The following homework will be discussed on May 18th to 20th:

I. Concentration inequalities

1. Show that Hoeffding's inequality follows from McDiarmid's inequality.
2. What do you think is the conceptual benefit in using McDiarmid's rather than Hoeffding's inequality in PAC learning bounds. (Compare the main ingredients in the proofs and the restrictions in the assumptions of the theorems.)

II. Rademacher complexities

1. Consider a hypotheses class $\mathcal{F} \subseteq \{-1,1\}^{\mathcal{X}}$, $L(y,y') := \mathbb{1}_{y \neq y'}$ as loss function and $\mathcal{G} := \{(x,y) \mapsto L(y,h(x)) \mid h \in \mathcal{F}\}$. Denote the restriction of $S = \big((x_i,y_i)\big)_{i=1}^n \in (\mathcal{X} \times \{-1,1\})^n$ to $\mathcal{X}$ by $S_{\mathcal{X}} := (x_i)_{i=1}^n$. Show that for any probability measure P on $\mathcal{X} \times \{-1,1\}$ with marginal p on $\mathcal{X}$ we have for the (empirical) Rademacher complexities:

$$\hat{\mathcal{R}}_S(\mathcal{G}) = \frac{1}{2}\hat{\mathcal{R}}_{S_{\mathcal{X}}}(\mathcal{F}) \quad \text{and} \quad \mathcal{R}_{n,P}(\mathcal{G}) = \frac{1}{2}\mathcal{R}_{n,p}(\mathcal{F}). \tag{1}$$

   Here the additional indices p, P and $S, S_{\mathcal{X}}$ label the probability measures and sets w.r.t. which the (empirical) Rademacher complexities are taken.

2. Let $\mathcal{G}, \mathcal{G}_1, \mathcal{G}_2 \subseteq \mathbb{R}^{\mathcal{Z}}$ be classes of real-valued functions on $\mathcal{Z}$ and $z \in \mathcal{Z}^n$. Show that the following holds for the empirical Rademacher complexities w.r.t. z:

   (a) If $c \in \mathbb{R}$, then $\hat{\mathcal{R}}(c\mathcal{G}) = |c|\hat{\mathcal{R}}(\mathcal{G})$.

   (b) $\mathcal{G}_1 \subseteq \mathcal{G}_2$ implies $\hat{\mathcal{R}}(\mathcal{G}_1) \leq \hat{\mathcal{R}}(\mathcal{G}_2)$.

   (c) $\hat{\mathcal{R}}(\mathcal{G}_1 + \mathcal{G}_2) = \hat{\mathcal{R}}(\mathcal{G}_1) + \hat{\mathcal{R}}(\mathcal{G}_2)$.

   (d) $\hat{\mathcal{R}}(\mathcal{G}) = \hat{\mathcal{R}}(\text{conv } \mathcal{G})$, where conv denotes the convex hull.

   (e) If $\varphi : \mathbb{R} \to \mathbb{R}$ is $L-$Lipschitz, then $\hat{\mathcal{R}}(\varphi \circ \mathcal{G}) \leq L\, \hat{\mathcal{R}}(\mathcal{G})$.

   Note: (a), (b) and (c) are simple, (d) is a little bit harder and (e) is hard. You should at least try (a)-(d).