Project 5: The Hubness Phenomenon in High Dimensional Spaces

July 21, 2017

Project 5 Participants

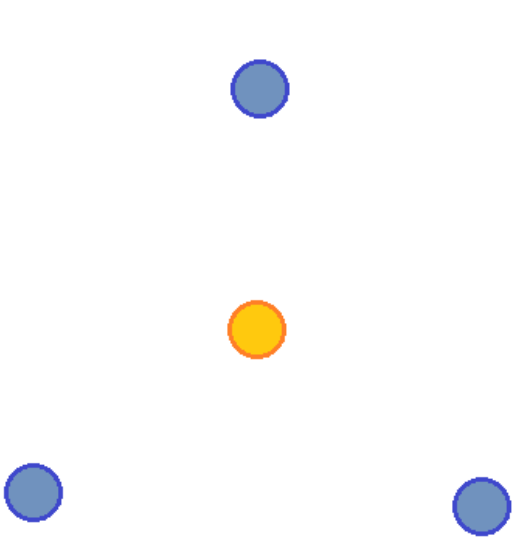
- Carlotta Domeniconi (George Mason University)
- Sibel Tari (Middle East Technical University)

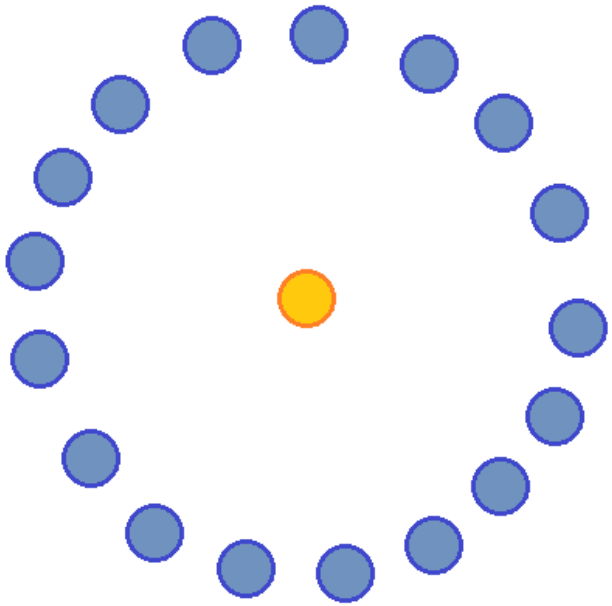
- Gülce Bal (Middle East Technical University)
- Libby Beer (Institute for Defense Analyses)
- Hillary Fairbanks (University of Colorado Boulder)
- Priya Mani (George Mason University)
- Jesse Metcalf-Burton (Department of Defense)
- Marilyn Vazquez (George Mason University)

