University of Amsterdam
Informatics Institute

AMLAB
Amsterdam Machine Learning Lab

Machine Learning 1

Lecture 4.3 - Supervised Learning
Bayesian Linear Regression - Gaussian Posteriors

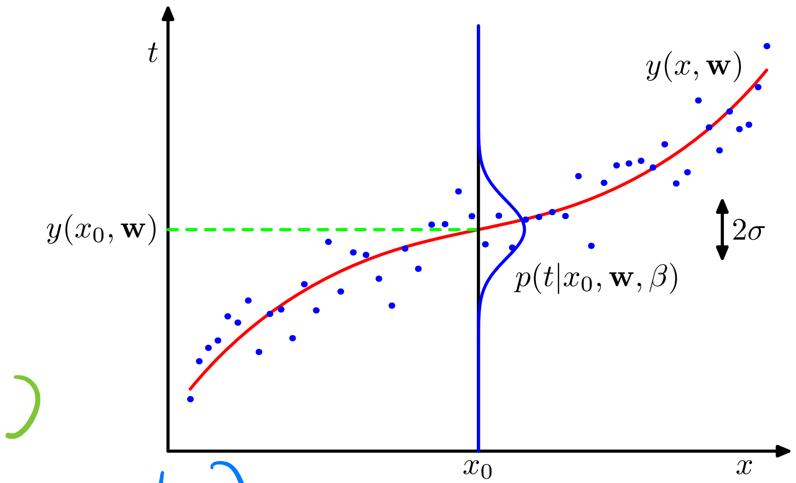
Erik Bekkers

(Bishop 3.3.1 (and 2.3.3)

Image credit: Kirillm | Getty Images

Slide credits: Patrick Forré and Rianne van den Berg

Bayesian Linear Regression



- Data: $\mathbf{t} = (t_1, ..., t_N)^T \quad \mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N)^T$

- Likelihood: $p(t' \,|\, \mathbf{x}', \mathbf{w}, \beta) = \mathcal{N}(t' \,|\, \underbrace{\underline{w}^T \underline{\phi}(x')}_{y(x', \underline{w})}, \beta^{-1})$

verify

$$p(\mathbf{t}|\mathbf{X}, \mathbf{w}, \beta) = \prod_{n=1}^{N} \mathcal{N}(t_n | y(\mathbf{x}_n, \mathbf{w}), \beta^{-1}) = \mathcal{N}(\underline{t} \,|\, \underline{\Phi}\, \underline{w}, \beta^{-1} I)$$

- Conjugate prior: $p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_0, \mathbf{S}_0)$

$$\mathbf{\Phi} = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \dots & \phi_{M-1}(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_N) & \dots & \phi_{M-1}(\mathbf{x}_N) \end{pmatrix}$$

- Posterior distribution:

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

Ch 2.3

Bishop Eq. 2.116

- Maximum A Posteriori estimate:

$$\mathbf{w}_{\text{MAP}} = \underline{m}_N$$

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

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

Bayesian Linear Regression

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

$$\mathbf{m}_0 = \mathbf{0}$$

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

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

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

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

$$p(\mathbf{w}|\mathbf{X}, \mathbf{t}, \alpha, \beta) = \frac{1}{\sqrt{(2\pi)^M \cdot |S_N|}} e^{-\frac{1}{2}(\underline{w} - \underline{m}_N) S_N^{-1} (\underline{w} - \underline{m}_N)}$$

Bayesian Linear Regression

$$\mathbf{m}_N = \beta \mathbf{S}_N \mathbf{\Phi}^T \mathbf{t}$$
$$\mathbf{S}_N^{-1} = \alpha \mathbb{1} + \beta \mathbf{\Phi}^T \mathbf{\Phi}$$

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

- Infinitely broad prior: no restriction on **w**!

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

$$\lim_{\alpha \to 0} \mathbf{m}_N = \lim_{\alpha \to 0} \beta \left(\alpha \mathbb{1} + \beta \mathbf{\Phi}^T \mathbf{\Phi}\right)^{-1} \mathbf{\Phi}^T \mathbf{t} = \cancel{\frac{\beta}{\beta}} (\underline{\Phi}^T \underline{\Phi})^{-1} \underline{\Phi}^T \underline{t}$$

$$= \underline{w}_{ML}$$

- Infinitely narrow prior:

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

$$\lim_{\alpha \to \infty} \mathbf{m}_N = \lim_{\alpha \to \infty} \beta \left(\alpha \mathbb{1} + \beta \mathbf{\Phi}^T \mathbf{\Phi}\right)^{-1} \mathbf{\Phi}^T \mathbf{t} = \lim_{\alpha \to \infty} \frac{\beta}{\alpha} \underline{\Phi}^T \underline{t} = \underline{0} = \underline{m}_N$$

$$\lim_{\alpha \to \infty} \mathbf{S}_N = \lim_{\alpha \to \infty} \left(\alpha \mathbb{1} + \beta \mathbf{\Phi}^T \mathbf{\Phi}\right)^{-1} = \lim_{\alpha \to \infty} \frac{1}{\alpha} I = 0$$

zero matrix