# Homework 3
# APPM 7400 Spr 2020 Theoretical ML

**Due date**: Friday, Feb 7, before 1 PM
**Instructor**: Prof. Becker
**Theme**: Rademacher complexity

**Instructions** Collaboration with your fellow students is OK and in fact recommended, although direct copying is not allowed. The internet is allowed for basic tasks (e.g., looking up definitions on wikipedia) but it is not permissible to search for proofs or to *post* requests for help on forums such as `http://math.stackexchange.com/` or to look at solution manuals. Please write down the names of the students that you worked with.

An arbitrary subset of these questions will be graded.

**Reading** You are responsible for reading chapter 3.1 (about Rademacher complexity) in Foundations of Machine Learning, 2nd edition, by Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar (MIT Press, 2018, ISBN-13: 978-0262039406, $50 on Amazon). The authors host a free PDF of the book at their website.

**Problem 1:** Let $\mathcal{H} = \{\mathbf{x} \mapsto \langle \mathbf{w}, \mathbf{x} \rangle \mid \|x\|_2 \leq 1\}$ be a set of linear classifiers. Let $S_x = (\mathbf{x}_1, \ldots, \mathbf{x}_m) \subset \mathbb{R}^n$ be a collection of vectors, and define the Frobenius norm of this set as $\|S_x\|_F^2 = \sum_{i=1}^m \|\mathbf{x}_i\|_2^2$.

a) Show the (empirical) Rademacher complexity is bounded as follows:

$$\widehat{\mathfrak{R}}(\mathcal{H} \circ S) \leq \frac{1}{m} \mathbb{E}_\sigma \left\| \sum_{i=1}^m \sigma_i \mathbf{x}_i \right\|_2$$

b) Simplify the bound to the following:

$$\widehat{\mathfrak{R}}(\mathcal{H} \circ S) \leq \frac{1}{m} \|S_x\|_F$$

*(Hint: use Jensen's inequality)*

c) Suppose $\mathcal{D}$ is the multivariate normal distribution $\mathcal{N}(0, I_{n \times n})$ on $\mathbb{R}^n$. Compute the (expected) Rademacher complexity $\mathfrak{R}_m(\mathcal{H}) = \mathbb{E}_{S \sim \mathcal{D}^m} \widehat{\mathfrak{R}}(\mathcal{H} \circ S)$.

Note: It is not straightforward to compute $\widehat{\mathfrak{R}}(\ell \circ \mathcal{H} \circ S)$ where $\ell$ is the $0-1$ loss function composed with a thresholding function, since $\ell$ is not Lipschitz continuous.