This homework is due at **2pm, October 4**.

Problem 1

The *characteristic function* for a set $\mathscr{X} \subseteq \mathbb{R}^p$ is defined as

$$\chi_{\mathscr{X}}(x) := \begin{cases} 0, & \text{if } x \in \mathscr{X}, \\ +\infty, & \text{otherwise.} \end{cases}$$

Show that the function $\chi_{\mathscr{X}}$ is convex, if and only if $\mathscr{X}$ is a convex set. (10 points)

Problem 2

Consider the linear regression model:

$$y_i := \langle x_i, \beta^\natural \rangle + w_i, \quad i = 1, 2, \ldots, n,$$

where $x_1, \ldots, x_n$ and $\beta^\natural$ are in $\mathbb{R}^p$, and $w_1, \ldots, w_n$ are independent and identically distributed Gaussian random variables of zero mean and unit variance. The *lasso (least absolute shrinkage and selection operator)* is a famous approach to estimating $\beta^\natural$ given $(x_1, y_1), \ldots, (x_n, y_n)$. The lasso is given by

$$\hat{\beta} \in \operatorname*{arg\,min}_{\beta} \left\{ f(\beta) \,\middle|\, \beta \in \mathbb{R}^p \right\},$$

where

$$f(\beta) := \frac{1}{n} \sum_{i=1}^{n} \left(y_i - \langle x_i, \beta \rangle \right)^2 + \lambda \|\beta\|_1, \quad \forall \beta \in \mathbb{R}^p,$$

for some positive real number $\lambda > 0$.

1. **Show that the function $g(x) := \|x\|_1$ is convex.** (10 points)
2. **Show that the function f defined above is convex.** (10 points)

Problem 3

The logistic regression corresponds to minimizing a sum of the logistic losses

$$f_i(w) := \log\left(1 + \mathrm{e}^{-y_i \langle x_i, w \rangle}\right),$$

for some $(x_i, y_i) \in \mathbb{R}^p \times \{-1, +1\}$.

1. **Show that**

   $$\nabla f_i(w) = \frac{-y_i x_i}{1 + \mathrm{e}^{y_i \langle x_i, w \rangle}}.$$

   (10 points)

2. **Show that**

   $$\nabla^2 f_i(w) = \frac{\mathrm{e}^{y_i \langle x_i, w \rangle} x_i x_i^{\mathrm{T}}}{\left(1 + \mathrm{e}^{y_i \langle x_i, w \rangle}\right)^2} \geq 0.$$

   (10 points)

Problem 4

Let x_1 and x_2 be two distinct real numbers. Let ξ be a random variable taking values in the set $\{x_1, x_2\}$, following the probability distribution

$$\mathsf{P}(\xi = x_i) = p_i, \quad \forall i = 1, 2,$$

for some strictly positive real numbers p_1 and p_2 summing up to 1. Then the expectation of ξ is given by

$$\mathsf{E}[\xi] := \sum_{i=1}^{2} p_i x_i.$$

Consider the function

$$\varphi(t) := \log\left(\mathsf{E}\left[e^{t\xi}\right]\right), \quad \forall t \in \mathbb{R}.$$

1. **Show that** $\varphi''(t) > 0$ **for all** $t \in \mathbb{R}$**.** Therefore, the function φ is convex. (10 points)

2. **Show that**

$$|\varphi'''(t)| \leq M\varphi''(t), \quad \forall t \in \mathbb{R},$$

where $M := |x_1 - x_2|$**.** (10 points)

3. **Show that**

$$\varphi''(t_2) \geq e^{-M|t_2 - t_1|}\varphi''(t_1), \quad \forall t_1, t_2 \in \mathbb{R}.$$

(10 points)

HINT: You may apply the result that $\varphi''(t) > 0$ for all t even if you cannot find the proof, and consider the function $\psi(t) := \log\big(\varphi''(t)\big)$.

4. **Show that**

$$\varphi'(t_2) - \varphi'(t_1) \geq -\left(\frac{e^{-M|t_2 - t_1|} - 1}{M|t_2 - t_1|}\right)\varphi''(t_1)(t_2 - t_1), \quad \forall t_1, t_2 \in \mathbb{R} \textbf{ such that } t_2 > t_1.$$

(10 points)

5. **Show that**

$$\varphi(t_2) \geq \varphi(t_1) + \varphi'(t_1)(t_2 - t_1) + \frac{e^{-M|t_2 - t_1|} + M|t_2 - t_1| - 1}{M^2}\varphi''(t_1), \quad \forall t_1, t_2 \in \mathbb{R} \textbf{ such that } t_2 > t_1.$$

(10 points)

Notice that the three inequalities above are sharper than the positive semi-definiteness of the Hessian, monotonicity of the gradient, and the linear lower bound formula in Lecture 2, respectively.