Learning theory for matrix inverse problems with gaussian priors

Jonas Adler — KTH, Elekta

Olivier Verdier — KTH

Suppose $A : \mathbb{R}^n \to \mathbb{R}^m$ is a forward operator and we have data of the form

$$x \in \mathcal{N}(0, \Sigma)$$
$$y \in \mathcal{N}(Ax, \Gamma)$$

Theorem 1: The maximum a posteriori (MAP) estimate of x given y

$$A^{-1}_{MAP}(y) = \arg\max_{x'} P(x'|y)$$

is

$$A^{-1}_{MAP}(y) = (A^T \Gamma^{-1} A + \Sigma^{-1})^{-1} A^T \Gamma^{-1} y$$

Proof: We have

$$P(x'|y) \propto \underbrace{P(y|x')}_{\mathcal{N}(Ax', I)} \; \underbrace{P(x')}_{\mathcal{N}(0, I)}$$

Instead of maximizing the posterior, we minimize the log of the posterior

$$\arg\max_{x'} P(x'|y) = \arg\min \log P(x'|y) = \arg\min \left(\log P(y|x') + \log P(x')\right)$$

where we let

$$\begin{aligned} f(x') = &\ \log P(y|x') + \log P(x') \\ = &\ \frac{1}{2}\|Ax - y\|^2_{\Gamma^{-1}} + \frac{1}{2}\|x\|^2_{\Sigma^{-1}} \end{aligned}$$

we compute

$$\nabla f = A^T \Gamma^{-1}(Ax - y) + \Sigma^{-1} x$$

Setting this to zero gives the conditions for a maximum

$$(A^T \Gamma^{-1} A + \Sigma^{-1}) x = A^T \Gamma^{-1} y$$

which gives a maximum likelihood solution as

$$x = (A^T \Gamma^{-1} A + \Sigma^{-1})^{-1} A^T \Gamma^{-1} y$$

which is the Tichonov regularized solution under Gaussian priors.

Theorem 2: We have

$$\arg\min_B \mathbb{E}_{(x,y)} \|B(y) - x\|_2^2 = \mathbb{E}_x\big(x \mid y\big)$$

where the minimization is taken over all operators $\mathbb{R}^m \to \mathbb{R}^n$.[1]

[1] This is also given in Bishop, 1.5.5

Proof: By the law of total probability

$$\mathbb{E}_{(x,y)}\|B(y)-x\|_2^2 = \mathbb{E}_y\big(\mathbb{E}_x\big(\|B(y)-x\|_2^2 \mid y\big)\big)$$

By the monotonicity of the expectation we have

$$B(y) = \arg\min_z \mathbb{E}_x\big(\|z-x\|_2^2 \mid y\big)$$

we find this by differentiating and setting to zero

$$0 = \mathbb{E}_x\big(2(z-x) \mid y\big) = 2\mathbb{E}_x\big(z \mid y\big) - 2\mathbb{E}_x\big(x \mid y\big)$$

which gives

$$B(y) = \mathbb{E}_x\big(x \mid y\big)$$