Martin Burger, Janic Föcke, Daniel Tenbrinck, Philipp Wacker

Inverse Problems - Worksheet 10 Winter term 2019/2020

1 Consolidation assignment

Consider the following denoising problem:

$$f = u + \varepsilon$$

with $\varepsilon \sim N(0, \sigma^2)$ independent Gaussian noise on each pixel. We assume the image to be 1-d and discrete to make computations easier.

Find a way to interpret this in a Bayesian way, i.e.

a) What is the maximum likelihood estimator?

b) Consider the classical solution with a generalized Tikhonov regularization:

$$u^\star = \frac{1}{2}\|u - f\|^2 + \frac{\alpha}{2}\|\nabla u\|^2.$$

Show that $u^\star$ is equal to the maximum-a-posteriori estimator for some prior. What is that prior?

2 Preparation assignment

We continue with the setting of linear regression with a polynomial regression function: Given data $z \in \mathbb{R}^{2\times N}$ (i.e. each of the N data points z_i has two coordinates which we abbreviate by $z_i = (x_i, y_i)$).

We assume that the data was generated by noisy evaluation of a polynomial with n unknown coefficients, i.e.

$$y_i = a_0 + a_1 \cdot x_i + a_2 \cdot x_i^2 + \cdots + a_{n-1} \cdot x_i^{n-1} + \varepsilon_i$$

where $\varepsilon_i \sim N(0, \sigma^2)$ are independent noise terms (with known variance σ^2.

For reference we state the following facts: We consider multivariate $X \sim N(m, \Sigma)$ for some $m \in \mathbb{R}^n, \Sigma \in \mathbb{R}^{n\times n}$ and $E \sim N(0, \Gamma)$ for some $\Gamma \in \mathbb{R}^{m\times m}$. With a matrix $A \in \mathbb{R}^{m\times n}$ we set

$$Y = AX + E.$$

Then

$$\begin{aligned}
Y|X = x &\sim N(A \cdot x, \Gamma) \\
Y &\sim N(A \cdot m, \Gamma + A\Sigma A^T) \\
(X, Y) &\sim N\left(\begin{pmatrix} m \\ Am \end{pmatrix}, \begin{bmatrix} \Sigma & \Sigma A^T \\ A\Sigma & \Gamma + A\Sigma A^T \end{bmatrix}\right) \\
X|Y = y &\sim N(m + K(y - Am), \Sigma - KA\Sigma), \text{ with } K := \Sigma A^T(\Gamma + A\Sigma A^T)^{-1}
\end{aligned}$$

a) We set a prior on the parameters $\vec{a}$ by $N(0, \text{diag}(w_k^2)_k)$ with $w_k = 1/k$. This means that higher polynomials are "penalized" with a lower prior standard deviation. Implement this in Matlab and plot a set of 50 polynomials with coefficents sampled from this distribution.

b) Choose a set of measurement locations x_i (it is interesting to cluster them a bit and leave some parts unobserved, so take e.g. `x = [0.01, 0.1, 0.6, 0.65, 0.7, 0.72, 0.8, 0.9]`).

c) Choose a vector $\vec{a}$ (i.e., a polynomial).

d) Generate data y with a given measurement noise standard deviation (start with $\sigma = 0.05$).

e) Use the formulae above to obtain the posterior $\vec{a}|y$. Implement this in Matlab and plot a set of 50 (or more) polynomials with coefficients sampled from this posterior distribution. Plot this together with the posterior mean and the original function that was used to generate the data.

f) Discuss the following points:

- How well are the data points fitted?
- Are there areas of higher/lower uncertainty? Why?
- Using the samples: Could you make an educated guess regarding interpolation, i.e. What is the value of the polynomial at an unmeasured position (and can you give a measure of uncertainty?)
- What happens if you increase/decrease the measurement noise? What happens if you choose more/less data points x_i? What happens if you set the prior as $N(0, \text{diag}(1, 1, \ldots, 1))$? What happens if you choose a "ground truth" polynomial function which is considered unlikely by the prior (e.g. $100 \cdot x^2$)?

3 Notes / Insights from class