

Learning Analytics: A Hands-On Conceptual Introduction

Ruben R. Puentedura, Ph.D.

KNIME

The screenshot displays the KNIME software interface with a workflow diagram and the Entropy Scorer node description.

Workflow Diagram:

- Node 1:** File Reader
- Node 2:** Partitioning
- Node 3:** Logistic Regression (Learner)
- Node 4:** Logistic Regression (Predictor)
- Node 5:** Entropy Scorer

The workflow connects Node 1 to Node 2, Node 2 to Node 3, Node 2 to Node 4, and Node 3 to Node 4. Node 4 then connects to Node 5.

Node Description: Entropy Scorer

Scorer for clustering results given a reference clustering. Connect the table containing the reference clustering to the first input port (the table should contain a column with the cluster IDs) and the table with the clustering results to the second input port (it should also contain a column with some cluster IDs). Select the respective columns in both

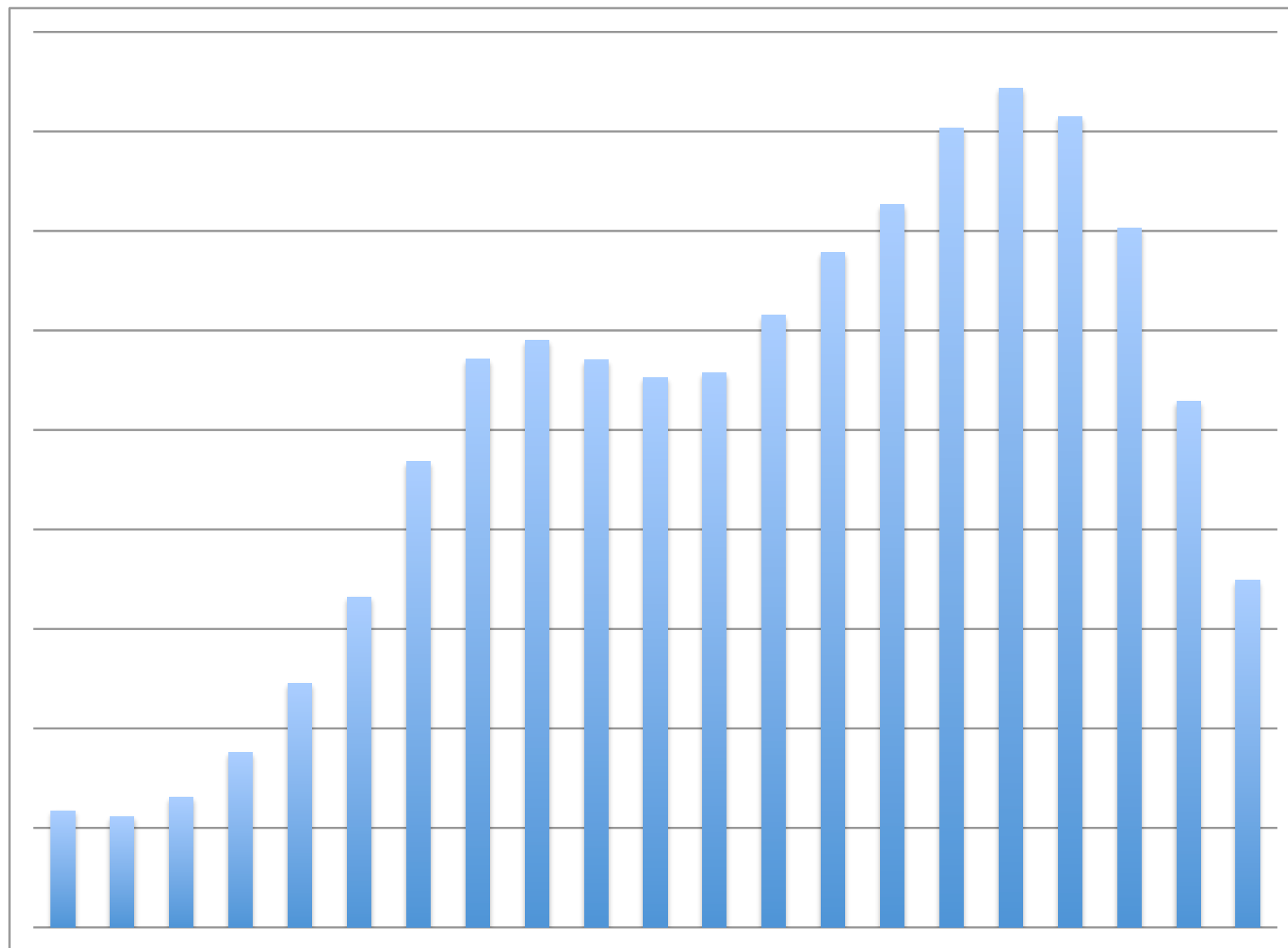
Node Repository:

- Mining
 - Bayes
 - Clustering
 - Rule Induction
 - Neural Network
 - Decision Tree
 - Misc Classifiers
 - Item Sets / Association Rules
 - MDS
 - PCA
 - SVM
 - Scoring
 - Enrichment Plotter
 - Entropy Scorer
 - ROC Curve
 - Scorer
- Chemistry
 - ChemAxon / Infocom
 - Distance Matrix
 - Meta
 - Misc
 - KNIME Labs
 - Time Series

