Foundations of Machine Learning
AI2000 and AI5000

FoML-11
Bayesian 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

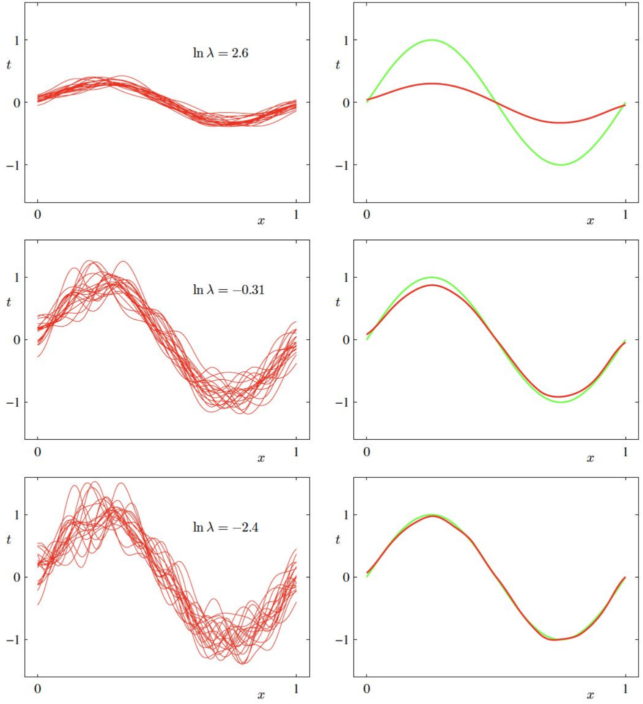
So far in FoML

- What is ML and the learning paradigms
- Probability refresher
- MLE, MAP, and fully Bayesian treatment
- Linear Regression with basis functions - and regularization
- Model selection
- Bias-Variance Decomposition/Trade-off

Bayesian Regression

We have seen that

- Model averaging may be a good thing to do
  - Across different datasets

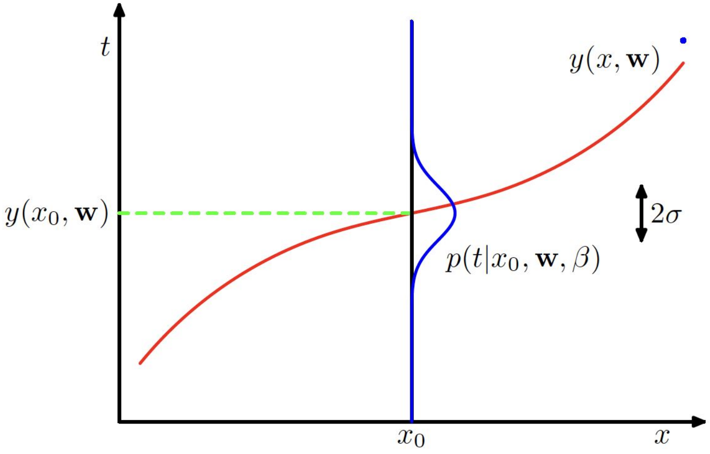


Bayesian Regression

- Instead of averaging over different datasets, we do it over different parameter sets

Bayesian Linear Regression

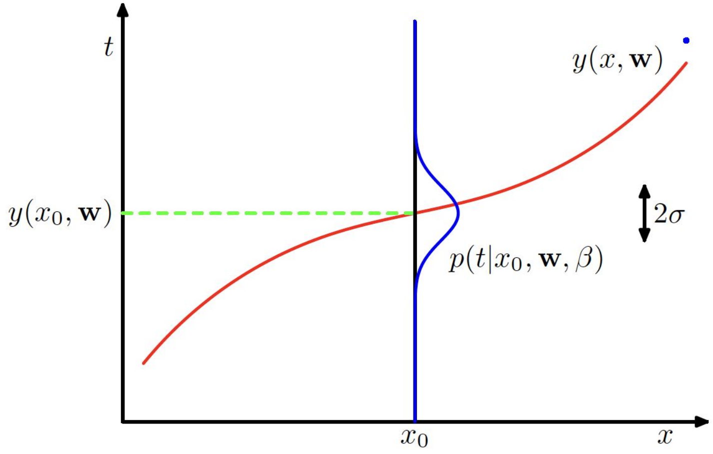
$$\text{Data} \quad \mathbf{t} = (t_1, \ldots t_N)^T \quad \mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^T$$



