Aalto University
School of Electrical
Engineering

Exercise Round 2

The deadline of this exercise round is **Wednesday January 22, 2020**. The solutions will be gone through during the exercise session in room T2 in Konemiehentie 2 (CS) on that day starting at 14:15.

The problems should be *solved before the exercise session*, and during the session those who have completed the exercises may be asked to present their solutions on the board/screen.

Exercise 1. (Regression as State Estimation)

Consider the regression problem

$$\begin{aligned} y_k &= a_1\, s_k + a_2\, \sin(s_k) + b + \varepsilon_k, \qquad k = 1, \ldots, T, \\ \varepsilon_k &\sim \mathrm{N}(0, R), \\ a_1 &\sim \mathrm{N}(0, \sigma_1^2), \\ a_2 &\sim \mathrm{N}(0, \sigma_2^2), \\ b &\sim \mathrm{N}(0, \sigma_b^2), \end{aligned} \tag{1}$$

where $s_k \in \mathbb{R}$ are the known regressor values, $R, \sigma_1^2, \sigma_2^2, \sigma_b^2$ are given positive constants, $y_k \in \mathbb{R}$ are the observed output variables, and ε_k are independent Gaussian measurement errors. The scalars a_1, a_2, and b are the unknown parameters to be estimated. Formulate the estimation problem as a linear Gaussian state space model.

Exercise 2. (Linear Bayesian Estimation)

Assume that in the linear regression model from the previous exercise round (Ex. 1.2), we set independent Gaussian priors for the parameters θ_1 and θ_2 as follows:

$$\begin{aligned} \theta_1 &\sim \mathrm{N}(0, \sigma^2), \\ \theta_2 &\sim \mathrm{N}(0, \sigma^2), \end{aligned}$$

where the variance σ^2 is known. The measurements y_k are modeled as

$$y_k = \theta_1\, x_k + \theta_2 + \varepsilon_k, \qquad k = 1, 2, \ldots, T,$$

where the ε_k are independent Gaussian error terms with mean 0 and variance 1, that is, $\varepsilon_k \sim \mathrm{N}(0, 1)$. The values x_k are fixed and known. The posterior distribution can be now written as

$$
\begin{aligned}
& p(\boldsymbol{\theta} \mid y_1, \ldots, y_T) \\
& \propto \exp\left(-\frac{1}{2} \sum_{k=1}^{T} (y_k - \theta_1\, x_k - \theta_2)^2 \right) \exp\left(-\frac{1}{2\sigma^2} \theta_1^2 \right) \exp\left(-\frac{1}{2\sigma^2} \theta_2^2 \right).
\end{aligned}
$$

The posterior distribution can be seen to be Gaussian and your task is to derive its mean and covariance.

(a) Write the exponent of the posterior distribution in matrix form as in Exercise 1 on round 1 (in terms of $\mathbf{y}$, $\mathbf{X}$, $\boldsymbol{\theta}$ and σ^2).

(b) Because a Gaussian distribution is always symmetric, its mean $\mathbf{m}$ is at the maximum of the distribution. Find the posterior mean by computing the gradient of the exponent and finding where it vanishes.

(c) Find the covariance of the distribution by computing the second derivative matrix (Hessian matrix) $\mathbf{H}$ of the exponent. The posterior covariance is then $\mathbf{P} = -\mathbf{H}^{-1}$ (why?).

(d) What is the resulting posterior distribution? What is the relationship with the least squares estimate in Exercise 2 on round 1?

Exercise 3. (Gaussian Identities)

Recall that the Gaussian probability density is defined as

$$
\mathrm{N}(\mathbf{x} \mid \mathbf{m}, \mathbf{P}) = \frac{1}{(2\,\pi)^{n/2}\, |\mathbf{P}|^{1/2}} \exp\left(-\frac{1}{2} (\mathbf{x} - \mathbf{m})^{\mathsf{T}}\, \mathbf{P}^{-1}\, (\mathbf{x} - \mathbf{m}) \right)
$$

Derive the following Gaussian identities:

(a) Let $\mathbf{x}$ and $\mathbf{y}$ have the Gaussian densities

$$
p(\mathbf{x}) = \mathrm{N}(\mathbf{x} \mid \mathbf{m}, \mathbf{P}), \qquad p(\mathbf{y} \mid \mathbf{x}) = \mathrm{N}(\mathbf{y} \mid \mathbf{H}\,\mathbf{x}, \mathbf{R}),
$$

then the joint distribution of $\mathbf{x}$ and $\mathbf{y}$ is

$$
\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \mathbf{m} \\ \mathbf{H}\,\mathbf{m} \end{pmatrix}, \begin{pmatrix} \mathbf{P} & \mathbf{P}\,\mathbf{H}^{\mathsf{T}} \\ \mathbf{H}\,\mathbf{P} & \mathbf{H}\,\mathbf{P}\,\mathbf{H}^{\mathsf{T}} + \mathbf{R} \end{pmatrix} \right)
$$

and the marginal distribution of $\mathbf{y}$ is

$$\mathbf{y} \sim \mathrm{N}(\mathbf{H}\,\mathbf{m}, \mathbf{H}\,\mathbf{P}\,\mathbf{H}^\mathsf{T} + \mathbf{R}).$$

Hint: Use the properties of expectation $\mathrm{E}[\mathbf{H}\,\mathbf{x} + \mathbf{r}] = \mathbf{H}\;\mathrm{E}[\mathbf{x}] + \mathrm{E}[\mathbf{r}]$ and $\mathrm{Cov}[\mathbf{H}\,\mathbf{x} + \mathbf{r}] = \mathbf{H}\;\mathrm{Cov}[\mathbf{x}]\,\mathbf{H}^\mathsf{T} + \mathrm{Cov}[\mathbf{r}]$ (if $\mathbf{x}$ and $\mathbf{r}$ independent).

(b) Write down the explicit expression for the joint and marginal probability densities above:

$$p(\mathbf{x}, \mathbf{y}) = p(\mathbf{y} \mid \mathbf{x})\, p(\mathbf{x}) = ?$$
$$p(\mathbf{y}) = \int p(\mathbf{y} \mid \mathbf{x})\, p(\mathbf{x})\, \mathrm{d}\mathbf{x} = ?$$

(c) If the random variables $\mathbf{x}$ and $\mathbf{y}$ have the joint Gaussian probability density

$$\begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} \boldsymbol{\mu}_x \\ \boldsymbol{\mu}_y \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{xy}^\mathsf{T} & \boldsymbol{\Sigma}_{yy} \end{pmatrix} \right),$$

then the conditional density of $\mathbf{x}$ given $\mathbf{y}$ is

$$\mathbf{x} \mid \mathbf{y} \sim \mathrm{N}(\boldsymbol{\mu}_x + \boldsymbol{\Sigma}_{xy}\, \boldsymbol{\Sigma}_{yy}^{-1}\, (\mathbf{y} - \boldsymbol{\mu}_y), \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\, \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{xy}^\mathsf{T}).$$

Hints:

- Denote the inverse covariance as $\boldsymbol{\Phi} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^\mathsf{T} & \mathbf{D} \end{pmatrix}$ and expand the quadratic form in the Gaussian exponent.
- Compute the derivative with respect to $\mathbf{x}$ and set it to zero. Conclude that due to symmetry the point where the derivative vanishes is the mean.
- Check from a linear algebra book that the inverse of $\mathbf{A}$ is given by the Schur complement:

  $$\mathbf{A}^{-1} = \boldsymbol{\Sigma}_{xx} - \boldsymbol{\Sigma}_{xy}\, \boldsymbol{\Sigma}_{yy}^{-1} \boldsymbol{\Sigma}_{xy}^\mathsf{T}$$

  and that $\mathbf{B}$ can be then written as

  $$\mathbf{B} = -\mathbf{A}\, \boldsymbol{\Sigma}_{xy}\, \boldsymbol{\Sigma}_{yy}^{-1}.$$

- Find the simplified expression for the mean by applying the identities above.
- Find the second derivative of the negative Gaussian exponent with respect to $\mathbf{x}$. Conclude that it must be the inverse conditional covariance of $\mathbf{x}$.
- Use the Schur complement expression above for computing the conditional covariance.