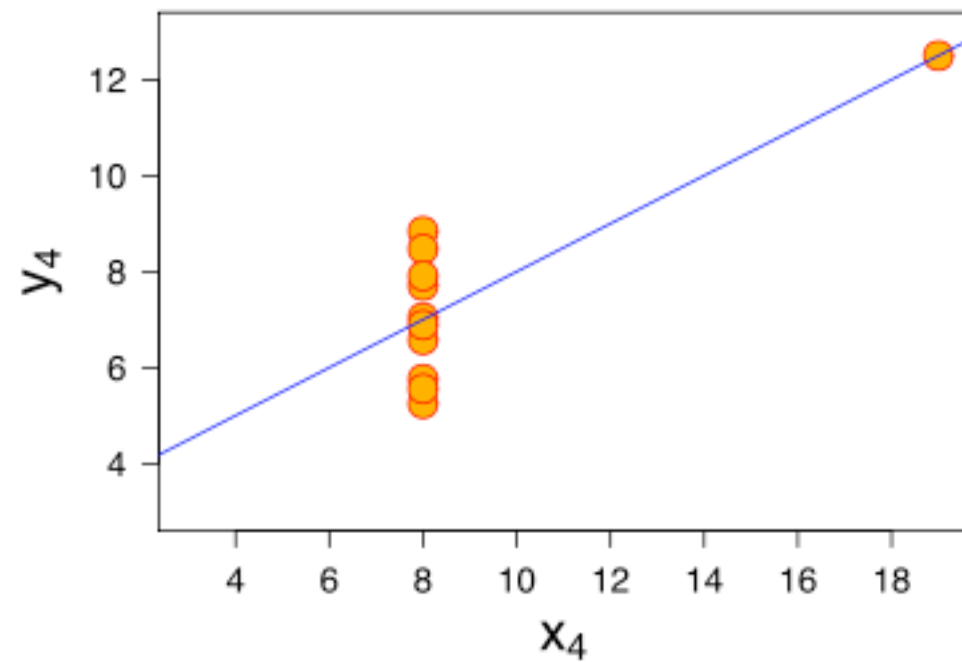
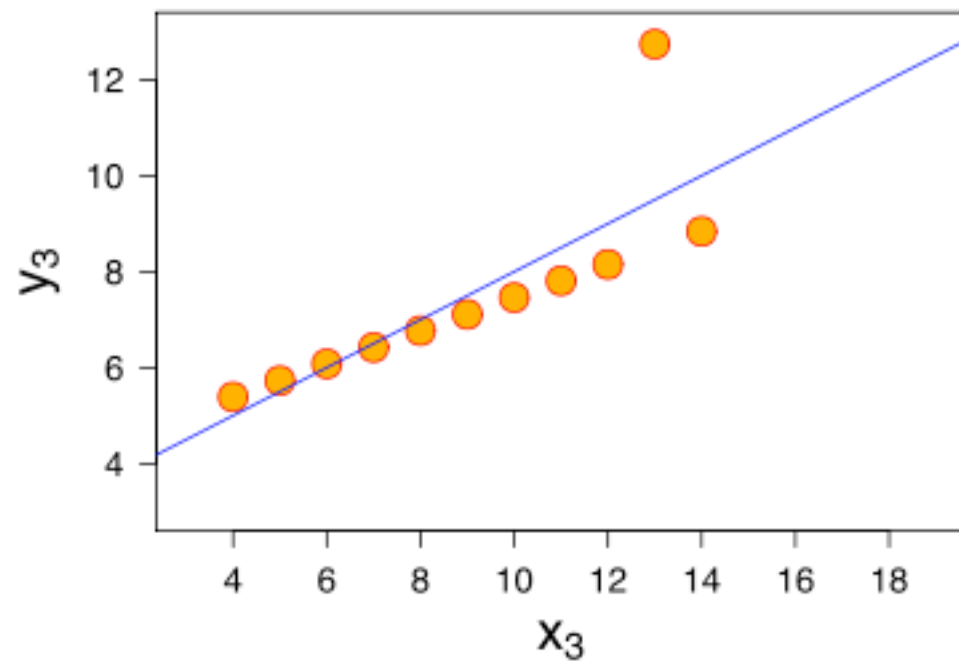
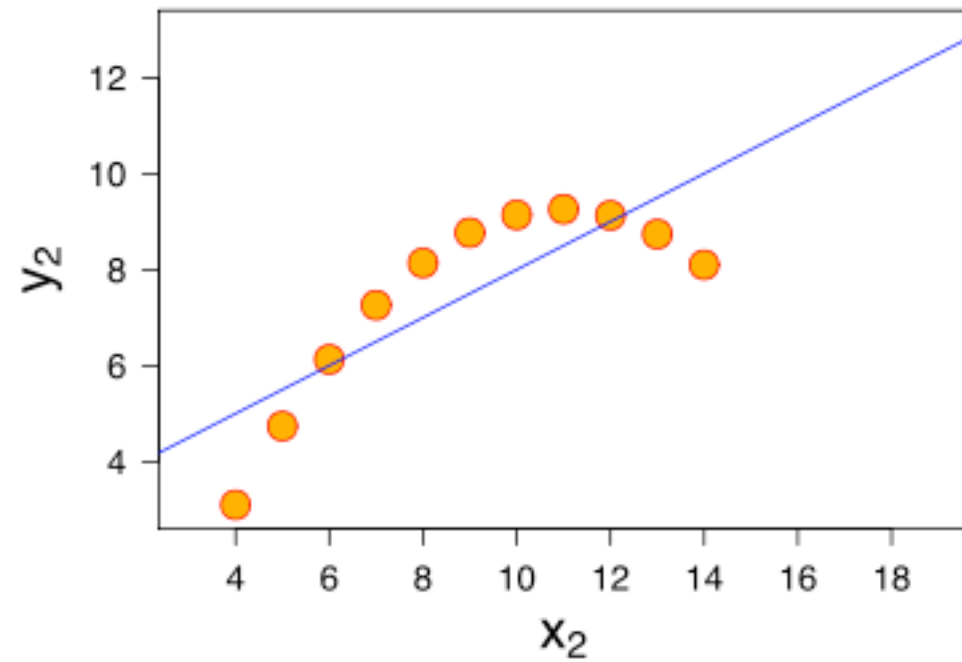
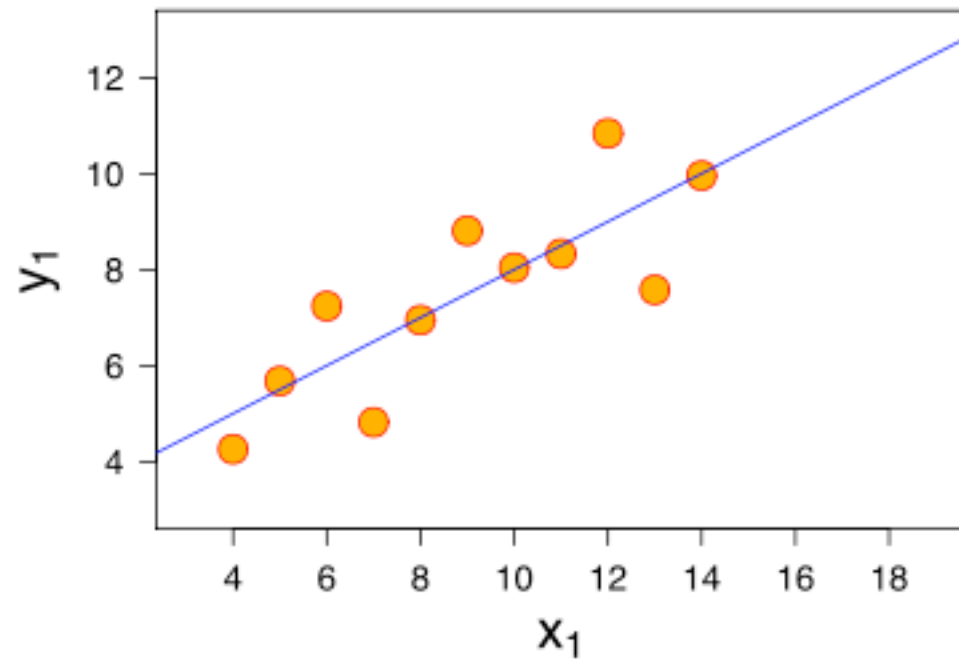
Console:

```
KNIME Console
Log file is located at: /Users/ruben/Documents/KnimeWorkspace/.metadata/knime/knime.log
ERROR NativeLibraryInstaller Unsupported system name / architecture: no OpenGL library for this system no
WARN File Reader No Settings available.
WARN Partitioning No sampling method selected
WARN Entropy Scorer No auto configuration available
Please configure in dialog.
WARN Entropy Scorer No auto configuration available
Please configure in dialog.
```

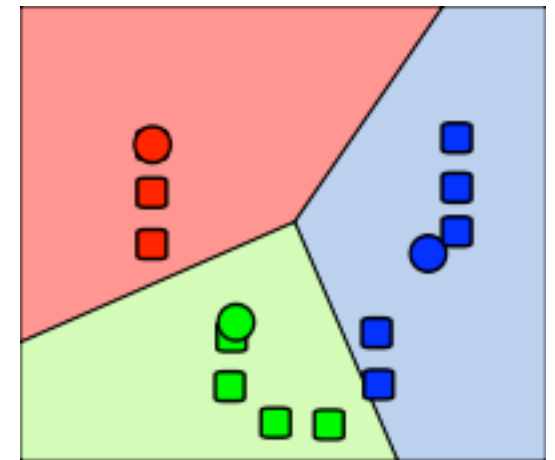
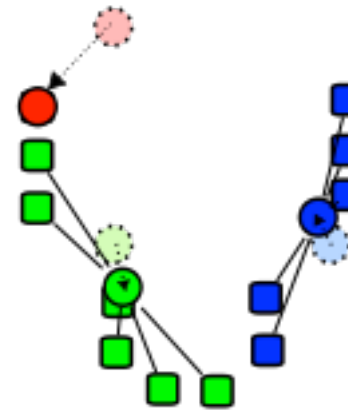
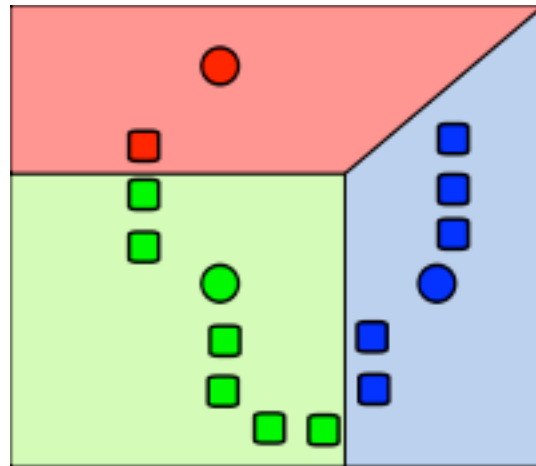
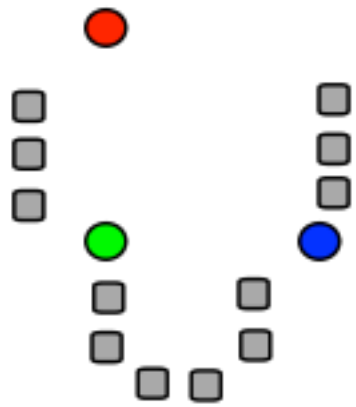
The Basics