Group 5 Overview

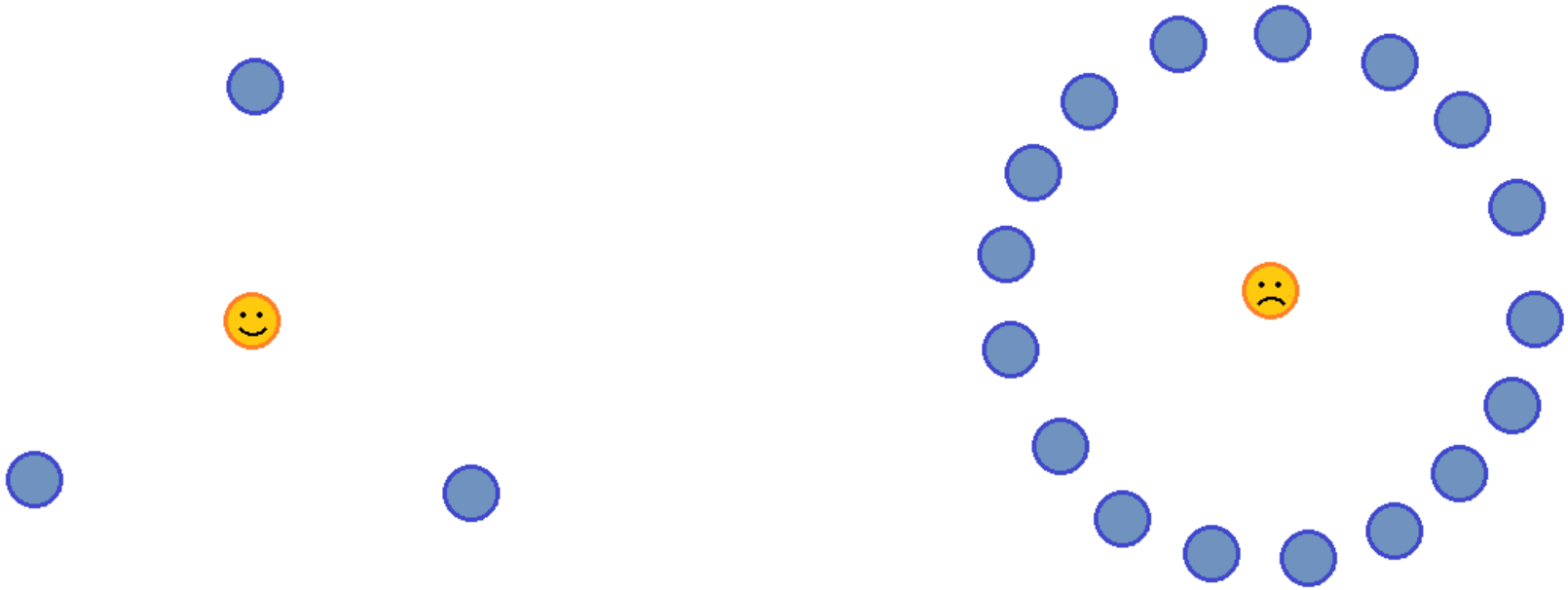
- Overview of concepts and datasets
- Question 1: Intrinsic dimensionality via hubness
  - Skewness vs. feature ranking - what is the right way to rank features?
  - Supervised vs. unsupervised methods
- Question 2: Hubs, density, clustering, and outliers
  - How does hubness relate to data density?
  - How are hubs distributed across clusters?
  - How do hubs relate to special points from various clustering methods (DBSCAN core points, outliers, etc.)?

Hubness - how many nodes have me as a near neighbor?

