CS2263 Assignment #4 - summer 2025

- .

Closest Points in N-Dimensional Space

In a 2-Dimensional space, the distance between 2 points (x_1, y_1, x_2, y_2) can be solved using the following Euclidean distance formula: $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$. This same logic can be used to find the distance between two points in a 3-Dimensional space: $\sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$. In fact, in any n-Dimensional space, the same logic can be applied: $\sqrt{\sum_{i=1}^{n}(p_2 - p_1)^2}$ where p is simply the axis (x, y, z, a, etc.) and we cycle through all axes. This knowledge can be used to make predictions for example, if we had a large data set of previous patients diagnosed with a condition of varying severity: (0) Not Diagnosed, (1) Diagnosed with Minimal Side Effects, and (2) Diagnosed. If we were to take the cell measurements of a new patient, we could use our data set to see which other data points our new patient's measurements are close to in order to predict the likelihood the new patient would be diagnosed.

For a simplified example, imagine we have the following data recorded:

Patient	M1	M2	M3	Diagnosis
1	1	1	1	1
2	3	3	3	2
3	4	3	3	2
4	10	4	6	3

Now, if a new patient comes in with the following measurements: 12, 5, 7, we could find the Euclidean distance between this new point and all other points in the data set. Whichever old data the new patient's data is closest to could be a prediction of what their condition may be. Note: In practice, this is not how we would predict as this prediction method may not be very accurate.

You have been given a .txt file on D2L that has various measurements (1 patient per line and comma-delimited list), with the following breakdown (M means measurement):

Patient Number, M1, M2, M3, M4, Diagnosis

Read in the .txt file and from command line and take a comma delimited list of the new patient's measurements from terminal input. Then, figure out which patient **with no missing data** is closest according to Euclidean distance to the new patient. Print out the diagnosis of the "closest" previous patient as a prediction.

Terminal input format: Patient Number, M1, M2, M3, M4, example:

```
21, 50.5, 135.1, 43.3, 68.2
```

Two functions you may find useful:

1. `float ** dataReadIn(char * fileName, int * numRows)`
2. `float euclideanDist(float * row1, float * row2, int length)`

Remember good coding etiquette. Create and submit a make file that compiles all code necessary to run your main file. The teaching assistant should just be able to type "**`make`**" in the terminal to compile your program.

Your electronic submission (submitted via Desire2Learn) will consist of two files.

1. A single pdf file containing your source code and make file from the assignment (include functions and testing).
2. A zip file containing your code and make file.