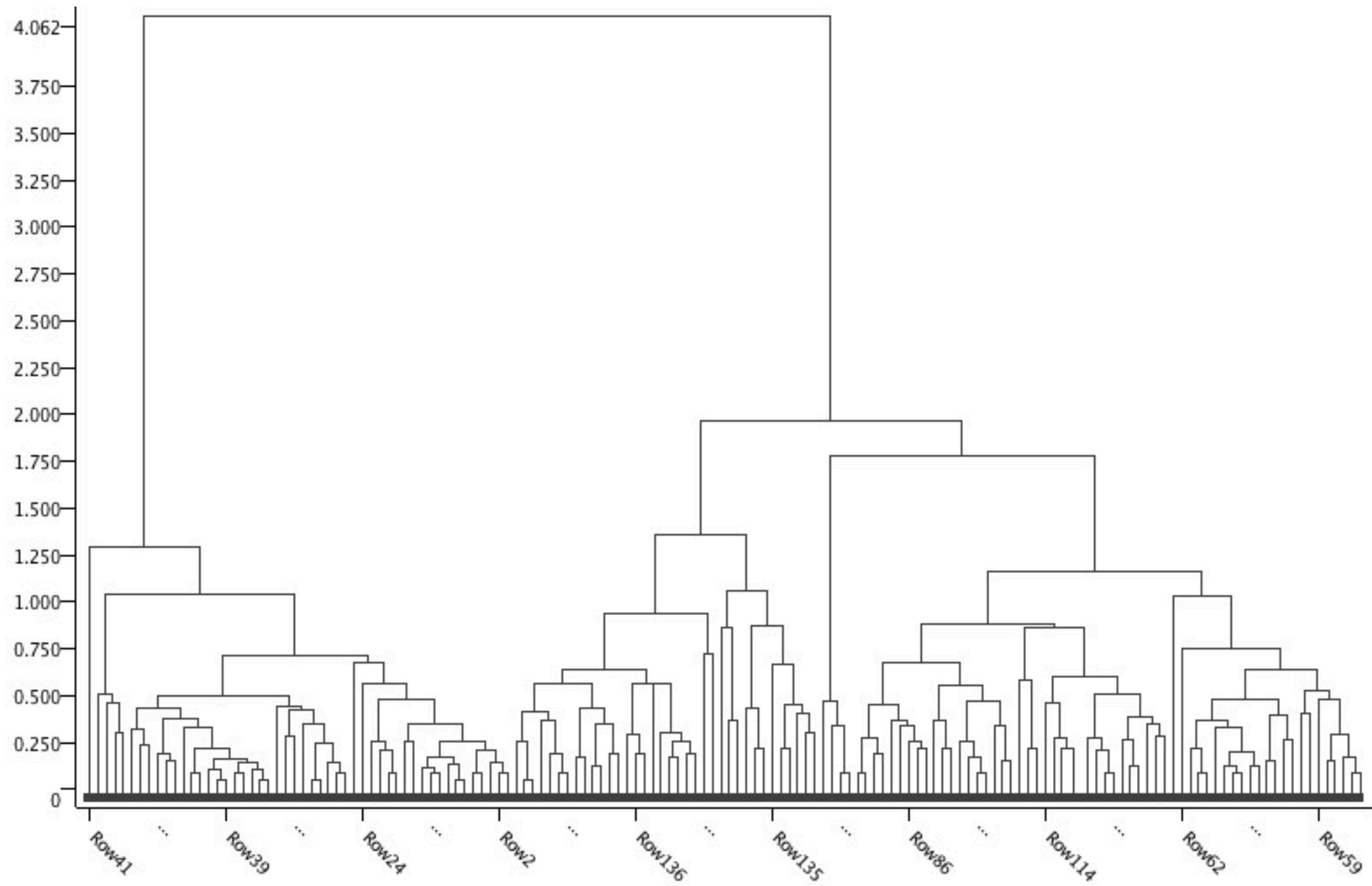
Anscombe's Quartet



Finding Groups: k-Means Clustering

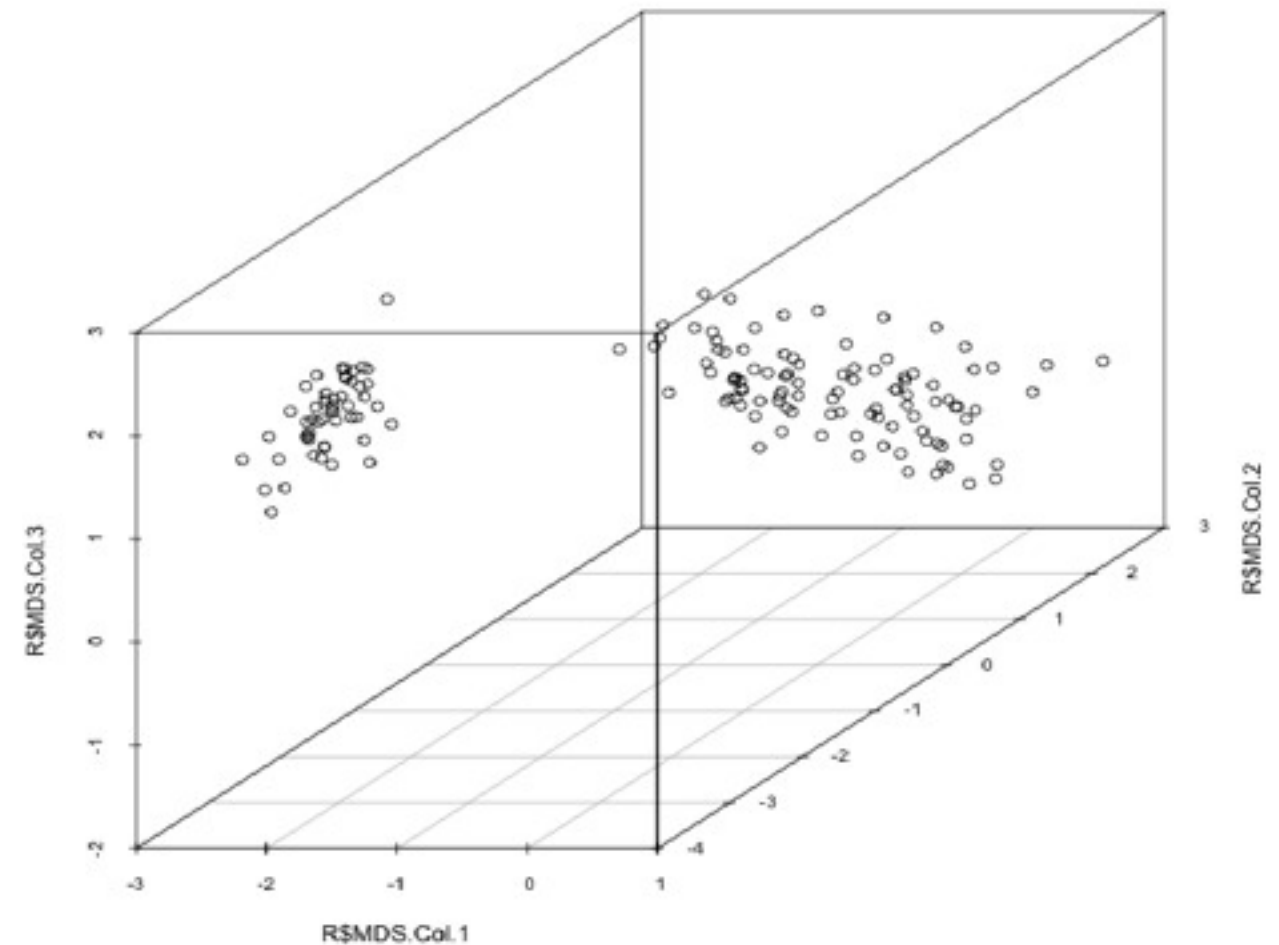


Finding Groups: Hierarchical Clustering

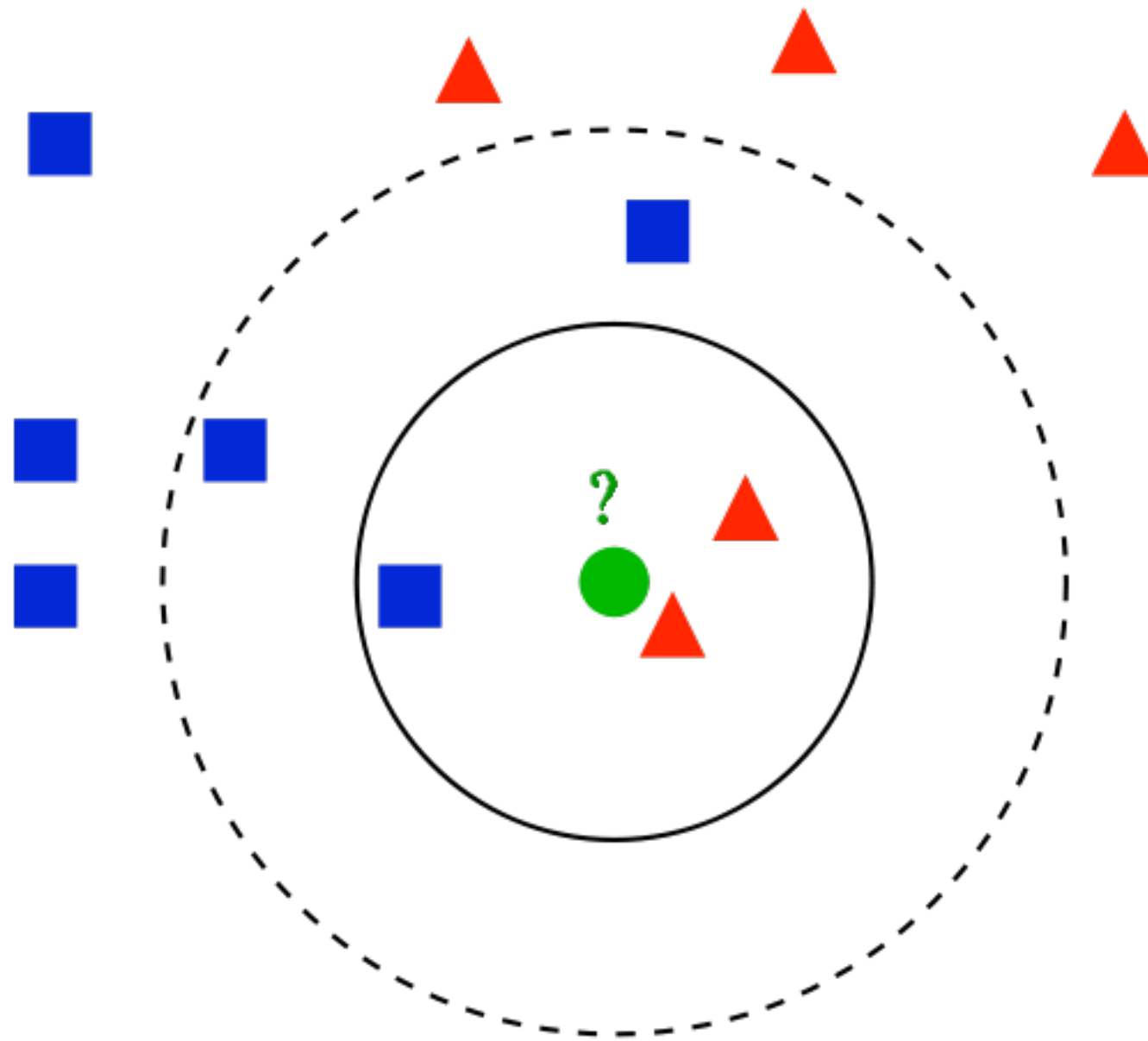


Finding Groups: Multidimensional Scaling

