**General Directions**

* Project two mostly covers chapter 10 and may also require methods covered in chapters 6 and 8 of your text depending your analysis. From the following website <http://bowtie-bio.sourceforge.net/recount/> scroll down and select “wang” dataset. Details of the dataset are available through the PMID accession code. In order to access data, you need to download both Count table and Phenotype table from the links available for this data. PMID accession code also allows you to read the main publication that describes this data.
* Apply a dimension reduction method to reduce the number of significant genes in the dataset.
* Apply suitable unsupervised learning methods to the dataset in order to make suitable interpretation of the patterns in the data.
* Finally apply suitable supervised learning methods if possible to better understand the information available in the data.
* This project should be completed in R Markdown editor. A suggested format is on the next page.
* File Title: Project2\_sam (the initials sam stand for sample)
* Submit Project two output file to Bb on or before the due date.
* If an extension is warranted, permission must be secured 48 hours prior to the due date.
* The specific tasks within the project and point values are shown on the next page.
* 50 points are possible on this project.

Please use the following format for Project 2

Name

Overview

|  |  |  |
| --- | --- | --- |
| **Task** | **Project 1 – Classification and Cross Validation** | **Points**  **Possible** |
| 1 | Provide an overview of the data set you selected. | 4 |
| 2 | Apply dimension reduction methods. | 10 |
| 3 | Apply unsupervised learning methods. | 10 |
| 4 | Apply supervised learning methods. | 10 |
| 5 | Interpret the results obtained in 2,3 and 4 in your own words. | 10 |
| 6 | Draw conclusion and state any challenges that you faced in analyzing this data. | 6 |
|  | Total Points | 50 |