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

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

is

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

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

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

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

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

Theorem 2: We have

$$\arg\min_B \mathbb{E}_{(x,y)}\|B(y) - x\|_2^2 = \left\{y \mapsto \int xP(x|y)dx\right\}$$

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

Proof: By the definition of the expectation and the chain rule for probabilities

$$
\begin{aligned}
\mathbb{E}_{(x,y)}\|B(y)-x\|_2^2 &= \int\int \|B(y)-x\|_2^2 P(x,y)dxdy \\
&= \int \left(\int \|B(y)-x\|_2^2 P(x|y)dx \right) P(y)dy
\end{aligned}
$$

Since $P(y) \geq 0$ by the monotonicity of the integral we have

$$
B(y) = \arg\min_z \int \|z-x\|_2^2 P(x|y)dx
$$

we find this by differentiating and setting to zero

$$
0 = \int 2(z-x)P(x|y)dx = 2z \underbrace{\int P(x|y)dx}_{1} - 2\int xP(x|y)dx
$$

which gives

$$
B(y) = \int xP(x|y)dx
$$