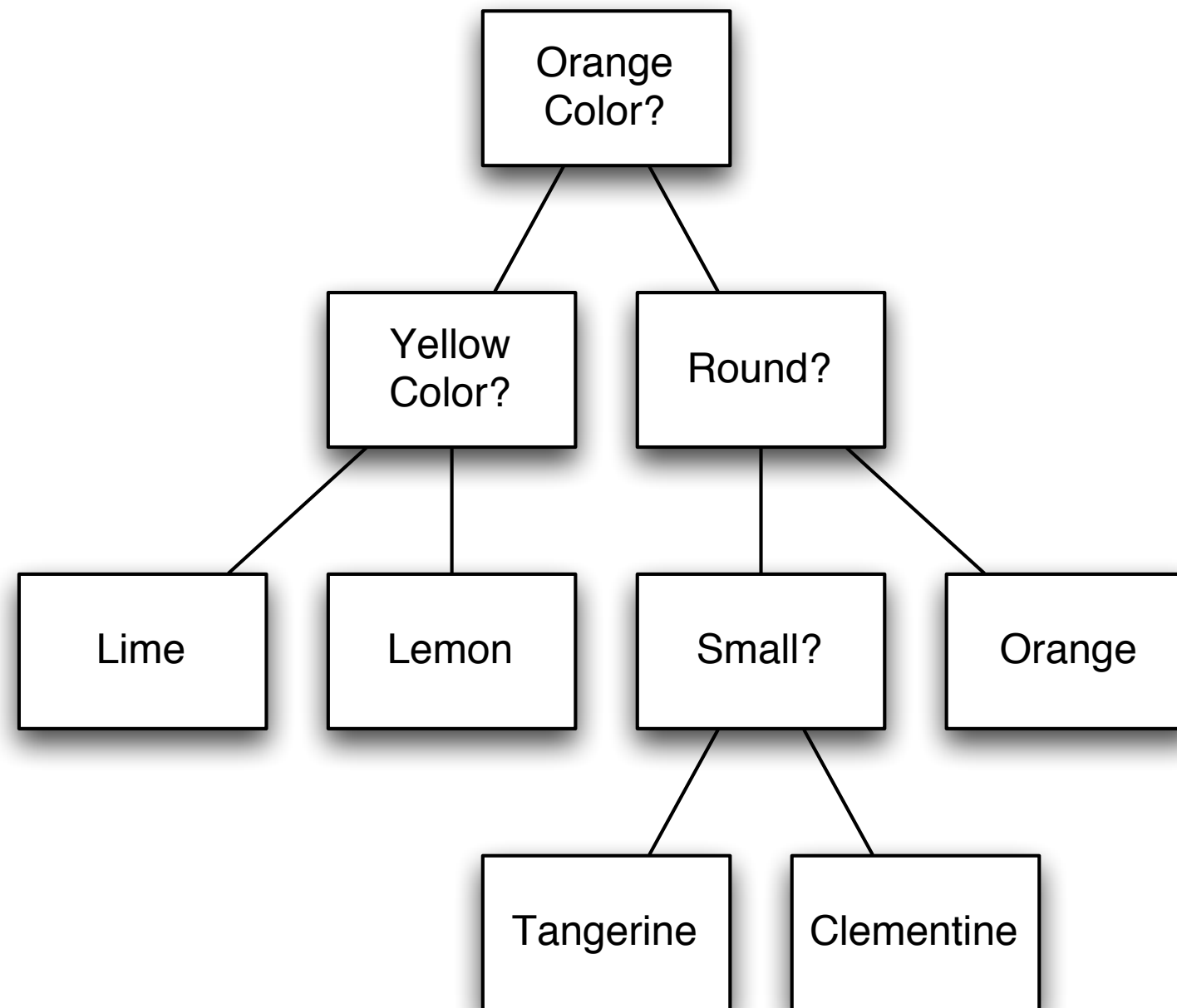
$$\Delta := \begin{pmatrix} \delta_{1,1} & \delta_{1,2} & \cdots & \delta_{1,I} \\ \delta_{2,1} & \delta_{2,2} & \cdots & \delta_{2,I} \\ \vdots & \vdots & & \vdots \\ \delta_{I,1} & \delta_{I,2} & \cdots & \delta_{I,I} \end{pmatrix}.$$



