Problem 1

Let $\mathsf{URad}(\cdot)$ denote the unnormalized Rademacher complexity. Show that for $n \in \mathbb{N}$, there exist sets S_n, S'_n containing vectors in $\mathbb{R}^n$ such that $|S_n| = |S'_n| = 2$ and $\mathsf{URad}(S_n) = \Omega(n^{1/2})$ while $\mathsf{URad}(S'_n) = 1$.

Problem 2

Part (a)

Consider the binary Gaussian mixture model with label flipping noise $p \in (0, 1/2)$: for some $\mu \in \mathbb{R}^d$, $\tilde{y} \sim \mathsf{Unif}(\{\pm 1\})$ and $x|\tilde{y} \sim \tilde{y}\mu + z$ for $z \sim \mathsf{N}(0, I_d)$, then $y = \tilde{y}$ w.p. $1 - p$, and $y = -\tilde{y}$ w.p. p. Let $S = \{(x_i, y_i)\}_{i=1}^n$ where (x_i, y_i) are i.i.d. from this distribution.

Let $\lambda_i > 0$ be a sequence of strictly positive numbers, and consider the estimator

$$v := \sum_{i=1}^{n} \lambda_i y_i x_i.$$

Let $\Lambda = \max_{i,j} \lambda_i/\lambda_j$. Let $\delta \in (0, 1/2)$. Assume that $\|\mu\| = d^\beta$ for $\beta \in (1/4, 1/2)$, and that $d = \omega(n^{2 \vee \frac{1}{1-2\beta}} \log^2(n/\delta))$.

Show that if Λ is an absolute constant (independent of n and d), and if p is smaller than some absolute constant (depending on Λ), then with probability at least $1 - \delta$, v exhibits *benign overfitting* in the sense that,

1. v perfectly fits the training data: for all $k \in [n]$, $y_k \langle v, x_k \rangle > 0$, and
2. v achieves near-optimal prediction error:

$$p \le \mathbb{P}(y \neq \operatorname{sign}(\langle v, x \rangle)) \le p + o_d(1).$$

Hint: The lecture notes provide a sketch of a proof when $\lambda_i = 1$ for all i.

Part (b)

Refer back to Homework 1 and the lecture notes. What does this tell us about what happens if we train by gradient flow on the logistic loss, initialized from the origin, using training data S? What if the loss is the squared loss instead?