

## Machine Learning Worksheet 9

Tomas Ladek, Michael Kratzer  
3602673, 3612903  
tom.ladek@tum.de, mkratzer@mytum.de

---

### Problem 1

Let us transform the row vectors of  $X$  to column vectors so that we can input them directly into the covariance function:

$$x_1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

$$x_2 = \begin{pmatrix} 0.5 \\ 2 \end{pmatrix}$$

Noise-free case means mean is zero:  $m(x_*) = 0 = \mu_*$ . Then:

$$\begin{aligned} \mathbf{K} &= \exp\left(-\frac{1}{2}(x - x')^T(x - x')\right) \\ &= \exp\left(-\frac{1}{2}\begin{pmatrix} x_1 - x_1 & x_1 - x_2 \\ x_2 - x_1 & x_2 - x_2 \end{pmatrix}^T \begin{pmatrix} x_1 - x_1 & x_1 - x_2 \\ x_2 - x_1 & x_2 - x_2 \end{pmatrix}\right) \\ &= \exp\left(-\frac{1}{2}\begin{pmatrix} 0 & -0.5 \\ 0 & -1 \\ 0.5 & 0 \\ 1 & 0 \end{pmatrix}^T \begin{pmatrix} 0 & -0.5 \\ 0 & -1 \\ 0.5 & 0 \\ 1 & 0 \end{pmatrix}\right) \\ &= \exp\left(-\frac{1}{2}\begin{pmatrix} 1.25 & 0 \\ 0 & 1.25 \end{pmatrix}\right) \\ &= \begin{pmatrix} 0.54 & 1 \\ 1 & 0.54 \end{pmatrix} \end{aligned}$$

$$\begin{aligned} \mathbf{K}_* &= \exp\left(-\frac{1}{2}\begin{pmatrix} x_1 - x_* \\ x_2 - x_* \end{pmatrix}^T \begin{pmatrix} x_1 - x_* \\ x_2 - x_* \end{pmatrix}\right) \\ &= \exp\left(-\frac{1}{2}\begin{pmatrix} -0.5 & 0 \\ 0 & 1 \end{pmatrix}^T \begin{pmatrix} -0.5 & 0 \\ 0 & 1 \end{pmatrix}\right) \\ &= \exp\left(-\frac{1}{2}\begin{pmatrix} 0.25 & 0 \\ 0 & 1 \end{pmatrix}\right) \\ &= \begin{pmatrix} 0.88 & 1 \\ 1 & 0.61 \end{pmatrix} = \mathbf{K}_*^T \end{aligned}$$

$$\begin{aligned} \mathbf{K}_{**} &= \exp\left(-\frac{1}{2}(x_* - x_*)^T(x_* - x_*)\right) \\ &= \exp(0) = 1 \end{aligned}$$


---

Therefore

$$f_{jt} = \begin{pmatrix} y_1 \\ y_2 \\ f(x_*) \end{pmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} \mathbf{K} & \mathbf{K}_* \\ \mathbf{K}_*^T & \mathbf{K}_{**} \end{bmatrix}\right) = \mathcal{N}\left(0, \begin{pmatrix} 0.54 & 1 & 0.88 & 1 \\ 1 & 0.54 & 1 & 0.61 \\ 0.88 & 1 & 1 & 0 \\ 1 & 0.61 & 0 & 1 \end{pmatrix}\right)$$

## Problem 2

?

## Problem 3

The joint distribution in the noisy case is given by

$$\begin{bmatrix} \mathbf{y} \\ f_* \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu \\ \mu_* \end{bmatrix}, \begin{bmatrix} \mathbf{K} + \sigma_n^2 \mathbf{I} & \mathbf{K}_* \\ \mathbf{K}_*^T & \mathbf{K}_{**} \end{bmatrix}\right)$$

which in our case evaluates to

$$\mathcal{N}\left(0, \begin{pmatrix} 0.54 + \sigma_1^2 & 1 & 0.88 & 1 \\ 1 & 0.54 + \sigma_2^2 & 1 & 0.61 \\ 0.88 & 1 & 1 & 0 \\ 1 & 0.61 & 0 & 1 \end{pmatrix}\right)$$