1. **PAC learnability of right subintervals**
   Consider the learning problem for the triple \((X, \mathcal{P}, \mathcal{C})\) where \(X = [0, 1]\), \(\mathcal{C} = \{[\tau, 1] : 0 \leq \tau \leq 1\}\), and \(\mathcal{P}\) is the set of all probability distributions on \(X\). Describe an ERM classifier \(C_n\) and show that it PAC learns \(\mathcal{C}\) in the realizable case to accuracy \(\epsilon\) with confidence at least \(1 - \delta\), if the sample size is at least \(n(\epsilon, \delta) = \left\lceil \frac{\log(1/\delta)}{\epsilon^2} \right\rceil\). The training set is given by \(Z^n = (Z_1, \ldots, Z_n)\), where \(Z_i = (X_i, Y_i) = (X_i, 1_{(X_i \in C \cap \tau, 1]} \) and \(C^* = [\tau^*, 1]\) is the target interval.

2. **The minimum of a uniform approximation is an approximate minimizer**
   Suppose we’d like to find a minimizer of a function \(G\) defined on some domain \(\mathcal{U}\), but the function \(G\) is not known. Suppose that \(\hat{G}\) is an \(\epsilon\) uniform approximation of \(G\) for some \(\epsilon > 0\), meaning that \(|G(u) - \hat{G}(u)| \leq \epsilon\) for all \(u \in \mathcal{U}\). Suppose that \(u^*\) is a minimizer of \(\hat{G}\), meaning that \(u^* \in \mathcal{U}\) and \(\hat{G}(u^*) \leq \hat{G}(u)\) for all \(u \in \mathcal{U}\). Prove that \(G(u^*) \leq \inf_{u \in \mathcal{U}} G(u) + 2\epsilon\).

3. **Subgaussian random variables**
   A random variable \(X\) is said to be subgaussian with scale parameter \(\nu\), if \(X\) has a finite mean and \(\mathbb{E} [e^{s(X - \mathbb{E}[X])}] \leq e^{\frac{s^2 \nu^2}{2}}\) for all \(s \in \mathbb{R}\). Sometimes \(\nu^2\) is called the proxy variance because a Gaussian random variable with variance \(\sigma^2\) is subgaussian for \(\nu^2 = \sigma^2\).
   
   (a) Suppose \(U\) is a random variable such that for some parameters \(a, b\), \(\mathbb{P}\{U \in [a, b]\} = 1\). What is the smallest value of \(\nu\), depending only on \(a\) and \(b\), for which it follows that \(U\) is subgaussian with scale parameter \(\nu^2\)? (Hint: Consider the Hoeffding inequality.)
   
   (b) Suppose \(S_n = X_1 + \cdots + X_n\), where \(X_1, \ldots, X_n\) are independent random variables, such that \(X_i\) is subgaussian with scale parameter \(\nu_i\). Show that \(S_n\) is subgaussian with proxy variance given by \(\nu^2 = \nu_1^2 + \cdots + \nu_n^2\). Does this fact continue to hold if the \(X\)’s are not independent? Justify your answer with either a proof or counter example.
   
   (c) Using the methodology of the Chernoff inequality, show that if \(X\) is subgaussian with scale parameter \(\nu\), then \(\mathbb{P}\{X - \mathbb{E}[X] \geq nt\} \leq e^{-\frac{t^2 \nu}{2}}\) for all \(t \geq 0\).
   
   (d) Suppose \(X_1, \ldots, X_n\) are mean zero, and each is subgaussian with scale parameter \(\nu\). Show that \(\mathbb{E}[\max_i X_i] \leq \nu \sqrt{2 \log n}\). Does this bound require the \(X\)’s to be independent? (Hint: Use Jensen’s inequality for \(e^x\), and the inequality \(e^{sx}X_i \leq \sum_i e^{sX_i}\) for \(s \geq 0\).)

4. **Bounded differences vs. Lipschitz continuity**
   Fix \(p > 0\) and define \(f : [0, 1]^n \rightarrow \mathbb{R}_+\) by \(f(x) = \|x\|_p = (\sum_{i=1}^n |x_i|^p)^{1/p}\). If \(p \geq 1\) then \(\|x\|_p\) is a norm called the \(\ell^p\) norm. Suppose \(n \geq 2\).
   
   (a) Find the smallest constant \(c > 0\) so that \(f\) has the bounded differences property (as appearing in McDiarmid’s inequality). Your answer may depend on \(n\) and \(p\). Check your answers for \(p = 1\) and \(p = 2\).
   
   (b) Find the smallest constant \(L > 0\) such that \(f\) is \(L\)-Lipschitz continuous (meaning \(|f(x) - f(y)| \leq L\|x - y\|_2\)). Your answers may depend on \(n\) and \(p\). Hint: For continuously differentiable functions on convex sets, the smallest \(L\) is the max of \(\|\nabla f\|_2\) over the domain of \(f\). If no finite \(L\) works, then \(f\) is not Lipschitz continuous.

5. **Problem 2.1 from the book**