Viterbo Giuseppe Homework 3

Matricola: 2086516

February 8, 2024

1 Ex. 10

We want to see how the MAP estimate changes according to the choice of the prior for the linear fit problem.

In this problem we have a dataset $(d_1, \ldots, d_n)$ of measurement taken at time/position $\vec{x}_i$. The average of the d_i is $\langle d_i \rangle = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij}$. We assume that noise is Gaussian. Since the error is distributed according to a Gaussian, in this framework we are going to assume that the Likelihood is going to be:

$$
\begin{aligned}
\mathcal{L}(\vec{d}|\vec{\beta}) &= \prod_{i}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(d_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2}{2\sigma^2}\right) \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\sum_{i}^{n}(d_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2}{2\sigma^2}\right)
\end{aligned}
\tag{1}
$$

Assuming a uniform prior $\pi(\vec{\beta})$, it is clear that the MAP, beeing the Likelihood an exponential with a negative exponent, it is found by minimazing the RSS:

$$
\mathbf{RSS} = \sum_{i}^{n}(d_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 \tag{2}
$$

Let's try to change the Prior $\pi(\vec{\beta})$.

1.1 Laplace Prior

We are going to assume that the parameters are i.i.d according to the following Laplacian distribution:

$$
\begin{aligned}
\pi(\vec{\beta}) &= \prod_{i}^{p} \frac{1}{b} \exp\left(-\frac{|\beta_i|}{b}\right) \\
&= \frac{1}{b} \exp\left(-\frac{\sum_{i}^{p} |\beta_i|}{b}\right)
\end{aligned}
\tag{3}
$$

where b is some fixed scale parameter. With this choice of Prior we can write the Posterior as:

$$
\begin{aligned}
P(\vec{\beta}|\vec{d}) &\propto \mathcal{L}(\vec{d}|\vec{\beta}) \times \pi(\vec{\beta}) \\
&\propto \frac{1}{b\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\sum_i^n (d_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2}{2\sigma^2} - \frac{\sum_{k=0}^p |\beta_k|}{b}\right) \\
&\propto \frac{1}{b\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_i^n (d_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \frac{2\sigma^2}{b}\sum_{k=0}^p |\beta_k|\right]\right)
\end{aligned} \tag{4}
$$

As we can clearly see the Posterior is an exponential with negative exponent, so in order to obtain the MAP we can just to minimize the expression in square brackets with respect to $\vec{\beta}$. In the end we obtain the MAP estimation for the parameters can be found as:

$$
\begin{aligned}
\hat{\vec{\beta}} &= \arg\min_{\vec{\beta}} \left(\sum_i^n (d_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \frac{2\sigma^2}{b}\sum_{k=0}^p |\beta_k|\right) \\
&= \arg\min_{\vec{\beta}} \left(\text{RSS} + \frac{2\sigma^2}{b}\sum_{k=0}^p |\beta_k|\right)
\end{aligned} \tag{5}
$$

This expression is called "LASSO regression".

1.2 Gaussian prior

We are going to assume that the parameters are i.i.d according to the following Gaussian distribution:

$$
\begin{aligned}
\pi(\vec{\beta}) &= \prod_{i=0}^p \frac{1}{\sqrt{2\pi c}} \exp\left(-\frac{\beta_i^2}{2c}\right) \\
&= \frac{1}{\sqrt{2\pi c}} \exp\left(-\frac{\sum_{i=0}^p \beta_i^2}{2c}\right)
\end{aligned} \tag{6}
$$

where c represent the variance of the Gaussian distribution. With this choice of Prior we can write the Posterior as:

$$
\begin{aligned}
P(\vec{\beta}|\vec{d}) &\propto \mathcal{L}(\vec{d}|\vec{\beta}) \times \pi(\vec{\beta}) \\
&\propto \frac{1}{\sqrt{2\pi c\sigma^2}} \exp\left(-\frac{\sum_i^n (d_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2}{2\sigma^2} - \frac{\sum_{k=0}^p \beta_k^2}{2c}\right) \\
&\propto \frac{1}{\sqrt{2\pi c\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}\left[\sum_i^n (d_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \frac{\sigma^2}{c}\sum_{k=0}^p \beta_k^2\right]\right)
\end{aligned} \tag{7}
$$

For the same argument of the Laplacian prior we obtain the MAP estimation for the parameters can be found as:

$$
\begin{aligned}
\hat{\vec{\beta}} &= \arg\min_{\vec{\beta}} \left(\sum_{i}^{n} (d_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \frac{\sigma^2}{c} \sum_{k=0}^{p} \beta_k^2 \right) \\
&= \arg\min_{\vec{\beta}} \left(\text{RSS} + \frac{\sigma^2}{c} \sum_{k=0}^{p} \beta_k^2 \right)
\end{aligned} \tag{8}
$$

This expression is called "Ridge regression".