Classifying Members: k-Nearest Neighbors



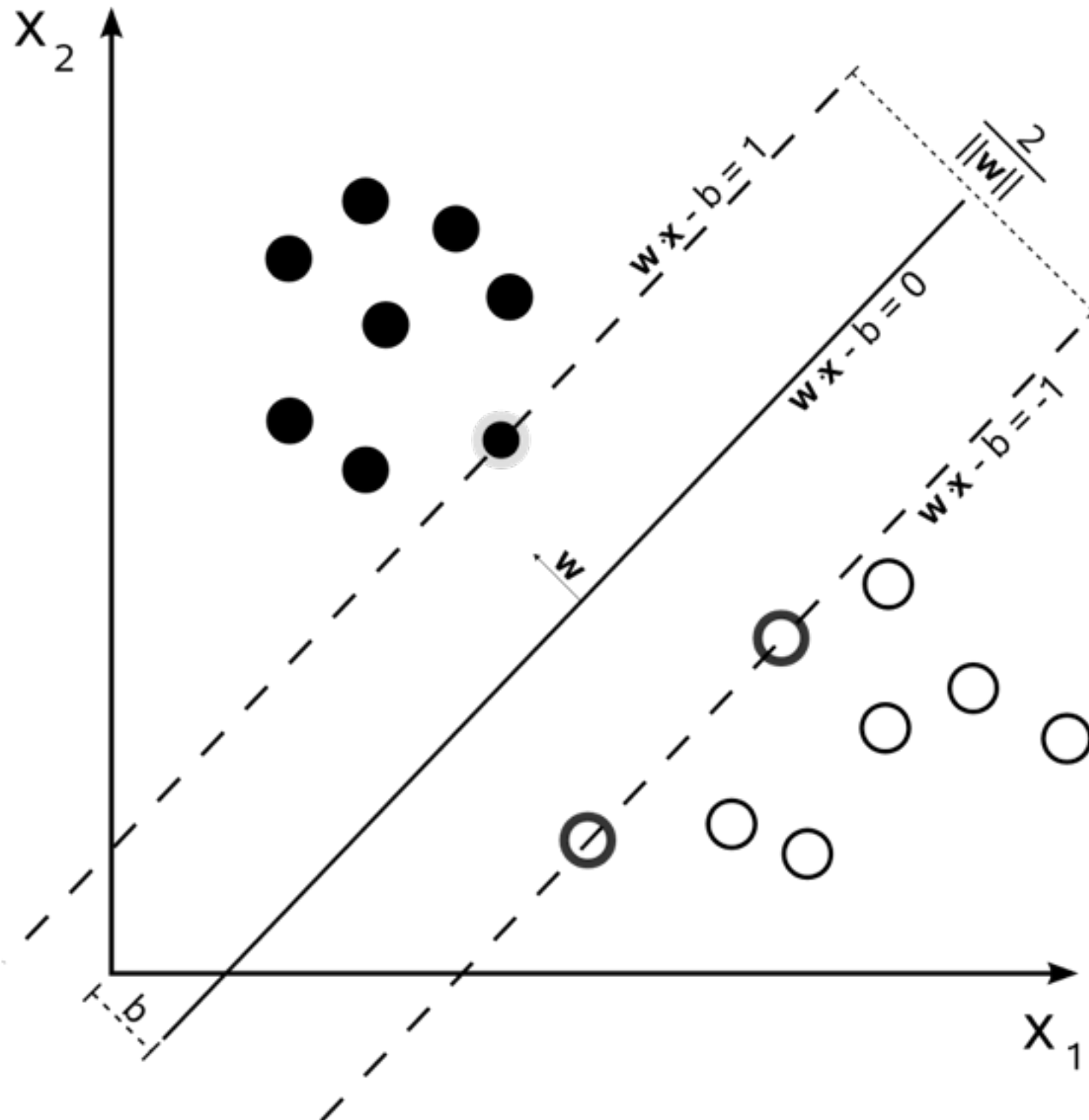
Classifying Members: Decision Tree



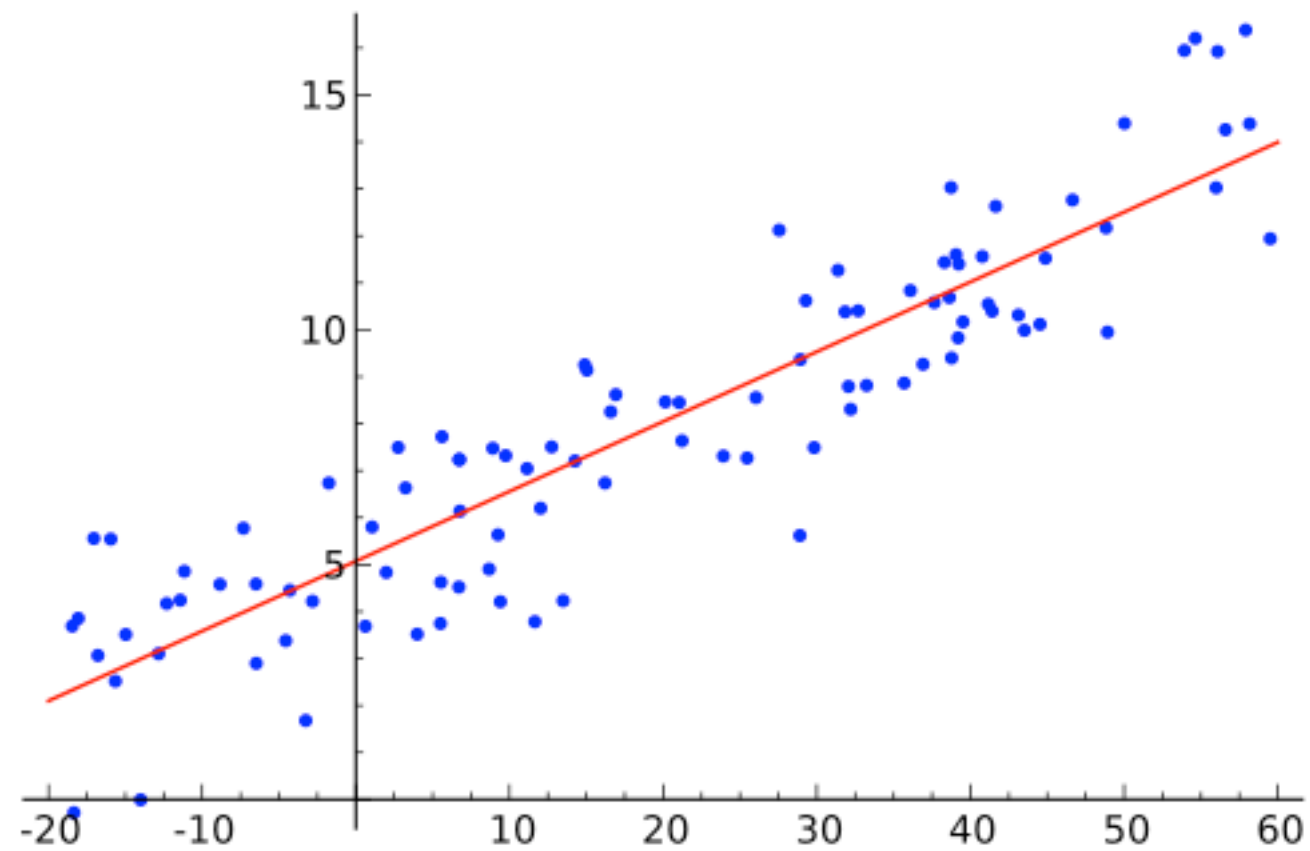
Classifying Members: Naive Bayes

$$p(C|F_1, \dots, F_n) = \frac{p(C) p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)}$$

