University of Amsterdam
Informatics Institute

AMLAB Amsterdam Machine Learning Lab

Machine Learning 1

Lecture 2.5 - Maximum A Posteriori

Erik Bekkers

(Bishop 1.2.5 - 1.2.6)

Image credit: Kirillm | Getty Images

Slide credits: Patrick Forré and Rianne van den Berg

Maximum A Posteriori Estimates

- Dataset $D = (x_1, x_2, \ldots, x_N)$ of N independent observations.

- ML estimate: choose ***w*** such that data likelihood is maximized:

$$\mathbf{w}_{ML} = \arg\max_{w} \; p(D \mid w)$$

- MAP estimate: choose most probable ***w*** given the data.

$$\mathbf{w}_{MAP} = \arg\max_{w} \; p(w \mid D)$$

posterior distribution

Curve Fitting: Maximum A Posteriori Estimates

- Dataset $D = \{(x_1, t_1), ..., (x_N, t_N)\} = \{\mathbf{x}, \mathbf{t}\}$
- Model: $p(t|x, \mathbf{w}, \beta) = \mathcal{N}(t|y(x, \mathbf{w}), \beta^{-1}) = \sqrt{\frac{\beta}{2\pi}} \exp\left[-\frac{\beta}{2}(t - y(x, \mathbf{w}))^2\right]$
- ML estimate: choose ***w*** such that data likelihood is maximized:

$$\mathbf{w}_{ML} = \underset{\mathbf{w}}{\operatorname{argmax}}\, p(\mathbf{t} \,|\, \mathbf{x}, \mathbf{w}, \beta) = \underset{\mathbf{w}}{\operatorname{argmax}} \log p(\mathbf{t} \,|\, \mathbf{x}, \mathbf{w}, \beta)$$

- MAP estimate: choose most probable ***w*** given the data.

$$\mathbf{w}_{MAP} = \underset{\mathbf{w}}{\operatorname{argmax}}\, p(\mathbf{w} \,|\, \mathbf{x}, \mathbf{t}, \beta)$$

↳ posterior

Curve Fitting: Maximum A Posteriori Estimates

- Dataset $D = \{(x_1, t_1), ..., (x_N, t_N)\} = \{\mathbf{x}, \mathbf{t}\}$

- Model: $p(t|x, \mathbf{w}, \beta) = \mathcal{N}(t|y(x, \mathbf{w}), \beta^{-1}) = \sqrt{\frac{\beta}{2\pi}} \exp\left[-\frac{\beta}{2}(t - y(x, \mathbf{w}))^2\right]$

- Given a prior $p(\mathbf{w}|\alpha)$ the posterior distribution is

$$p(\mathbf{w}\,|\,\mathbf{x}, \mathbf{t}, \beta, \alpha) = \sim \frac{p(t|x, \underline{w}, \beta) \cdot p(\underline{w}|\alpha)}{p(t|x, \beta, \alpha)}$$

does not depend on w

- Maximum A Posteriori Estimate:

$$\mathbf{w}_{MAP} = \underset{\mathbf{w}}{\operatorname{argmax}}\, p(\mathbf{w}\,|\,\mathbf{x}, \mathbf{t}, \beta, \alpha) = \underset{w}{\arg\max} \log p(w|x, t, \beta, \alpha)$$

$$= \underset{w}{\arg\max} \log p(t|x, w, \beta) + \log p(w|\alpha) - \log p(t|x, \beta, \alpha)$$

Curve Fitting: Maximum A Posteriori Estimates

- Gaussian prior: $\mathbf{w} \in \mathbb{R}^M$

$$p(\mathbf{w}|\alpha) = \prod_{i=1}^{M} \mathcal{N}(w_i|0, \alpha^{-1}) = \left(\frac{\alpha}{2\pi}\right)^{M/2} \prod_{i=1}^{M} e^{-\frac{\alpha}{2} w_i w_i} = \left(\frac{\alpha}{2\pi}\right)^{\frac{M}{2}} e^{-\frac{\alpha}{2}\mathbf{w}^T\mathbf{w}}$$

$$\mathbf{w}_{MAP} = \underset{\mathbf{w}}{\operatorname{argmin}} - \log p(\mathbf{w} \,|\, \mathbf{x}, \mathbf{t}, \beta, \alpha) = \underset{\mathbf{w}}{\operatorname{argmin}} - \log p(\mathbf{t} \,|\, \mathbf{x}, \mathbf{w}, \beta) - \log p(\mathbf{w} \,|\, \alpha)$$

$$= \underset{w}{\operatorname{argmin}} -\log p(t|x, w, \beta) + \frac{\alpha}{2} w^T w$$

- Curve fitting a function with Gaussian noise and Gaussian prior:

$$p(t|x, \mathbf{w}, \beta) = \sqrt{\frac{\beta}{2\pi}} \exp\left[-\frac{\beta}{2}\left(t - y(x, \mathbf{w})\right)^2\right]$$

$$\mathbf{w}_{\text{MAP}} = \arg\min_{\mathbf{w}} \frac{\beta}{2} \sum_{i=1}^{N} \left(t - y(x_i, w)\right)^2 + \frac{\alpha}{2}\mathbf{w}^T\mathbf{w}$$

- Predictive distribution:

$$p(t' \,|\, x', \mathbf{w}_{\text{MAP}}, \beta) = \mathcal{N}(t' \,|\, y(x', w_{MAP}), \beta^{-1})$$

$$\mathbb{E}[t' \,|\, x', \mathbf{w}_{\text{MAP}}, \beta] = y(x', w_{MAP})$$