# MSBD 5004 Mathematical Methods for Data Analysis
# Homework 1

Due date: 27 February, 9pm, Wednesday

1. Find a norm other than $|\cdot|$ (the absolute value function) on $\mathbb{R}$ as a vector space over $\mathbb{R}$.
2. Consider the vector space $\mathbb{R}^n$.
   (a) Check that $\|\boldsymbol{x}\|_2 = \left(\sum_{i=1}^n |x_i|^2\right)^{1/2}$ is indeed a norm on $\mathbb{R}^n$.
   (b) Prove the equivalence
   $$\|\boldsymbol{x}\|_\infty \leq \|\boldsymbol{x}\|_1 \leq n\|\boldsymbol{x}\|_\infty, \quad \forall \boldsymbol{x} \in \mathbb{R}^n.$$
3. Let $(V, \|\cdot\|)$ be a normed vector space.
   (a) Prove that, for all $\boldsymbol{x}, \boldsymbol{y} \in V$,
   $$|\|\boldsymbol{x}\| - \|\boldsymbol{y}\|| \leq \|\boldsymbol{x} - \boldsymbol{y}\|.$$
   (b) Let $\{\boldsymbol{x}_k\}_{k\in\mathbb{N}}$ be a convergent sequence in $V$ with limit $\boldsymbol{x} \in V$. Prove that
   $$\lim_{k\to\infty} \|\boldsymbol{x}_k\| = \|\boldsymbol{x}\|.$$
   (*Hint: Use part (a).*)
4. Suppose that the vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N$ in $\mathbb{R}^n$ are clustered using the $k$-means algorithm, with group representatives $\boldsymbol{z}_1, \ldots, \boldsymbol{z}_k$.
   (a) Suppose the original vectors $\boldsymbol{x}_i$ are nonnegative, i.e., their entries are nonnegative. Explain why the representatives $\boldsymbol{z}_j$ output by the $k$-means algorithm are also nonnegative.
   (b) Suppose the original vectors $\boldsymbol{x}_i$ represent proportions, i.e., their entries are nonnegative and sum to one. (This is the case when $\boldsymbol{x}_i$ are word count histograms, for example.) Explain why the representatives $\boldsymbol{z}_j$ output by the $k$-means algorithm also represent proportions, i.e., their entries are nonnegative and sum to one.
   (c) Suppose the original vectors $\boldsymbol{x}_i$ are Boolean, i.e., their entries are either 0 or 1. Give an interpretation of $(\boldsymbol{z}_j)_i$, the $i$-th entry of the $j$ group representative.