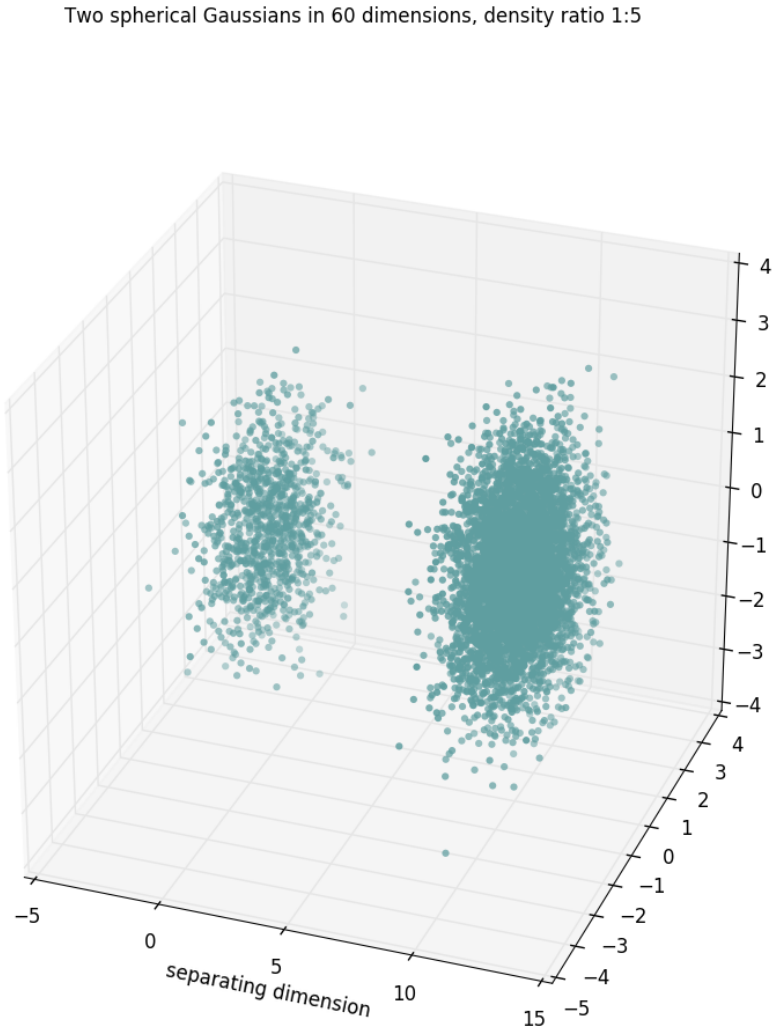




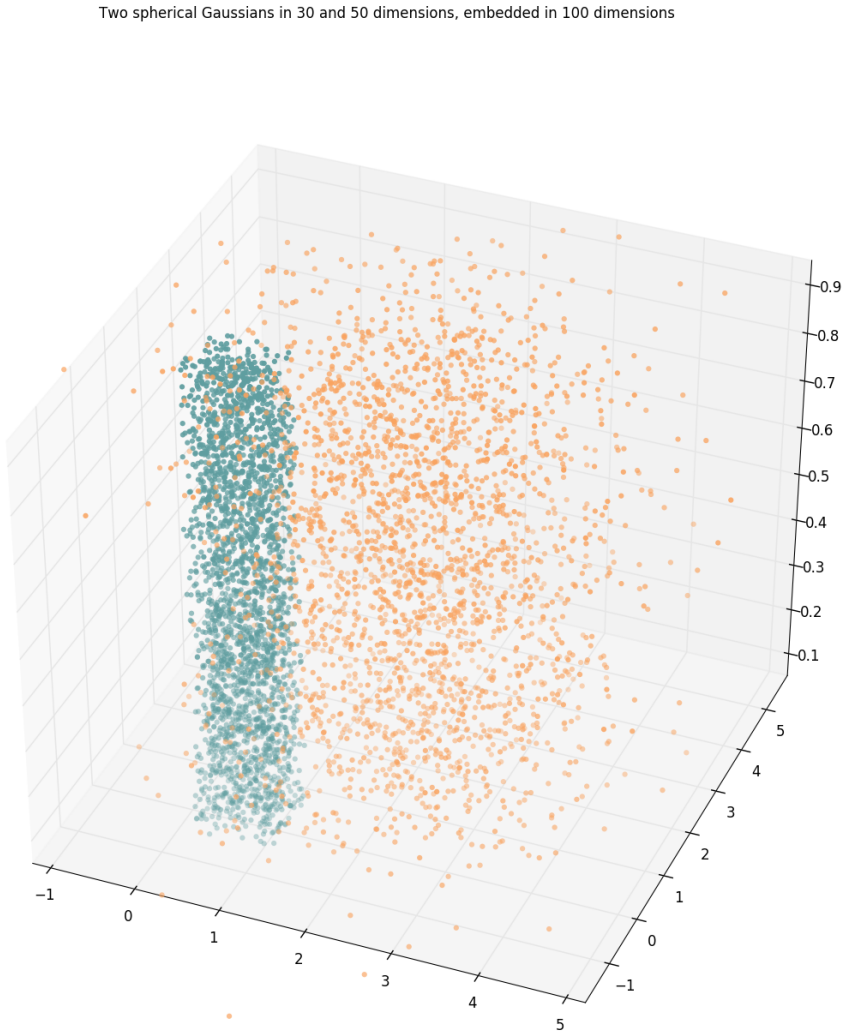
Hubness - how many nodes have me as a near neighbor?



Synthetic data – density difference



Synthetic data – dimensionality difference



Real-world data: spam classification