BIOSTATISTICS TRAINING FOR BASIC BIOMEDICAL RESEARCH (T32 GM074934)



37.2% of all interactions

COVER ART

1. Adam Richards - The four panels exemplify the major steps of the spectral clustering method. This a toy example of a gene set with quantitative distances for each of the gene pairs. The D Matrix panel is simply a heatmap of the input NxN distance matrix. The A matrix or affinity matrix is the distance matrix transformed using a Gaussian kernel in order to modulate the relative importance of the distances between genes. The X matrix are the k=3 eigenvectors corresponding to the largest eigenvalues associated with the Laplacian matrix. The last panel shows the data projected into eigenvector space along with estimates of their centroids as found using K-means clustering. The labels can then be mapped back to the original genes effectively partitioning the data into 3 subgroups.

2. Mary Shotwell - Location of bottlenose dolphin strandings (Codes 1-3) from 1993-2007. Red circles indicate fully measured animals, and black indicate those that are missing total body weight.

3. Adam Richards - The figure is a gene-regulatory network from a time-course microarray experiment. The mouse used in these studies has a mutation in the beta-subunit of the Phosphodiesterase 6 gene (Pde6b) which causes retinal degeneration. It is used as a model for diseases such as Retinitis pigmentosa and Macular degeneration. The graph consists of genes (circles) and transcription factors (diamonds) and the edges between represent putative gene regulatory interactions between them. The colored genes were discovered by a statistical method called correlation integral and this figure shows all the possible paths between two genes of biological interest. Further subnetwork analyses were performed to validate the use of correlation integral in the gene expression domain and to further provide biological evidence for novel genes associated with retinal degeneration. These work were published in the journal of computational and graphical statistics.

4. Matt Shotwell - The first 20 principle component (PC) "images" of a dataset consisting of 400 images of human faces. Each PC image represents the linear projection of an image consisting of 112x92 pixels onto a single PC "pixel". The first 20 PC pixels account for more than 70% percent of the dataset variability. This method was used to reduce the dimensionality of each original image for further analysis. That is, 70% percent of the variability in the original 400 images may be captured by just 20 PC pixels per image, rather than the original 112x92 = 10304 pixels per image. The dark regions in each PC image contribute most to the corresponding PC pixel. Hence, dark regions of the PC images are considered important discriminating features.

5. Matt Shotwell - An electrocardiogram trace before and after filtering with a convolution low-pass filter. This method is used to reduce high frequency noise that is common in electrophysiological recordings. This recording captures the electric potentials produced by the heart (sharp ticks), as well as the bias due to respiration (slow waves).

6. Brian Muller - A Steiner tree of a yeast KEGG pathway. Black nodes represent yeast proteins. Blue nodes represent Gene Ontology terms. A minimal spanning tree was constructed using the Gene Ontology for these terms.

7. Beth Wolf - replace X4 & X5 with X5 & X11) Within a Logic Forest, we say an interaction persists in the forest if it occurs in increasingly larger (includes more predictors) interactions within the forest. If an interaction persists in the forest, it is more likely that it is truly associated with the outcome. The plot, ÄúPersist for X4 & X5,Äù shows the persistence of the interaction term (X4 & X5) in a logic forest constructed for simulated data where Y = (X4 & X5) OR (X5 & X11). From the plot we can see in which interactions (X4 & X5) persists.

8. Beth Wolf - This plot shows the persistence of five interactions of size 2 that occur in a Logic Forest. These data are looking at the association between periodontal disease and several SNPs and health factors. The y-axis represents the top 5 interactions of size 2 (ranked by an interaction importance measure). The x-axis represents their persistence within interactions of equal or larger size. The diameter of the circles on the main plot represents the count of number of interactions of each size that occur within the forest, the darker bar represents how many of these interactions the 5 shown in the plot account for in each interaction size group. The histogram to the right of the plot shows the frequency distribution of all interactions of size 2 that occur in the forest and the darker bars specifically represent the interactions shown on the main plot.

9. Beth Wolf - The center circle represents the frequency of occurrence within the forest of (X4 & X5) exactly as (X4 & X5). The concentric rings represent increasing interaction sizes of 3, 4, and 5. The circles on the rings represent the additional terms that comprise larger interaction containing (X4 & X5). Circle size represents the frequency of occurrence within the forest of these interactions relative to (X4 & X5). - in this case of all interactions that occur in the forest, 37.2% of them include the term (X4 & X5).

10. Josh Swearingen – Dimension reduction via prinicple components analysis is used to show the increase in separation of both behavior (above) and associated neural acitivity (below) in the first day of rats acquiring new task behavior. Early indicates analysis on the first 100 trials, Late on the next 100 trials. X and Y axis are prinicple component 1 and 2. Circles represent scores of observations (time) in this space. 40 seconds preceded stimulus presentation (gray circles), with 40 seconds post stimulus (red circles), stimulus indicates to the animal that the trial has begun.