Exercise 1: Risk Minimization and Gradient Descent

You want to estimate the relationship between a continuous response variable $\boldsymbol{y} \in \mathbb{R}^n$ and some feature $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ using the linear model with an appropriate loss function L.

(a) Describe the model f used in this case, its hypothesis space $\mathcal{H}$ and the theoretical risk function.

(b) Given $f \in \mathcal{H}$, explain the different parts of the Bayes regret if (i) $f^* \in \mathcal{H}$; if (ii) $f^* \notin \mathcal{H}$.

(c) Define the empirical risk and derive the gradients of the empirical risk.

(d) Show that the empirical risk is convex in the model coefficients. Why is convexity a desirable property? Hint: Compute the Hessian matrix $\boldsymbol{H} \in \mathbb{R}^{p \times p}$ and show that $\boldsymbol{z}^\top \boldsymbol{H} \boldsymbol{z} \geq 0 \, \forall \boldsymbol{z} \in \mathbb{R}^p$, i.e., show that the Hessian is positive semi-definite (psd).