# CSC411: *k*-Nearest Neighbors

January 29, 2016

# *k*-Nearest Neighbors

- *k*-NN is a simple algorithm which stores all available training examples and predict value/class of an unseen instance based on a similarity measure
  - ➢ *k* = 1
    - ▪ predict the same value/class as the nearest instance in the training set
  - ➢ *k* > 1
    - ▪ find the k closet training examples
    - ▪ *predict class:* majority vote
    - ▪ *predict value:* average weighted by inverse distance
- memory based, no explicit training or model

# *k*-NN Classification

- similarity measure: Euclidean distance, etc.
  - assumption behind Euclidean distance: uncorrelated inputs with equal variances
- predict class: majority vote
- *k* preferably odd to avoid ties for binary classification
- choice of *k:*
  - smaller *k*: higher variance (less stable)
  - larger *k*: higher bias (less precise)
  - cross-validation can help
- MATLAB demo