To be completed individually or in groups of two people. **Please be sure matriculation numbers are clearly included at the top of your submission.** Submissions can be handwritten or in LaTeX formatting, but hard-to-read handwritten submissions will not be graded.

Please submit via Ilias. Submissions should be a single PDF document (note that Jupyter notebooks can and should also be downloaded as PDFs, and not submitted as .ipynb files).

Each question will be graded "pass" (full points) or "fail" (no points). We award 0.5 bonus points for the exam for each theory and practical question solved. You must complete 50% of all exercises to enter the final exam.

1. **EXAMple Question** *(Bayes Estimators)* Consider a classification problem, where you can assign a datapoint x to one of C classes. There is an additional **reject option**, that is, the action for the datapoint x_i can range between $\{1 \dots C, C+1\}$, where $C+1$ means that you have rejected to classify the datapoint. (If the cost of rejects is less than the cost of false classification, this may be the optimal action.)

   Given x, Let $\theta = j$ denote the true class, and $\hat{\theta}$ the action. The loss function for *posterior risk* in this case is defined as

$$L(\hat{\theta} = i, \theta = j) = \begin{cases} 0 & \text{if } i = j \text{ and } i, j \in \{1, \dots, C\} \\ r & \text{if } i = C+1 \\ 1 & \text{otherwise} \end{cases}$$

   (a) Assuming a discrete posterior distribution on the classes $p(\theta = j|x)$, write down the *posterior risk* $\mathbb{E}_{p(\theta|x)}[L(\hat{\theta}, \theta)]$. You can use the indicator functions $\mathbf{I}[\cdot]$ with boolean arguments inside.

   (b) Show that for $j \in \{1 \dots C\}$, the optimal action corresponding to minimum risk is $\hat{\theta} = j$, *only if*

      i. $p(\theta = j|x) \geq p(\theta = k|x)$ for all $k \neq j$, *and*

      ii. $p(\theta = j|x) \geq 1 - r$.

      Also show that in case these conditions are false, the optimal action is to instead reject.

   (c) What is the behaviour of the optimal action when rejection cost r varies from 0 to 1?

2. **Theory Question** *(Conjugate Prior Inference)* Assume we are given n independent draws $\mathbf{x}_i \in \mathbb{R}^d$, $i = 1, \dots, n$ from the *multivariate* Gaussian distribution.

$$p(\mathbf{x}_i|\Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mu)^T \Sigma^{-1} (\mathbf{x}_i - \mu)\right)$$

   Assume that we know the mean $\mu \in \mathbb{R}^d$, but we do *not* know the symmetric positive definite covariance matrix Σ. The Gaussian distribution is a member of the exponential family, and all distributions in the exponential family have conjugate priors.

   (a) Show that the *Wishart distribution*, the exponential family with p.d.f.

$$\begin{aligned} \mathcal{W}(\Sigma^{-1}; V, \nu) &:= p(\Sigma^{-1}|V, \nu) \text{ with } \nu > (d-1) \in \mathbb{R}, \text{ and s.p.d. } V \in \mathbb{R}^{d \times d} \\ &= \frac{1}{2^{\nu d/2}|V|^{\nu/2}\Gamma_d(\nu/2)} |\Sigma^{-1}|^{(\nu-d-1)/2} \exp\left(-\mathrm{tr}\left(\Sigma^{-1}V^{-1}\right)/2\right) \end{aligned}$$

is the conjugate prior for the *inverse* covariance (aka. precision) matrix Σ^{-1}. Here Γ_d is the multivariate Gamma function[1], and $\mathrm{tr}(Z) = \sum_i [Z]_{ii}$ is the trace operation.

(b) What is the posterior distribution $p(\Sigma^{-1}|\mathbf{x}_1, \ldots, \mathbf{x}_n)$ over Σ^{-1}?

3. **Practical Question** See `Exercise_08.ipynb`.

[1]The multivariate Gamma function is defined as $\Gamma_d(a) = \pi^{d(d-1)/4} \prod_{j=1}^{d} \Gamma(a + (1-j)/2)$, but this is not needed for a solution to this exercise.