**Written Report – 6.419x Module 2**

**Name:** (Felipe Mehsen Tufaile)

* **Problem 2**
* **Part 1: Visualization**

*1. (3 points) Provide at least one visualization which clearly shows the existence of three main brain cell types as described by the scientist, and explain how it shows this. Your visualization should support the idea that cells from different groups can differ greatly. (Maximum 200 words)*

**Solution:**

The following visualization was obtained by plotting the “X\_log” data (that is the log2(X+1) transformation of the original data) on the first two principal components of the t-SNE embedding for gene expression. It is possible to note that t-SNE keeps similar type of brain cells together and different types of brain cells relatively far apart. We see that there is very little overlap in the intervals that define each of the clusters meaning that each cell type is defined by very different domains regarding PC1 and PC2 (in other words *different groups differ greatly*):

* Brain cell type I (Pink-cluster): , , Center: (-11.2, -2.1);
* Brain cell type II (Purple-cluster): , , Center: (12.1, 7.3);
* Brain cell type III (Beige-cluster): , , Center: (13.3, -32.9);

Nevertheless, it is still possible to note outliers cells that are positioned far away of their expected cluster. For example, some Brain cell type II (purple-cluster) are positioned closest to the Brain cell type I cluster (pink-cluster) and to the Brain cell type III cluster (beige-cluster).

A screenshot of a computer game

Description automatically generated

*2. (4 points) Provide at least one visualization which supports the claim that within each of the three types, there are numerous possible sub-types for a cell. In your visualization, highlight which of the three main types these sub-types belong to. Again, explain how your visualization supports the claim.*

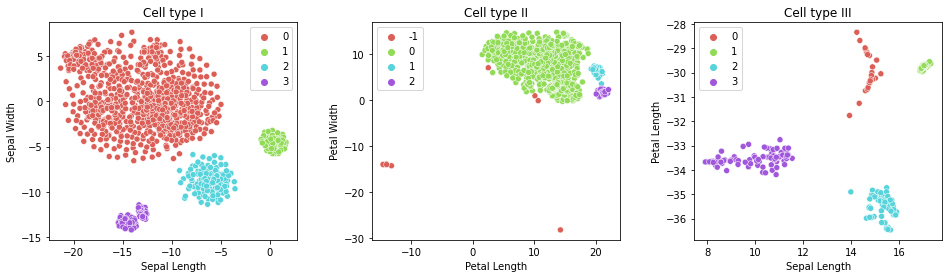
**Solution:**

The same visualization as before can be used to show the sub-types for a cell. If we look at the Brain cell type I (Pink-cluster) we can see a large aggregation of points surrounded by 3 other smaller groups of points. Each one of these 4 aggregation of points mentioned (1 large group and 3 other smaller groups) would represent sub-types of the Brain cell type I (Pink-cluster). The same reasoning applies to the Brain cell type III (Beige-cluster): we can see that this cluster is formed by 4 small aggregations of points, meaning 3 sub-types of the Brain cell type III (Beige-cluster).

A screenshot of a computer game

Description automatically generated

To better visualize these sub-types of cells, we could perform a DBSCAN clustering using the two first components of the t-SNE embedding. This will result in the following visualizations:



We can confirm our previous assumption that Brain cell type I (Original Pink-cluster) has 4 sub-type cells and Brain cell type III (Original Beige-cluster) also has 4 sub-type cells. What wasn’t so clear before is that Brain cell type II (Original Purple-cluster) is constituted by 3 sub-type cells. The red cluster in the cell type II seems to be indicating the outliers of this cluster as pointed out in the previous question.

**Reference**

[1] D. Freedman, R. Pisani and R. Purves, “Statistics,” Fourth Edition, Norton & Company, 2007.

[2] H. K., Beecher, “Measurement of Subjective Responses,” Oxford University Press, 1959, pp.66-67.