Deadline: Fr. July 5 (online) or Mo. July 8, 10:00 AM (hand-in)
Drop your printed or legible handwritten submissions into the boxes at Samelsonplatz, or upload a `.pdf` file via LearnWeb.

Remark 1. This is the last hand in tutorial for this semester. For each 20 points you earned, you will get 0.5 bonus points in the exam. (capped at 4)

1 Proximal Gradients (10 points)

A. [10p] Apply SPARSA (ISTA with homotopy) to solve the LASSO problem

$$\text{minimize } f(\theta) = L(\theta) + \lambda R(\theta) \qquad L(\theta) = \tfrac{1}{2}\|y - X\theta\|_2^2 \qquad R(\theta) = \|\theta\|_1$$

on the dataset from Table 1 below. Use the parameters $s = 0.5$ and $\lambda = 0.1$ and the acceptance criterion $f(\theta) < f(\theta_{\text{old}})$ (called monotone criterion).

Hint: The algorithm should terminate after 4 outer loops

Algorithm 1: SPARSA algorithm (ISTA with homotopy) for LASSO

input : parameters $\lambda \geq 0$, $s \in (0, 1)$
output : A partition of the bitmap
init : $\theta = 0$, $\alpha = 1$, $\lambda_0 = \infty$
while $\lambda_t \neq \lambda$ **do**
  $\lambda_t = \max(s\|\nabla L(\theta)\|_\infty, \lambda)$
  repeat
    $\theta_{\text{old}} = \theta$
    $g = \nabla L(\theta)$
    $u = \theta - \frac{1}{\alpha} g$
    $\theta = \text{soft}(u, \frac{\lambda_t}{\alpha})$
    $\alpha = \text{BB}(\theta, \theta_{\text{old}})$
  until θ *satisfies acceptance criterion*
end
return : θ

Table 1

x_1	x_2	x_3	y
0	1	0	1
0.5	0.5	1	0.75
1	0	0	0.5

2 Laplace Priors (10 points)

Hint: Let X be a Random Variable (RV) with density function $f(x)$. Then $Y = X^{-1}$ has the density function $\frac{1}{y^2} f(\frac{1}{y})$.

A. [2p] Explain the role of the prior $p(\beta)$ in Bayesian Linear Regression

B. [4p] Given $p(\beta_i \mid \tau_i^2) = \mathcal{N}(\beta_i \mid 0, \tau_i^2)$ and $p(\tau_i^2) = \text{Exp}(\tau_i^2 \mid \frac{1}{2}\lambda^2)$, show that $z_i = \frac{1}{\tau_i^2}$ has the conditional distribution

$$p(z_i \mid \beta_i) = \text{InvGauss}\left(z \,\middle|\, \sqrt{\frac{\lambda^2}{\beta_i^2}}, \lambda^2\right) \tag{1}$$

C. [2p] Given that $p(y \mid X, \beta, \sigma^2) = \mathcal{N}(y \mid X\beta, \sigma^2 I)$ and $p(\sigma^2) = \text{IG}(\sigma^2 \mid a, b)$, show that σ^2 has the conditional distribution

$$p(\sigma^2 \mid X, y, \beta, \tau^2) = \text{IG}(\sigma^2, a', b')$$

with $a' = a + \frac{1}{2}N$ and $b' = b + \frac{1}{2}\|y - X\beta\|_2^2$.

D. [2p] Show that in general $\mathbb{E}[X] \neq \mathbb{E}[X^{-1}]^{-1}$ by providing an example.