2 Ex. 12

We assume that our data in $D = \{d_1, \dots, d_N\}$ are modelled by a sinusoidal signal $f(t)$ plus Gaussian noise $\vec{n}$ with known variance σ^2. So the data are fitted by:

$$
d_i = \underbrace{B_1 \cos(\omega t_i) + B_2 \sin(\omega t_i)}_{f(t_i)} + n_i \tag{9}
$$

Given this assumption, and the fact that we know that each measurement is independent, we can write the Likelihood of D as the product of the Likelihood of each d_i as:

$$
\begin{aligned}
\mathcal{L}(D|\omega, B_1, B_2) &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(d_i - f(t_i))^2}{2\sigma^2} \right) \\
&= (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{\sum_{i=1}^{N} (d_i - f(t_i))^2}{2\sigma^2} \right) \\
&= (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{\sum_{i=1}^{N} (d_i^2 + f^2(t_i) - 2 d_i f(t_i))}{2\sigma^2} \right) \\
&= (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{Q}{2\sigma^2} \right)
\end{aligned} \tag{10}
$$

Since we want to reconstruct a specific form of the numerator of the exponential we are gonna isolate and manipulate it:

$$
\begin{aligned}
Q &= \sum_{i=1}^{N} d_i^2 + f^2(t_i) - 2d_i f(t_i) \\
&= N\bar{d_i^2} - 2\left[B_1 \underbrace{\sum_{i=1}^{N} d_i \cos(\omega t_i)}_{R(\omega)} + B_1 \underbrace{\sum_{i=1}^{N} d_i \sin(\omega t_i)}_{I(\omega)}\right] + \\
&+ B_1^2 \underbrace{\sum_{i=1}^{N} \cos^2(\omega t_i)}_{s} + B_2^2 \underbrace{\sum_{i=1}^{N} \sin^2(\omega t_i)}_{c} + 2B_1B_2 \sum_{i=1}^{N} \cos(\omega t_i)\sin(\omega t_i) = \\
&= N\bar{d_i^2} - 2\left[B_1 R(\omega) + B_2 I(\omega)\right] + B_1^2 s + B_2^2 c + 2B_1B_1 \sum_{i=1}^{N} \cos(\omega t_i)\sin(\omega t_i) \\
&= N\bar{d_i^2} - 2\left[B_1 R(\omega) + B_2 I(\omega)\right] + B_1^2 s + B_2^2 c + B_1B_1 \sum_{i=1}^{N} \sin(2\omega t_i)
\end{aligned}
\tag{11}
$$

In order to show that the expression of Q that we have found is the one that we expected, and that both c and s are independent of ω, we need to consider the fact that we are working in a large frequency limit for the signal. This approximation implies both the $\Delta < \omega^{-1}$ and that we have sampled many periods of the signal during the detection. In practise we assume that our ωt_i (modulus 2π) data are uniformly distributed in the $[0, 2\pi]$ interval, so the in the end we get that:

$$
\begin{aligned}
\sum_{i=1}^{N} \sin(2\omega t_i) &\approx N\langle \sin(2\omega t)\rangle = 0 \\
c = \sum_{i=1}^{N} \sin^2(\omega t_i) &\approx N\langle \sin^2(\omega t_i)\rangle = \frac{N}{2} \\
s = \sum_{i=1}^{N} \cos^2(\omega t_i) &\approx N\langle \cos^2(\omega t_i)\rangle = \frac{N}{2}
\end{aligned}
\tag{12}
$$

With this consideration we can rewrite Q as:

$$
Q = N\bar{d_i^2} - 2\left[B_1 R(\omega) + B_2 I(\omega)\right] + B_1^2 \frac{N}{2} + B_2^2 \frac{N}{2} \tag{13}
$$

