COSC 681 Bioinformatics

Project 3 Gene Expression Analysis

2016/12/15

William Gillespie

**Part 1: Supervised Classification**

1. **Apply holdout method to divide samples into training set and testing set.**

I chose 3 different holdouts for the train/test parts: 50/50, 65/35, 75/25. I chose these because they each give a decent amount of training data while still having at lease 10 samples in the test data.

1. **Apply SAM to select informative genes**

I used the MeV software to select 78 informative genes. I chose 78 because that was the lowest number above 75 that I could select with the slider bar. The requirements were to select around 25-75. I wanted to select as many as possible while still being around that range.