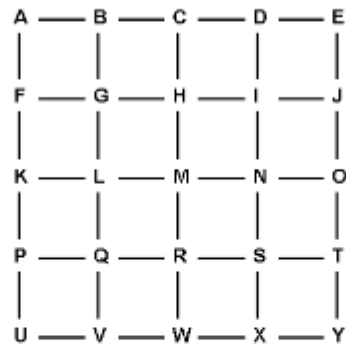


In the following diagram, A through Y are all data points. Each data point has features (x,y) corresponding to its coordinate in the grid.



- (3 points) What are the $k = 5$ nearest neighbors of data point M using Euclidean distance? Break any ties with alphabetical ordering.

G,H,L,N,R

- (3 points) If data points A through L belong to class 1 and data points N through Y belong to class 2, what is the classification of M when using the $k = 5$ nearest neighbors?

Class 1

- (4 points) Suppose if in addition to the (x,y) position features, we also had an additional “weight” w attached to each data point. Assume that the standard deviation of x , y , and w are as follows: $\sigma_x = 2$, $\sigma_y = 2$, $\sigma_w = 200$. Would Euclidean distance still be an appropriate distance metric? Briefly describe why or why not.

No. In order to use Euclidean distance, we would want to *standardize* each feature by its standard deviation before combining them. Otherwise, the “weight” feature would overwhelm the “x-position” and “y-position” features in the distance calculations.
