Exercise sheet 6

Exercise 6 - 1

It was shown in the lecture that for arbitrary signal-, noise-, and data-statistics with known correlations $\langle ss^\dagger \rangle_{(s,d)}$, $\langle ds^\dagger \rangle_{(s,d)}$, and $\langle dd^\dagger \rangle_{(s,d)}$, the optimal linear filter is given by

$$m = \langle sd^\dagger \rangle_{(s,d)} \langle dd^\dagger \rangle_{(s,d)}^{-1} d. \tag{1}$$

A linear response matrix R and a noise covariance matrix N can be *defined* via the following identifications:

$$\langle ss^\dagger \rangle_{(s,d)} \equiv S \tag{2}$$
$$\langle ds^\dagger \rangle_{(s,d)} \equiv RS \tag{3}$$
$$\langle dd^\dagger \rangle_{(s,d)} \equiv RSR^\dagger + N \tag{4}$$

Find expressions for R and N in terms of $\langle ss^\dagger \rangle_{(s,d)}$, $\langle ds^\dagger \rangle_{(s,d)}$, and $\langle dd^\dagger \rangle_{(s,d)}$(2 points).

Exercise 6 - 2

You are interested in three numbers, $s = (s_1, s_2, s_3) \in \mathbb{R}^3$. Your measurement device, however, only measures three differences between the numbers, according to

$$d_1 = s_1 - s_2 + n_1 \tag{5}$$
$$d_2 = s_2 - s_3 + n_2 \tag{6}$$
$$d_3 = s_3 - s_1 + n_3 \tag{7}$$

with some noise vector $n \in \mathbb{R}^3$. Assume a Gaussian prior $\mathcal{P}(s) = \mathcal{G}(s, S)$ for s and a Gaussian PDF for the noise, $\mathcal{P}(n) = \mathcal{G}(n, N)$, with $N_{ij} = \sigma^2 \delta_{ij}$.

a) Assume that the prior is degenerate, i.e., $S^{-1} \equiv 0$. Write down the response matrix, try to give the posterior $\mathcal{P}(s|d)$, and explain why this is problematic (2 points).

b) Now assume that $S_{ij} = \sigma^2 \delta_{ij}$. Work out the posterior $\mathcal{P}(s|d)$ in this case (1 point).

<u>Note</u>: Using a computer algebra system for the matrix operations is okay.

Exercise 6 - 3

The numbers quantifying the degree of industrialization $\tilde{\iota}$ of a society, its fertility rate $\tilde{f}$, and the stork population $\tilde{s}$ on its territory are random variables assumed to belong to a joint three-dimensional Gaussian distribution. Consider the fluctuations around the respective mean values $\iota = \tilde{\iota} - \langle \tilde{\iota} \rangle_{(\tilde{\iota})}$, $f = \tilde{f} - \langle \tilde{f} \rangle_{(\tilde{f})}$, and $s = \tilde{s} - \langle \tilde{s} \rangle_{(\tilde{s})}$. It is known that both the stork index s and the fertility index f are anticorrelated with the degree of industrialization. The normalized correlation coefficients are $c_{s\iota} = -0.85$ and $c_{f\iota} = -0.70$, where

$$c_{ab} = \frac{\langle ab \rangle_{(a,b)}}{\sqrt{\langle aa \rangle_{(a)} \langle bb \rangle_{(b)}}}. \tag{8}$$

Assume further that there is no direct correlation between f and s, i.e., $\mathcal{P}(s|f, \iota) = \mathcal{P}(s|\iota)$ and $\mathcal{P}(f|s, \iota) = \mathcal{P}(f|\iota)$. Derive an expression for $\langle sf \rangle_{(s,f,\iota)}$. Use this to calculate the normalized correlation coefficient c_{sf} (2 points).

Exercise 6 - 4

You have conducted a measurement of a quantity at n positions $\{x_i\}_i$, yielding n data points $\{(x_i, d_i)\}_i$. Now you want to fit some function to these data points. To this end, you write the function as a linear combination of m basis functions $\{f_j(x)\}_j$, i.e.,

$$f(x) = \sum_{j=1}^{m} s_j f_j(x). \tag{9}$$

If, for example, you were to fit a second order polynomial, you could choose the monomials as basis functions, i.e., $f(x) = s_2 x^2 + s_1 x + s_0$.
The fitting process now comes down to determining the coefficients $\{s_j\}_j$, allowing for some Gaussian and independent measurement error, i.e.,

$$d_i = \sum_{j=1}^{m} s_j f_j(x_i) + n_i. \tag{10}$$

Assume that you do not know anything about the coefficients a priori, i.e. take a flat improper prior, $S^{-1} \equiv 0$, where $S_{ik} = \langle s_i s_k \rangle_{\mathcal{P}(s)}$.

a) Write down the response matrix for this problem (1 point).

b) For a given set of m basis functions, how many data points n are at least necessary for the calculation of the posterior mean of the coefficients (2 points)?

c) Now let's make a linear fit. Assuming $N_{ik} = \langle n_i n_k \rangle_{\mathcal{P}(n)} = \eta^{-1}\delta_{ik}$, choose two basis functions and work out the explicit formula for the posterior mean of the two coefficients (3 points).

This exercise sheet will be discussed in the tutorials.
group A, Wednesday 16:00 - 18:00 Theresienstr. 39 - room B 101,
group B, Thursday, 10:00 - 12:00, Geschw.-Scholl-Pl. 1 (B) - room B 015
group C, Thursday, 16:00 - 18:00, Theresienstr. 37 - room A 449
backup slot, Wednesday, 8:00 - 10:00, Theresienstr. 37 - A 450
(to be used whenever Thursday is a vacation day)
www.mpa-garching.mpg.de/∼ensslin/lectures