Please submit your solutions via Ilias. The submission is not a formal requirement for passing the exam but doing the exercises will be very helpful to do so. Submissions should be a single PDF document (note that Jupyter notebooks can and should also be downloaded as PDFs, and not only submitted as .ipynb files).

1. **Theory Question 1** You want to estimate the room temperature and buy three thermometers. These three thermometers measure the room temperature at different accuracies; the manufacturers specify accuracy as the standard deviation of the measurements (in degrees):

   T1: $\sigma_1 = 5.0$, T2: $\sigma_2 = 1.0$, T3: $\sigma_3 = 0.1$

   Upfront, you assume an improper (i.e. unnormalized) uniform prior across all temperatures.

   (a) (a) You measure the temperature with T1 and the measured value is 32 degrees. What is the posterior distribution over the room temperature?

   (b) (b) T2 measures a distribution of 31 degrees. Given measurements of T1 and T2, what is the posterior distribution over the room temperature?

   (c) (c) T3 measures a temperature of 22 degrees. Given measurements of T1, T2, and T3, what is the posterior distribution over the room temperature?

   (d) (d) Do you trust the inference result? Do you believe that the accuracies reported by the manufacturers are accurate?

2. **Theory Question 2** Consider the Gaussian random variable $\boldsymbol{w} \in \mathbb{R}^F$ with probability density function $p(\boldsymbol{w}) = \mathcal{N}(\boldsymbol{w}; \boldsymbol{\mu}, \Sigma)$ where $\boldsymbol{\mu} \in \mathbb{R}^F$ and symmetric positive definite $\Sigma \in \mathbb{R}^{F \times F}$. You have access to data $\boldsymbol{y} \in \mathbb{R}^N$ assumed to be generated from $\boldsymbol{w}$ through a linear map $\Phi \in \mathbb{R}^{F \times N}$ according to the likelihood

   $$p(\boldsymbol{y}|\boldsymbol{w}) = \mathcal{N}(\boldsymbol{y}; \Phi^T \boldsymbol{w}, \Lambda),$$

   where $\Lambda \in \mathbb{R}^{N \times N}$ is symmetric positive definite.

   Consider the special case $\Lambda = \sigma^2 I$ with $\sigma^2 \in \mathbb{R}_+$ (that is, iid. observation noise).

   (a) Show that the **maximum likelihood estimator** for $\boldsymbol{w}$ is given by the **ordinary least-squares** estimate
   $$\boldsymbol{w}_{ML} = (\Phi\Phi^T)^{-1}\Phi\boldsymbol{y}.$$

   (b) Show that the **maximum a-posteriori estimator** is identical to the posterior mean, $\boldsymbol{w}_{\text{MAP}} = \mathbb{E}_{p(\boldsymbol{w}|\boldsymbol{y})}(\boldsymbol{w})$ (you can use the fact that the posterior is Gaussian).

   (c) There exists an important relationship between the regularization of least squares estimates and the choice of the prior in probabilistic linear regression. Given the Gaussian prior $p(\boldsymbol{w})$ for the particular choice $\boldsymbol{\mu} = 0, \Sigma = I_F, \Lambda = \sigma^2 I$, show that the MAP estimator calculated in part (b) is equivalent to the **$\boldsymbol{l}_2$-regularized least-squares** estimator (aka ridge regression)

   $$\boldsymbol{w}_{\boldsymbol{l}_2} = (\Phi\Phi^T + \alpha I)^{-1}\Phi\boldsymbol{y},$$

   and give the corresponding value of the regularization parameter α.

   (d) Which choice of prior would a LASSO ($\boldsymbol{l}_1$) regularization correspond to?

3. **Coding Question** See `Exercise_03.ipynb`.