t|x, \mathbf{w}, \beta) =$$

Data matrix

Targets vector

ML: sum of squares error

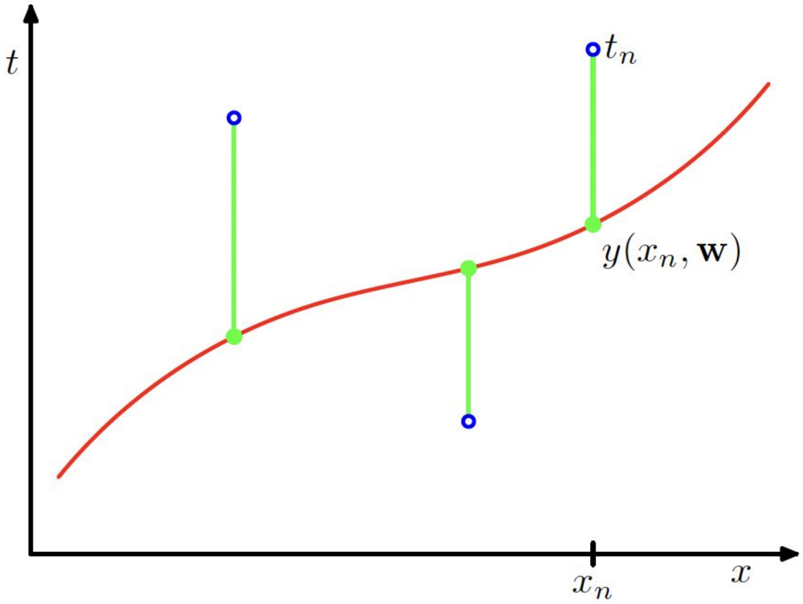
- Likelihood

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

NLL =

Sum-of-squared error E_D (w) =

ML: sum of squares error



ML Estimates

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

ML Estimates

- Optimal w* satisfies

$$\mathbb{E}[t'|\mathbf{x}', \mathbf{w_{ML}}] =$$

Next
SGD