2.1 Marginalized Posterior

Considering uniform priors on the three parameters of the problem (ω, B_1, B_2), and treating the amplitudes as nuisance parameters, we can find the marginalized 1-parameter posterior for ω as follows:

$$
\begin{aligned}
P(\omega|D,I) &= \int dB_1\, dB_2\, P(\omega, B_1, B_2|D) \propto \\
&\propto \int dB_1\, dB_2\, \mathcal{L}(D|\omega, B_1, B_2)\, \underbrace{\pi(\omega, B_1, B_2)}_{\text{cost.}} \\
&\propto \int dB_1\, dB_2\, \underbrace{(2\pi\sigma^2)^{\frac{N}{2}}}_{\text{cost.}} \exp\left(-\frac{Q}{2\sigma^2}\right) \\
&\propto \int dB_1\, dB_2\, \exp\left(-\frac{1}{2\sigma^2}\left(\underbrace{N\bar{d_i^2}}_{\text{cost.}} - 2\left[B_1 R(\omega) + B_2 I(\omega)\right] + B_1^2 s + B_2^2 c\right)\right) \\
&\propto \int dB_1\, dB_2\, \exp\left(-\frac{1}{2\sigma^2}\left(\underbrace{B_1^2 s - 2B_1 R(\omega) + \frac{R^2(\omega)}{s}}_{s\left(B_1 - \frac{R(\omega)}{s}\right)^2} - \frac{R^2(\omega)}{s} + \underbrace{B_2^2 c - 2B_2 I(\omega) + \frac{I^2(\omega)}{c}}_{c\left(B_2 - \frac{I(\omega)}{c}\right)^2} - \frac{I^2(\omega)}{c}\right)\right) \qquad (14) \\
&\propto \exp\left(\frac{1}{2\sigma^2}\left(\frac{R^2(\omega)}{s} + \frac{I^2(\omega)}{c}\right)\right) \underbrace{\int dB_1\, \exp\left(-\frac{s}{2\sigma^2}\left(B_1 - \frac{R(\omega)}{s}\right)^2\right) \int dB_2 \exp\left(-\frac{c}{2\sigma^2}\left(B_2 - \frac{I(\omega)}{c}\right)^2\right)}_{\text{for each of them we use } \int dx e^{-a(x-b)^2} = \sqrt{\frac{2\pi}{a}}} \\
&\propto \exp\left(\frac{1}{2\sigma^2}\left(\frac{R^2(\omega)}{s} + \frac{I^2(\omega)}{c}\right)\right) \sqrt{\frac{4\pi\sigma^2}{s}} \sqrt{\frac{4\pi\sigma^2}{c}} \\
&\propto \frac{1}{\sqrt{sc}} \exp\left(\frac{1}{2\sigma^2}\left(\frac{R^2(\omega)}{s} + \frac{I^2(\omega)}{c}\right)\right) \\
&\propto \frac{2}{N} \exp\left(\frac{1}{N\sigma^2}(R^2(\omega) + I^2(\omega))\right)
\end{aligned}
$$

We have brought the constants ('cost.') outside the integral and in the last expression we have used the results obtain in eq. 12.

We want to show that the MAP estimate of the frequency of the signal can be obtain as the value of ω which maximise the Periodogram $C(\omega)$. The calculation are as follow:

$$
\begin{aligned}
P(\omega|D,I) &\propto \frac{2}{N}\exp\left(\frac{1}{N\sigma^2}(R^2(\omega)+I^2(\omega))\right) \\
&\propto \frac{2}{N}\exp\left(\frac{1}{N\sigma^2}\left[\left(\sum_{i=1}^{N} d_i\cos(wt_i)\right)^2+\left(\sum_{i=1}^{N} d_i\sin(wt_i)\right)^2\right]\right) \\
&\propto \frac{2}{N}\exp\left(\frac{1}{N\sigma^2}\left|\sum_{i=1}^{N} d_i\left(\cos(wt_i)+i\sin(wt_i)\right)\right|^2\right) \\
&\propto \frac{2}{N}\exp\left(\frac{1}{N\sigma^2}\left|\sum_{i=1}^{N} d_i\,\exp(-i\omega t_i)\right|^2\right)
\end{aligned} \tag{15}
$$

The last expression maximum is also the maximum of the Periodogram $C(\omega)$.

2.2 Least squares fitting

Least squares fitting means that we need to minimize $\chi^2 = Q/\sigma^2$. A Gaussian likelihood for ω can be guaranteed by the following conditions:

1. linear dependace of the model $f(t)$ on ω.
2. i.i.d. Gaussian noise on the data.

The second condition is guaranteed from our assumptions. The first condition is not always guaranteed for all the values of t_i, but if we remain in a interval where $\omega t_i \approx 0$, then we can expand our model to first order $f(t_i) \approx B_1 + B_2\omega t_i$. In general the Likelihood is not Gaussian, and the model is not linear.