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** *(Logistic regression)*

   a) Show that $\sigma(s) = 1/(1+\exp(-s)) = \exp(s)/(1+\exp(s))$.

   b) Show that the logistic function satisfies $\sigma(-s) + \sigma(s) = 1$.

   c) Show that the first derivative of $\sigma(s)$, $\sigma'(s) = \sigma(s)(1-\sigma(s))$.

   d) Plot $\sigma(s)$ as well as $\log(\sigma(s))$ as a function of s (either using Python or with pen and paper, a rough plot which captures the qualitative features of the functions is sufficient).

   e) Explain why, for large $s > 0$, $\log(\sigma(s)) \approx 0$ and $\log(\sigma(-s)) \approx -s$.

   f) Suppose that we have data from two classes, and the data within each class is Gaussian distributed with the same covariance, i.e. $x|t=1 \sim \mathcal{N}(\mu_+, \Sigma)$ and $x|t=-1 \sim \mathcal{N}(\mu_-, \Sigma)$, and that the two classes the same prior probabilities $\pi_+ = P(t=+1) = \pi_- = P(t=-1) = 0.5$. Show that the conditional probability of belonging to the positive class can be written as a logistic function $P(t=1|x) = \sigma(\omega^\top x + \omega_o)$ and identify the corresponding parameters ω and ω_o.

2. **Theory Question** *(Generalized linear models with exponential family models)*

   Suppose we have a generalized linear model with exponential observation model

$$p(\boldsymbol{y}|\theta) = \frac{1}{Z(\theta)} h(\boldsymbol{y}) \, e^{\boldsymbol{y}^\top \theta}.$$

   We set set $\theta = w^\top x$ to get the GLM likelihood

$$p(y|w, \boldsymbol{x}) = p_{\mathrm{EF}}(\boldsymbol{y}|\theta = w^\top \boldsymbol{x}).$$

   a) For exponential family models, the expectation can be derived from differentiating the log-likelihood: Show that $\mathrm{E}_{p(\boldsymbol{y}|\theta)}[\boldsymbol{y}] = \nabla_\theta \log Z(\theta)|_{\theta = w^\top \boldsymbol{x}}$.

   b) Our goal is to find the Laplace approximation,

$$p(w|\boldsymbol{y}, \boldsymbol{X}) = p(w) \prod_{n=1}^{N} p_{\mathrm{EF}}(\boldsymbol{y}|\theta = w^\top \boldsymbol{x}_n) \approx \mathcal{N}(w|w_*, \Sigma_w).$$

   Show that the exponential family log-likelihood is given by $\log p(\boldsymbol{y}|w, X) = \sum_{n=1}^{N} \big(\phi(y) w^\top \boldsymbol{x}_n - \log Z(w^\top \boldsymbol{x}_n) + \log h(y)\big)$.

   c) To find mode w_* of $\log p(\boldsymbol{y}|w, X) + \log p(w)$ by gradient ascent, we need the gradient $\nabla_w \log p(\boldsymbol{y}|w, X)$– derive it!

   d) We also need the Hessian $\mathcal{H} = \nabla_w \nabla_w \ln p(w|y, X)$– derive it!

3. **Practical Question** See `Exercise_06.ipynb`.