

HW02: Nearest neighbor classifier

Hand in via moodle at: <https://moodle.umass.edu/course/view.php?id=20836>. Remember that only PDF submissions are accepted. We encourage using L^AT_EX to produce your writeups. See `hw00.tex` for an example of how to do so. You can make a `.pdf` out of the `.tex` by running “`pdflatex hw00.tex`”. You’ll need `mydefs.sty` and `notes.sty` which can be downloaded from the course page.

1. Give an example of a low dimensional (approx. 20 dimensions), medium dimensional (approx. 1000 dimensions) and high dimensional (approx. 100000 dimensions) problem that you care about.
2. What does the decision boundary of 1 nearest neighbor classifier for 2 points (one positive, one negative) look like?
3. Clustering was introduced as a way to speed up k nearest neighbor (kNN) classification. Is it possible that clustering can lead to a better classifier? Briefly explain why.
4. Give two examples of data where the Euclidean distance is not the right metric.
5. Does the accuracy of a kNN classifier using the Euclidean distance change if you (a) translate the data (b) scale the data (i.e., multiply the all the points by a constant), or (c) rotate the data? Explain. Answer the same for a kNN classifier using Manhattan distance¹.

¹http://en.wikipedia.org/wiki/Taxicab_geometry