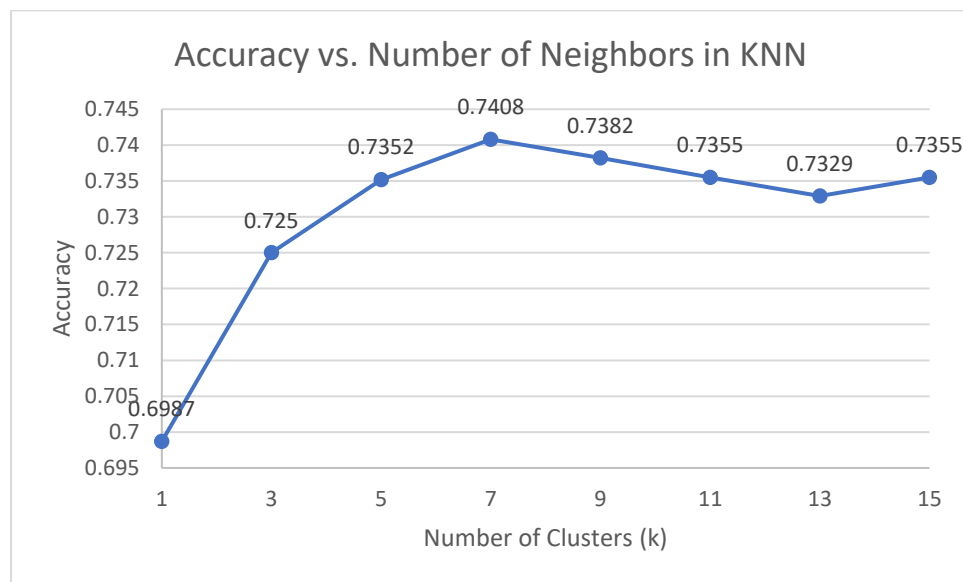


KNN (K Nearest Neighbors)

Below are the results of normal, unweighted KNN with $k = \{1, 3, 5, \dots, 15\}$. Accuracies are plotted for each of the eight simulations, as they correspond to each data point on the plot. We clearly see an increase in accuracy as the number of nearest neighbors increases by a significant amount until $k = 7$. However, any higher value of k results in a similar or worse accuracy at the cost of added computations. Thus, $k = 7$ proves to be the best number of neighboring data points for the Pima dataset.



Weighted KNN

The results for weighted KNN do not showcase a significant boost in accuracy with variation of sigma. There is some increase in accuracy as sigma increases, which follows the general idea of the weight formula (to an extent, this varies from dataset to dataset). As sigma increases, neighbors further away from the current test sample are weighted less:

$$w_i = e^{\frac{-||\mathbf{x} - \mathbf{x}_i||^2}{\sigma^2}}$$

Note, in the formula, as sigma decreases, the denominator becomes exponentially smaller, thus e is raised to an even larger negative power, resulting in a lower weight value for all samples. In the process, larger distances, which are squared in the numerator of the exponent, are penalized even further as sigma decreases. But, this only holds until a sigma, specific to the dataset, is reached where the perfect balance between weights in far samples and near samples exists.

The results are not significant, but showcase a nice lesson of the impact on accuracy as the value of sigma varies. A sigma value of 0.5 results in a slightly better accuracy compared to sigma values 0.2 and 0.01 .

