Homework 2

1. Check whether the following norm L is legitimate:

   $L(v) = 0$, if $v=0$

   $L(v) = 1$, if $v \neq 0$

2. Find the nearest vector for each vector from the following group of vectors:

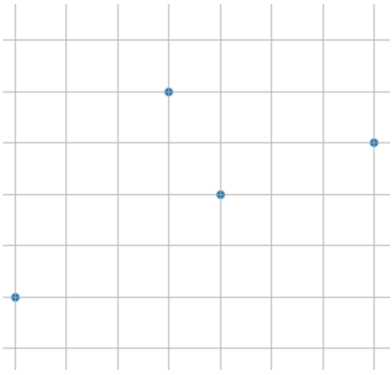
   $a = (0, -1, 1)$, $b = (0, 4, 0)$, $c = (2, 1, 4)$

   Measure distances using 3 metrics: L_1, L_2, L_∞.

   $||x||_\infty = \max_j(|x_j|)$

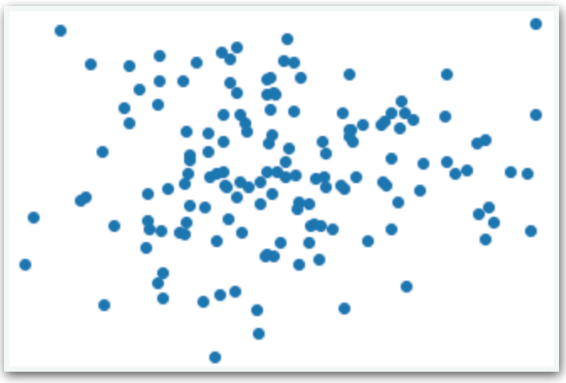
3. For x from $x_1=0$ to $x_2=8$ plot the function y using KNN. Take k=2 and the dataset of pairs (x, y).

   data = [(0, 1), (3, 5), (4, 3), (7, 4)]

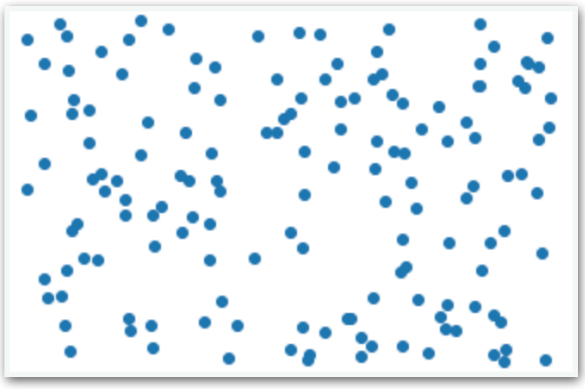


4. What would you use normalization or standardization for the datasets below. Why?

   A



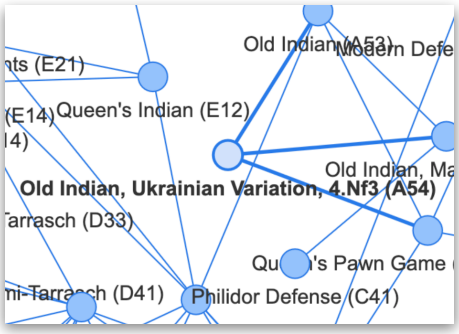
   B



5. **Optional**:

   Find clusters in the dataset contained in clusters.csv file. You can use k-means or any other available clustering methods.

6. **Optional**:

   Encyclopedia of Chess Openings includes about one thousand codes (for details see https://en.wikipedia.org/wiki/Encyclopaedia_of_Chess_Openings).

   In the dataset in chess_openings.csv, each opening is represented by a vector. Vectors are in the rows of the dataset. For your convenience there is another csv file (debuts.csv) that contains the names of the openings and their codes. For the opening **A54** - "Old Indian, Ukrainian Variation, 4.Nf3 (A54)", find 10 nearest vectors (similar openings) for both L1 and L2 distance metrics.