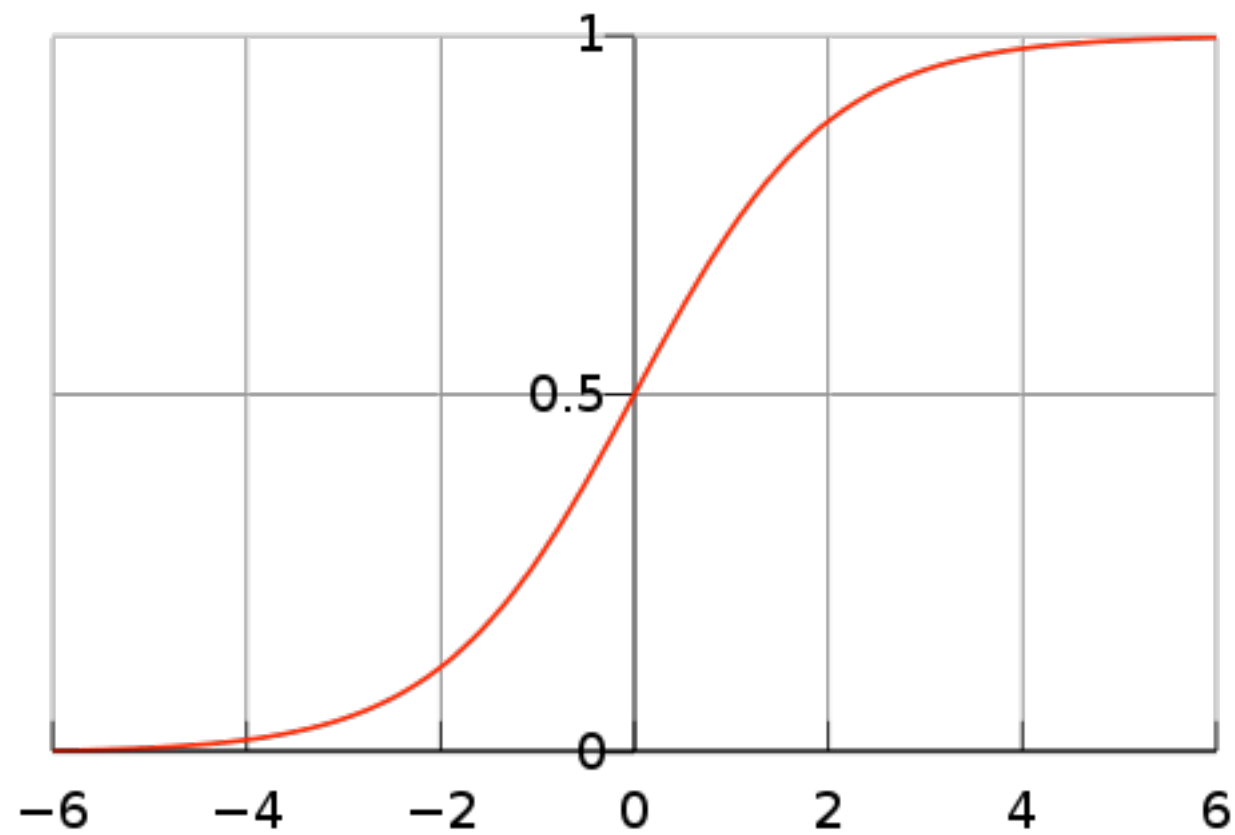
Classifying Members: Support-Vector Machines



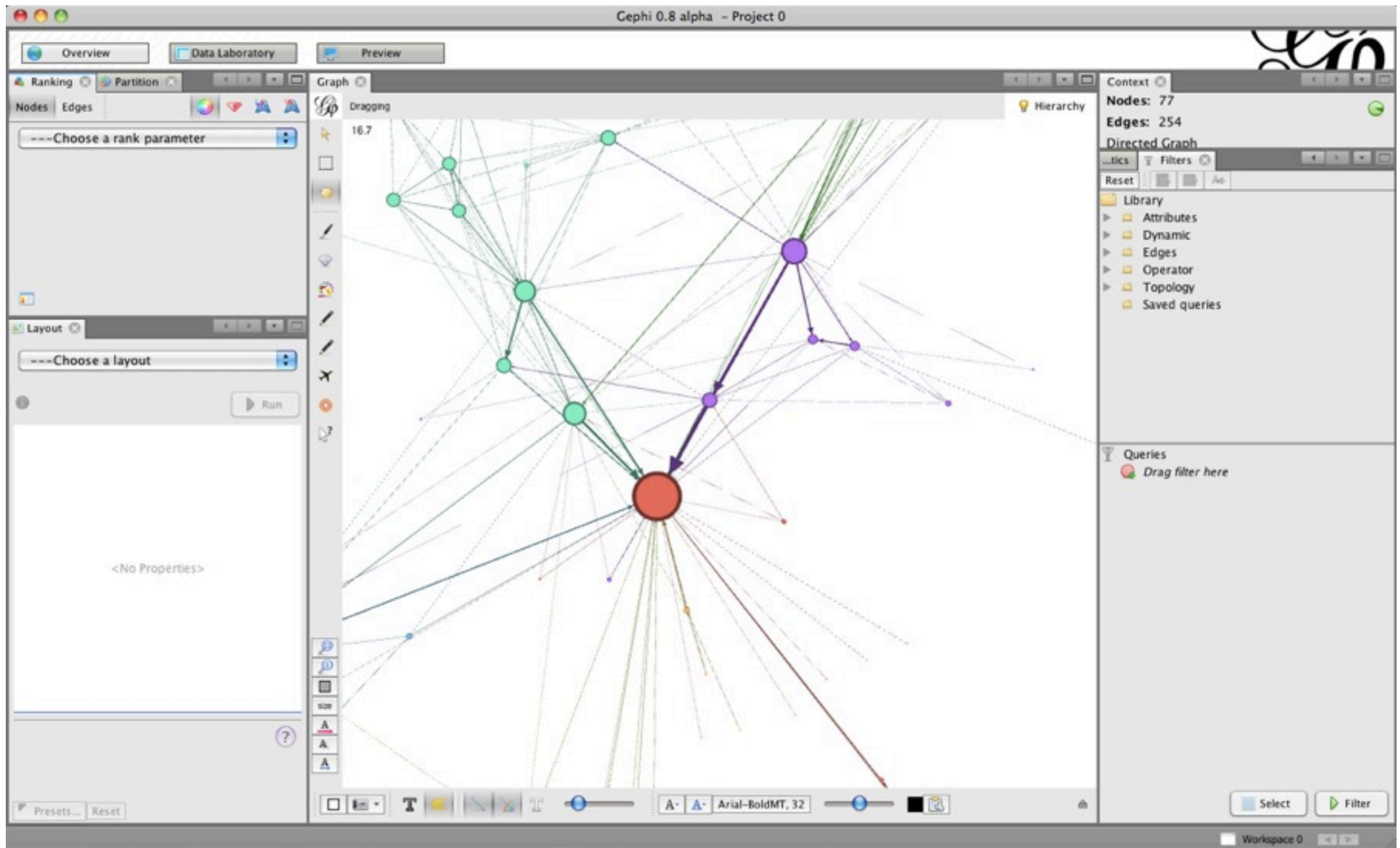
Modeling: Linear Regression



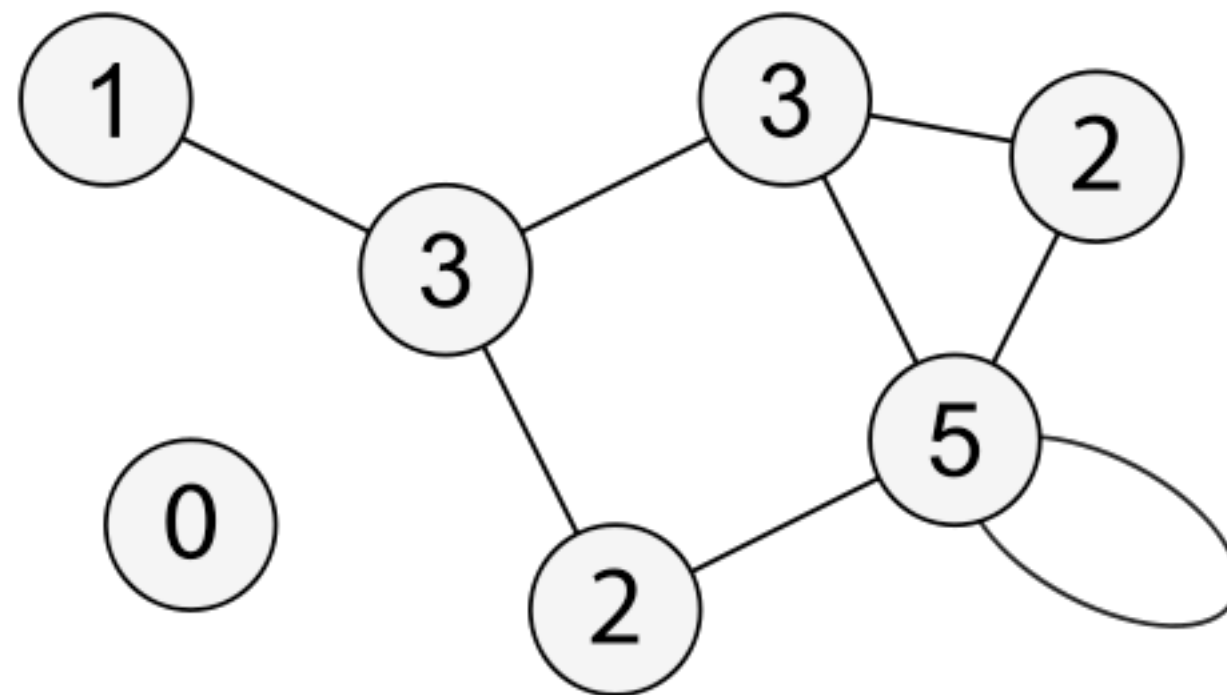
Classifying and Modeling: Logistic Regression



