COGSCI 131 – Assignment 2
DUE: July 8 at class start

In this problem set, you will implement multidimensional scaling (MDS) from scratch. You may use standard matrix/vector libraries (e.g. numpy) but you must implement two dimensional MDS itself on your own and not use an existing software package.

Recall that MDS starts with a similarity matrix and attempts to find an arrangement of points such that the distances match the observed similarities. In our implementation, we will try to minimize a quantity called the *stress*, which is the squared difference between the *distances* of items in psychological space and in MDS space.

Stress S = $\sum_{i \neq j} (d_{ij} - dist(x_i, x_j))^2$

where d_{ij} is the psychological distance between stimulus i and j that was reported by subjects and x_i and x_j are the positions of stimulus i and j in MDS space (here, $\mathbb{R}^2$). Thus, x_i and x_j are vectors. To minimize the stress,we will write code that will numerically compute the gradient using a multidimensional version of the simple rule for derivatives,

$$\frac{df}{dx}(p) \approx \frac{f(p+\delta) - f(p-\delta)}{2\delta}$$

where δ is small. We'll be using a version of this rule to compute the derivative of stress with respect to each coordinate of each x_i.

Please limit your responses to any short answers below to just a few sentences.

1. [5pts] The dataset provides similarities, not distances. Write down three ways you could convert a similarity to a distance d_{ij} and choose one to use in the code. Explain why you chose it over the others.

2. [20pts] Write the code that follows a gradient in order to find positions that minimize the stress. You may use the template provided if you want, but you do not need to. Plot the sport names at the resulting coordinates. Do the results agree with your intuitions about how this domain might be organized? Why or why not?

3. [5pts] In your run, make a scatter plot of the the pairwise distances MDS found vs. people's reported distances. Describe whether it looks good or bad.

3. [5pts] Plot the stress over iterations of your MDS. How should you use this plot in order to figure out how many iterations are needed?

4. [10pts] Run the MDS code you wrote 5 times and show small plots, starting from random initial positions. Are they all the same or not? Why?

5. [5pts] If you wanted to find one "best" answer but had run MDS 5 times, how would you pick the best? Why? Show a plot of the best and any code you used to find it.