Bayesian Linear Regression

Data $\mathbf{t} = (t_1, \ldots t_N)^T \quad \mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_N)^T$

Likelihood: $p(t' | \mathbf{x}', \mathbf{w}, \beta) =$



$$p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) = \prod_{i=1}^{N} \mathcal{N}(t_i | y(x_i, \mathbf{w}), \beta^{-1}) =$$

Posterior distribution: $p(\mathbf{w}|\mathbf{t}, \mathbf{X}) = \frac{p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) p(\mathbf{w})}{p(\mathbf{t}|\mathbf{X}, \beta)}$

Bayesian Linear Regression

Conjugate Prior: $p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|m_0, \mathbf{S}_0)$

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

Bayesian Linear Regression

Conjugate Prior: $p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|m_0, \mathbf{S}_0)$

$$p(\mathbf{w}|\mathbf{t}, \mathbf{X}) = \frac{p(\mathbf{t}|\mathbf{X},\mathbf{w},\beta)p(\mathbf{w})}{p(\mathbf{t}|\mathbf{X},\beta)} = \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N)$$

$$\mathbf{S}_N^{-1} = \mathbf{S}_0^{-1} + \beta\mathbf{\Phi}^T\mathbf{\Phi}$$
$$\mathbf{m}_N = \mathbf{S}_N(\mathbf{S}_0^{-1}\mathbf{m}_0 + \beta\mathbf{\Phi}^T\mathbf{t})$$

Bayesian Linear Regression

- Simple prior: $p(\mathbf{w}|\alpha) =$

$$\mathbf{m}_0 = \mathbf{0}$$
$$\mathbf{S}_0 = \alpha^{-1}\mathbf{I}$$

Bayesian Linear Regression

- Simple prior: $p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I})$

$$\mathbf{m}_0 = \mathbf{0}$$
$$\mathbf{S}_0 = \alpha^{-1}\mathbf{I}$$

- Posterior

$$p(\mathbf{w}|\mathbf{t}, \mathbf{X}, \alpha, \beta) = \mathcal{N}(\mathbf{w}|\mathbf{m}_N, \mathbf{S}_N)$$
$$\mathbf{S}_N^{-1} = \mathbf{S}_0^{-1} + \beta\boldsymbol{\Phi}^T\boldsymbol{\Phi} =$$
$$\mathbf{m}_N = \mathbf{S}_N(\mathbf{S}_0^{-1}\mathbf{m}_0 + \beta\boldsymbol{\Phi}^T\mathbf{t}) =$$

Bayesian Linear Regression

- Special prior: Infinitely broad prior (no restriction) on w

$$p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I}) \quad \alpha \to 0$$

- Posterior

$$\mathbf{S}_N^{-1} = \mathbf{S}_0^{-1} + \beta\mathbf{\Phi}^T\mathbf{\Phi} =$$

$$\mathbf{m}_N = \mathbf{S}_N(\mathbf{S}_0^{-1}\mathbf{m}_0 + \beta\mathbf{\Phi}^T\mathbf{t}) =$$

Bayesian Linear Regression

- Special prior: Infinitely narrow prior on w

$$p(\mathbf{w}|\alpha) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I}) \quad \alpha \rightarrow \inf$$

- Posterior

$$\mathbf{S}_N^{-1} = \mathbf{S}_0^{-1} + \beta\mathbf{\Phi}^T\mathbf{\Phi} =$$

$$\mathbf{m}_N = \mathbf{S}_N(\mathbf{S}_0^{-1}\mathbf{m}_0 + \beta\mathbf{\Phi}^T\mathbf{t}) =$$

Next
Decision Theory