Gephi



Degree



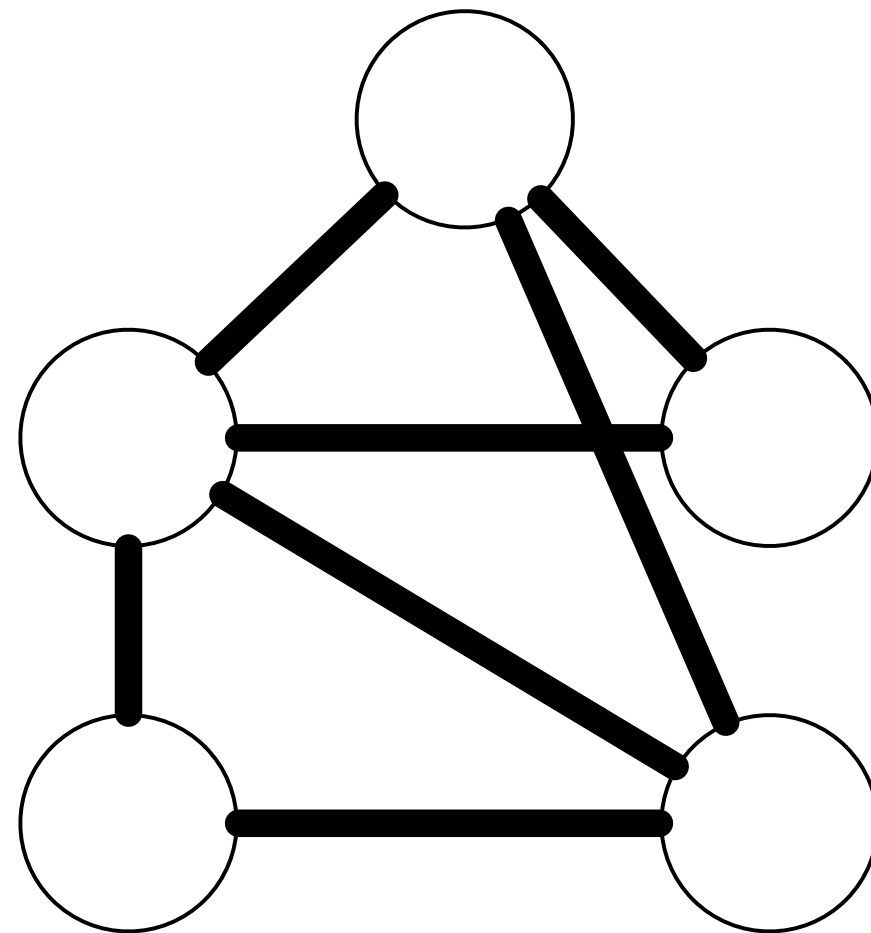
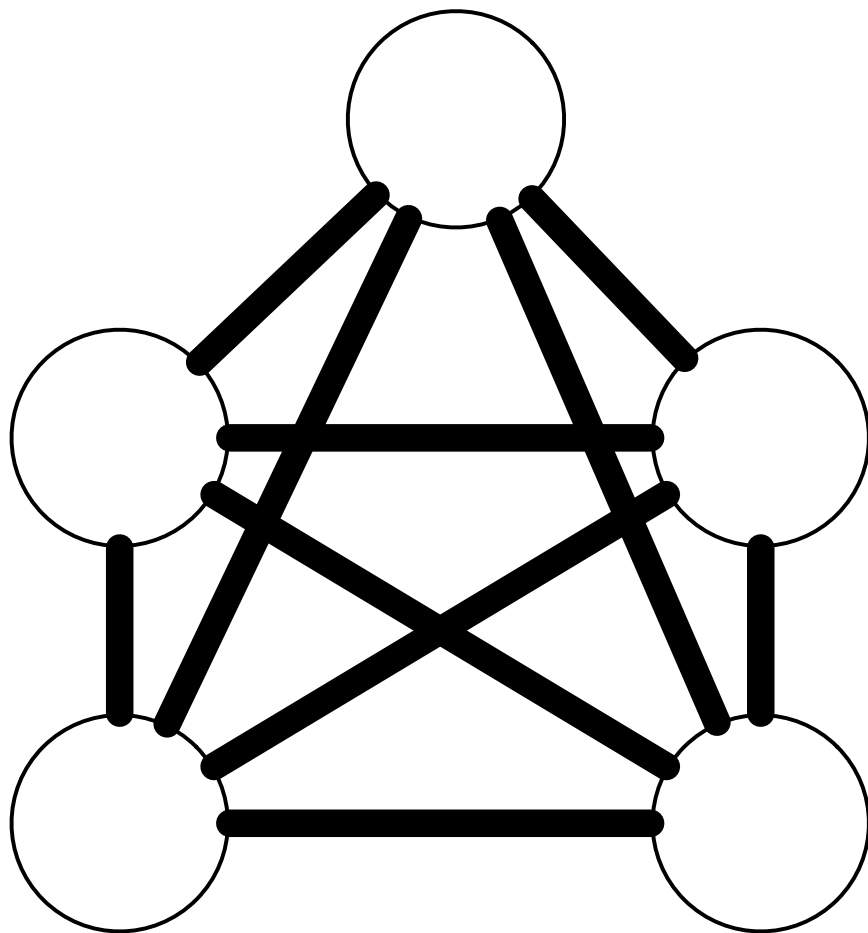
Paths and Diameter

- Average Path Length: average shortest path between all pairs of nodes
- Network Diameter: longest shortest path between two nodes

Centrality and Eccentricity

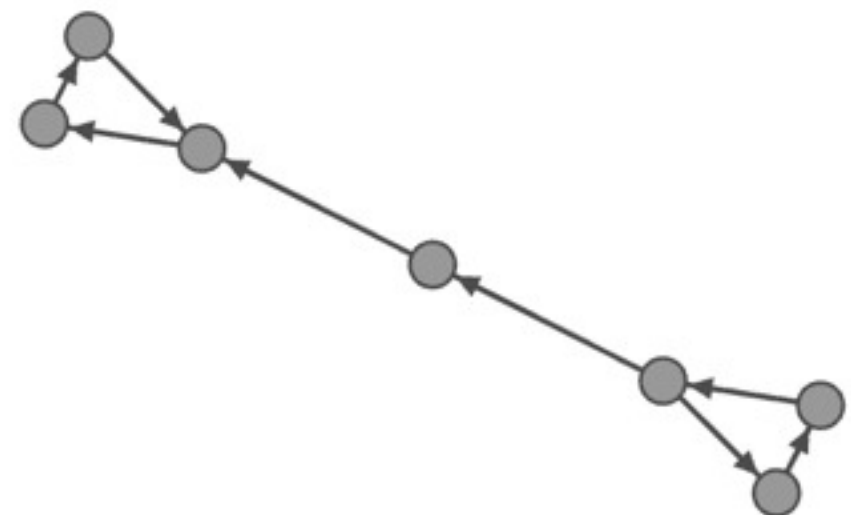
- Betweenness Centrality: how often does a node show up on shortest paths between nodes?
- Closeness Centrality: average distance from a given node to all other nodes in the network
- Eccentricity: maximum entry in the shortest path matrix for any given point

