Bayesian Interpretation

Fallstudien der mathematischen Modellbildung, Teil 2

20.10.2023 - 21.11.2023,

paul.catala@tum.de

Maximum Likelihood

- Stochastic forward model:

$$y = Ax + w$$

where $w \sim_{i.i.d.} \mathcal{N}(0, \sigma^2 \mathrm{Id})$, and hence $y \sim \mathcal{N}(Ax, \sigma^2 \mathrm{Id})$.

Maximum Likelihood

- Stochastic forward model:
$$y = Ax + w$$
where $w \sim_{i.i.d.} \mathcal{N}(0, \sigma^2 \mathrm{Id})$, and hence $y \sim \mathcal{N}(Ax, \sigma^2 \mathrm{Id})$.
- The inverse problem is a Maximum Likelihood Estimation (MLE) problem:
$$\hat{x} = \mathrm{argmax}_x \, p(y|x)$$
where $p(y|x)$ is the likelihood (of y knowing x). In our example
$$p(y|x) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \prod_{i=1}^{n} \exp\left(-\frac{(y_i - a_i^\top x)^2}{2\sigma^2}\right)$$
(where we used the i.i.d. hypothesis).

Maximum Likelihood

- Stochastic forward model:

$$y = Ax + w$$

where $w \sim_{i.i.d.} \mathcal{N}(0, \sigma^2 \mathrm{Id})$, and hence $y \sim \mathcal{N}(Ax, \sigma^2 \mathrm{Id})$.

- The inverse problem is a Maximum Likelihood Estimation (MLE) problem:

$$\hat{x} = \mathbf{argmax}_x \; p(y|x)$$

where $p(y|x)$ is the likelihood (of y knowing x). In our example

$$p(y|x) = \frac{1}{(\sqrt{2\pi\sigma^2})^n} \prod_{i=1}^{n} \mathbf{exp}\left(-\frac{(y_i - a_i^\top x)^2}{2\sigma^2}\right)$$

(where we used the i.i.d. hypothesis).

- It is convenient to work with the log-likelihood

$$\mathbf{log}\, p(y|x) = c_0 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - a_i^\top x)^2 = c_0 - \frac{1}{2\sigma^2} \|y - Ax\|^2$$

so $\hat{x}$ is the solution of a least-squares regression.
$\rightarrow$ A least-square fidelity is adapted for Gaussian i.i.d. noise

Bayesian Interpretation of Regularization

- Variational formulations fidelity + regularity can be interpreted in the Bayesian framework.

Bayesian Interpretation of Regularization

- Variational formulations fidelity + regularity can be interpreted in the Bayesian framework.
- Bayesian treat x as a random variable as well, and put a prior distribution $p(x)$ on it. The Maximum A Posteriori (MAP) seeks the best estimate knowing the observed data, *i.e.*

$$\tilde{x} = \text{argmax}_x \, p(x|y)$$

- Variational formulations fidelity + regularity can be interpreted in the Bayesian framework.
- Bayesian treat x as a random variable as well, and put a prior distribution $p(x)$ on it. The Maximum A Posteriori (MAP) seeks the best estimate knowing the observed data, *i.e.*

$$\tilde{x} = \mathrm{argmax}_x \, p(x|y)$$

- By Bayes rule,

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

$p(y\|x)$	:	likelihood (= model on the noise)
$p(x)$	:	prior distribution of the model
$p(y)$	:	marginal distribution of the data (normalizing constant)

- Variational formulations fidelity + regularity can be interpreted in the Bayesian framework.
- Bayesian treat x as a random variable as well, and put a prior distribution $p(x)$ on it. The Maximum A Posteriori (MAP) seeks the best estimate knowing the observed data, *i.e.*

$$\tilde{x} = \mathrm{argmax}_x\, p(x|y)$$

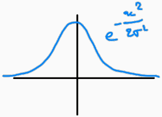
- By Bayes rule,

$$p(x|y) = \frac{p(y|x)p(x)}{p(y)}$$

$p(y\|x)$	:	likelihood (= model on the noise)
$p(x)$	:	prior distribution of the model
$p(y)$	:	marginal distribution of the data (normalizing constant)

- The MAP estimation becomes

$$\mathrm{argmax}\, p(y|x)p(x) \iff \mathrm{argmin}\, E(x) := -\log p(y|x) - \log p(x)$$

- 

Assume $x \sim_{i.i.d.} \mathcal{N}(0, \tau^2)$. Then

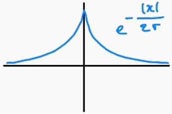
$$\log(p(x)) \propto -\frac{1}{2\tau^2} \sum_{i=1}^{n} x_i^2 = -\frac{1}{2\tau^2} \|x\|^2$$

and the MAP extimation reads gives

$$\operatorname{argmin} \frac{1}{2\sigma^2} \|y - Ax\|^2 + \frac{1}{2\tau^2} \|x\|_2^2$$

- Conclusion: Gaussian prior $\iff$ ℓ^2-regularization, with parameter $\lambda = \sigma^2 / \tau^2$

Laplacian Prior

- 

Laplace distribution has density

$$p(x_i) = \frac{1}{C} \exp\left(\frac{-|x_i - a|}{2\rho^2} \right)$$

where C is a normalizing constant.

- Assume $x \sim_{i.i.d.} \mathrm{Laplace}(0, \rho^2)$. Then

$$\log p(x) \propto -\frac{1}{2\rho^2} \sum_{i=1}^{n} |x_i| = -\frac{1}{2\rho^2} \|x\|_1$$

and the MAP estimation gives

$$\operatorname{argmin} \frac{1}{2\sigma^2} \|y - Ax\|^2 + \frac{1}{2\rho^2} \|x\|_1$$

- Conclusion: Laplace prior $\iff$ ℓ^1-regularization, with parameter $\lambda = \sigma^2/\rho^2$