DS2500
3/18-Tues.

Agenda

1. Supervised Learning → Slides
2. KNN Classification
3. Python

2. KNN Classifier
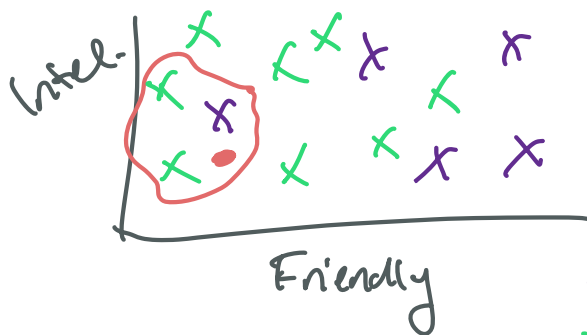
- just now: high level steps
- Next: pseudocode

{ What steps in code?
  What functions?
  What computations? }

training.csv   features, labels
testing.csv   features, (labels)

Intel- | (scatter plot)
       Friendly

x train dogs
x train cats
• test

- read in csv files
- train df, test df
- euclidean distance
- compute distance from a testing object
    to all training objects
        ↳ find k closest training objects
          from these, which label appears
          most often?

assign that label as predicted
for the test object

- do the same for _all_ test objects!
- plot/visualize!