The following homework will be discussed on June 29th to July 1st:

I. Rademacher complexity with margin

1. Let $\mathcal{H}$ be a real Hilbert space, $x := \{x_1, \ldots, x_n\} \subseteq \mathcal{H}$, $G_{ij} := \langle x_i, x_j \rangle$, $\rho \in \mathbb{R}_+$ and $\mathcal{G} := \{\mathcal{H} \ni z \mapsto \langle w, z \rangle \mid w \in \mathcal{H} \wedge ||w||^{-1} \geq \rho\}$. Show that the empirical Rademacher complexity of $\mathcal{G}$ w.r.t x can be bounded by

$$\hat{\mathfrak{R}}(\mathcal{G}) \leq \frac{\operatorname{tr}[G]^{1/2}}{n\rho}. \tag{1}$$

   Hint: you may use (in that order) Cauchy-Schwarz, Jensen's inequality and the fact that if $i \neq j$, then $\mathbb{E}_\sigma[\sigma_i \sigma_j] = 0$.

2. Use 1. to derive a PAC bound for support vector machine learning. Specify the assumptions that you need.

II. KKT and support vectors

Apply the convex KKT conditions (see lecture notes) to the optimization problem

$$\begin{aligned} &\min_{(b,w,\xi)} \frac{\lambda}{2}||w||^2 + \frac{1}{n}\sum_{i=1}^{n} \xi_i \\ &\text{s.t.} \quad y_i\big(\langle w, x_i \rangle + b\big) \geq 1 - \xi_i \quad \wedge \quad \xi_i \geq 0 \quad \forall i = 1, \ldots n, \end{aligned} \tag{2}$$

where $\lambda > 0$ is a free parameter and the minimum is taken over all $w \in \mathcal{H}, b \in \mathbb{R}$ and $\xi \in \mathbb{R}^n_+$. What can you say about the optimal w? Can you identify two types of *support vectors*?