Graph Density



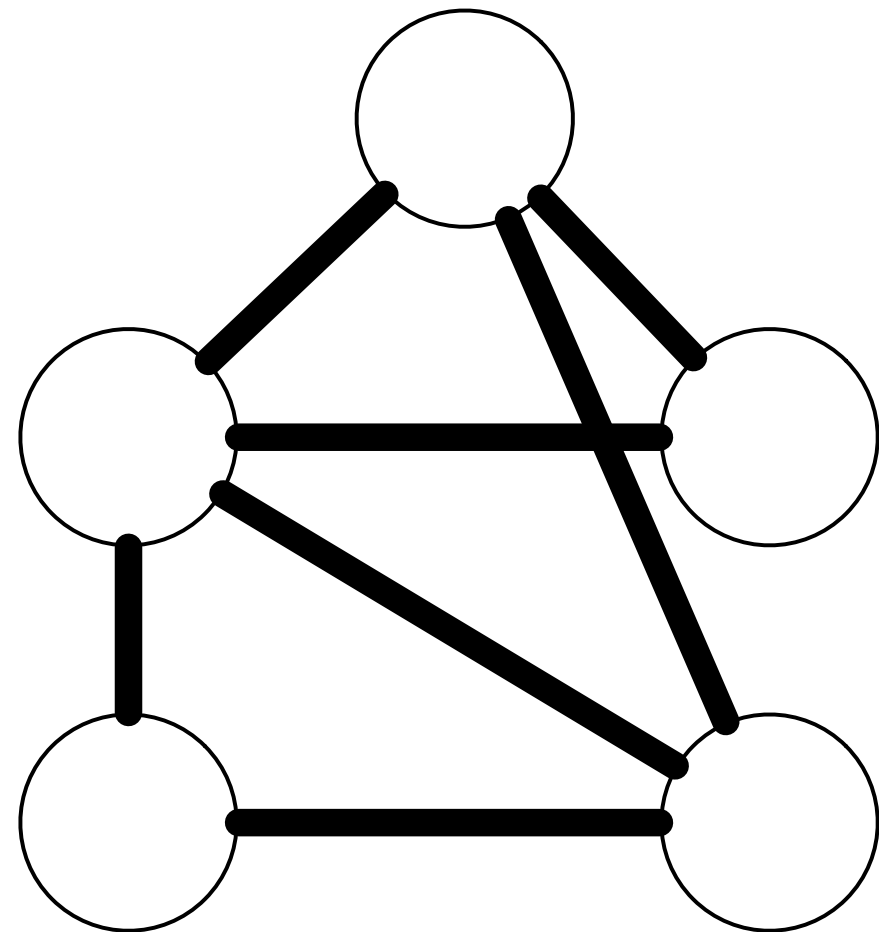
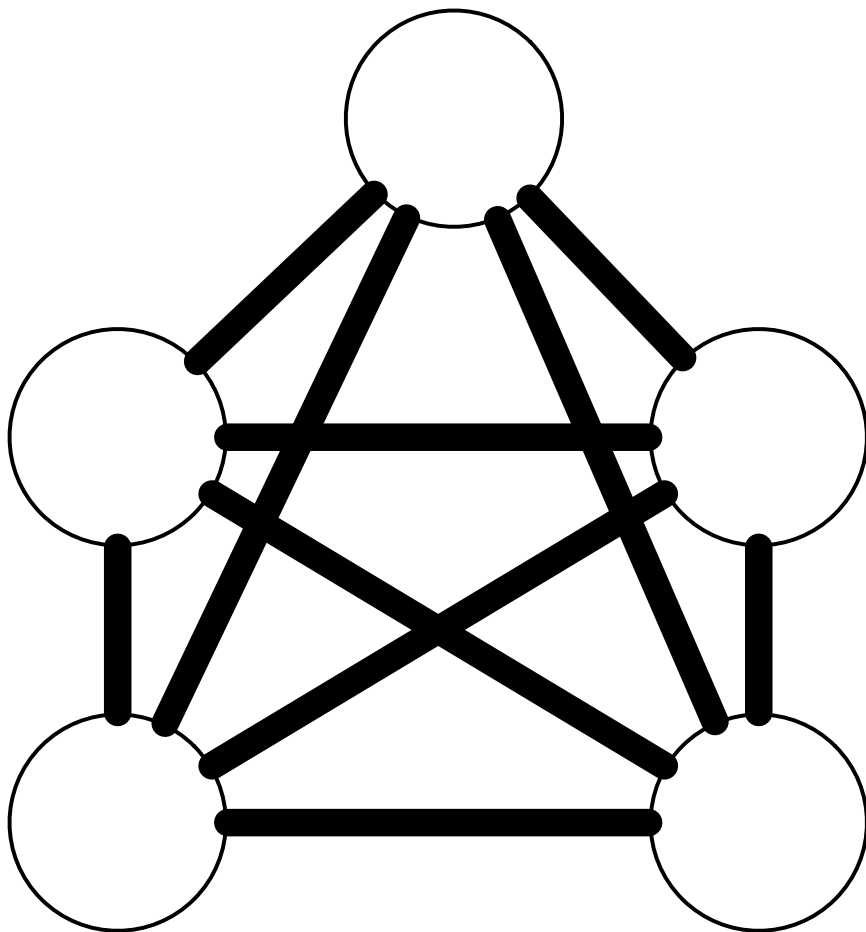
Modularity & Connectedness

- Modularity: measure of quality of division of network into modules (communities)
 - High modularity: corresponds to dense connections within modules, sparse connections between modules.
 - Modularity for a given division:
fraction of edges that fall within groups - fraction of edges if dist. at random
- Connected component: all nodes in the component are reachable by all other nodes in the same component
 - Strong: direction matters
 - Weak: direction does not matter



Clustering Coefficient

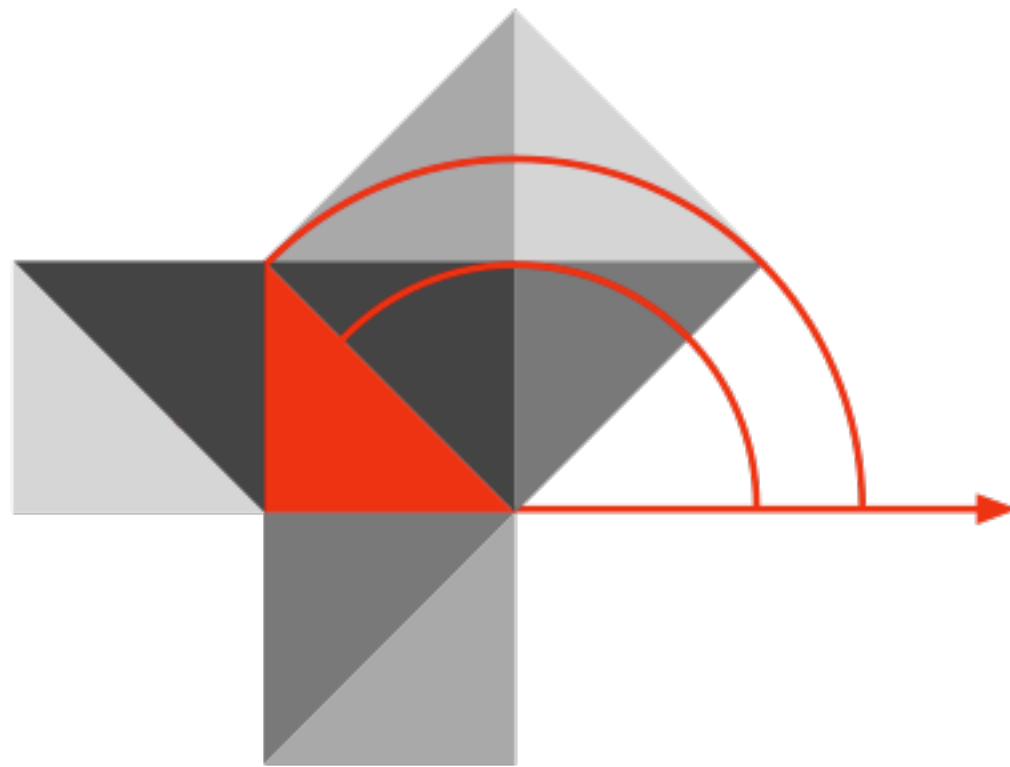
- Measures how close its neighbors are to being a clique (i.e., a complete graph)
- Ratio of number of triangles on a node (3 edges & vertices) to number of triples on a node (3 vertices, 2 edges)



Eigenvector Centrality

- Similar to PageRank
- Connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes
- A node is central to the extent that it is connected to other nodes that are central

Hippasus



<http://hippasus.com/rrpweblog/>
rubenrp@hippasus.com

This work is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 License.

