

Charlee Cobb - Transcriptomics, Exercise 8

Charlee Cobb

2023-03-26

- 1) ***What are the different ways one can come across a set of genes?*** We can find a set of genes on the GEO database we can also create a set of genes by how they are expressed in an experiment.
- 2) ***How can you learn from grouping genes with the same expression pattern?*** You can learn how they respond to certain elements or drugs, and you can see if the groups of genes has a specific function they all share.
- 3) ***What is the advantage, to the scientific community, for using GO-terms?*** Go terms help create a controlled vocabulary for specific terms in biology, so scientist can easily see and find the term for functions of a gene.
- 4) ***How can we use GO-terms to better understand our list of genes? What statistical method(s) are applicable for this analysis?*** To better understand our list of genes, we can use the Hypergeometric test on the genes in the set to determine what GO-terms best represent the set.
- 5) ***What measurement is needed in order to perform clustering?*** In order to perform clustering, we need a distance measure from the point (gene) to the correlation of the dataset.
- 6) ***What feature of gene expression is captured by Euclidean Distance.*** The feature of gene expression that is captured by Euclidean Distance is how far they are from each other.
- 7) ***What feature of gene expression is captured by Pearson Correlation.*** Pearson correlation captures how closely correlated a gene's expression is to another gene's expression.
- 8) ***How do we convert a correlation value to a distance value?*** You do this by simply subtracting the correlation value of two variables by 1.
- 9) ***What are the steps to create hierarchical clustering?*** First you find the shortest distance and group the two genes, then you recalculate the distances from the joined group to the rest of the matrix. After that, you find the next shortest distance, and then you repeat the previous two steps until it's complete.
- 10) ***What are the different linkage methods and what are they used for?*** Single linkage - looks at two groups and measures distance between closest member of the groups. Complete linkage - looks at furthest member if the groups. Centroid linkage - looks at the middle member of the groups. Average linkage - looks at the average calculation made between group members.
- 11) ***What are the steps for k-means clustering?*** First, make a random assignment and compute centroids. Then you assign points to the nearest centroid. And finally, recompute centroids and repeat this for a given number of iterations
- 12) ***What does the K in K-means mean?*** K stands for the number of clusters found.
- 13) ***What are the caveats of K-means clustering and how do you overcome them?*** The placement is completely random, you have to run this method a number of times to test the robustness.
- 14) ***Explain the silhouette width equation.*** The silhouette equation validates groups created in clustering. A value of or close to positive one means B is close to members of it's own group. And a value of negative one means it's close to members of another group.

15) *Is it possible to get clusters if you cluster random data?* Yes, it is possible. The equations simple perform on the given data.