Harnoor Singh

Machine Learning

HW2

1. It is possible to convert the DNA string into a 2D array where the letters ACTG are assigned to some arbitrary 2D value (1, 0), (0,1), (-1,0), or (0,-1). Therefore, the distance between any discrepancy within two strings at any point will be two, via the Manhattan distance. Calculating the total distance will be equal to the sum of these discrepancies.

If we will use Eucledian distances then graph would be identical to PCA graph. Minimizing the linear distances is same as maximizing the linear correlation

2. The results appear to be more or less equal. The clustering appears not to change significantly as a result of using either method.