// Munir Vafai, Collins Kariuki

1. Average Perceptron 10-fold cross validation

0.7464788732394367

0.7323943661971831

0.5211267605633801

0.8732394366197183

0.8309859154929577

0.7605633802816901

0.8028169014084507

0.8450704225352113

0.7323943661971831

0.8533333333333333

Average accuracy for 10 iterations: 0.7698403755868545

2. Average Perceptron non-binary

0.4084507042253521

0.5915492957746479

0.7323943661971831

0.5915492957746479

0.6338028169014084

0.6056338028169013

0.5352112676056339

0.5774647887323943

0.6197183098591549

0.6266666666666668

Average accuracy for 10 iterations: 0.5922441314553991

3. Average Knn 10-fold cross validation

0.6619718309859156

0.6197183098591549

0.5211267605633801

0.7323943661971831

0.8309859154929577

0.7746478873239437

0.7464788732394367

0.8028169014084507

0.704225352112676

0.6799999999999999

Average accuracy for 10 iterations: 0.7074366197183098

Average Knn non-binary

0.676056338028169

0.6619718309859156

0.7183098591549297

0.6197183098591549

0.6338028169014084

0.5633802816901409

0.5633802816901409

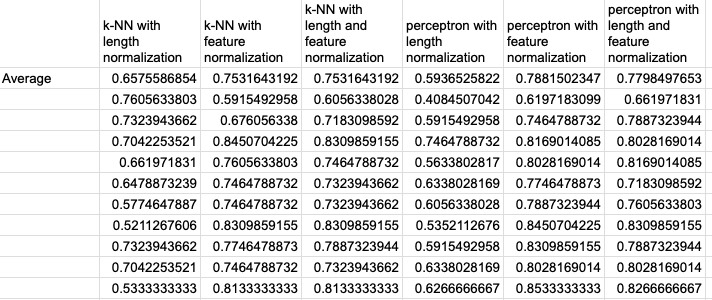
0.6338028169014084

0.7464788732394367

0.5199999999999999

Average accuracy for 10 iterations: 0.6336901408450704

4.

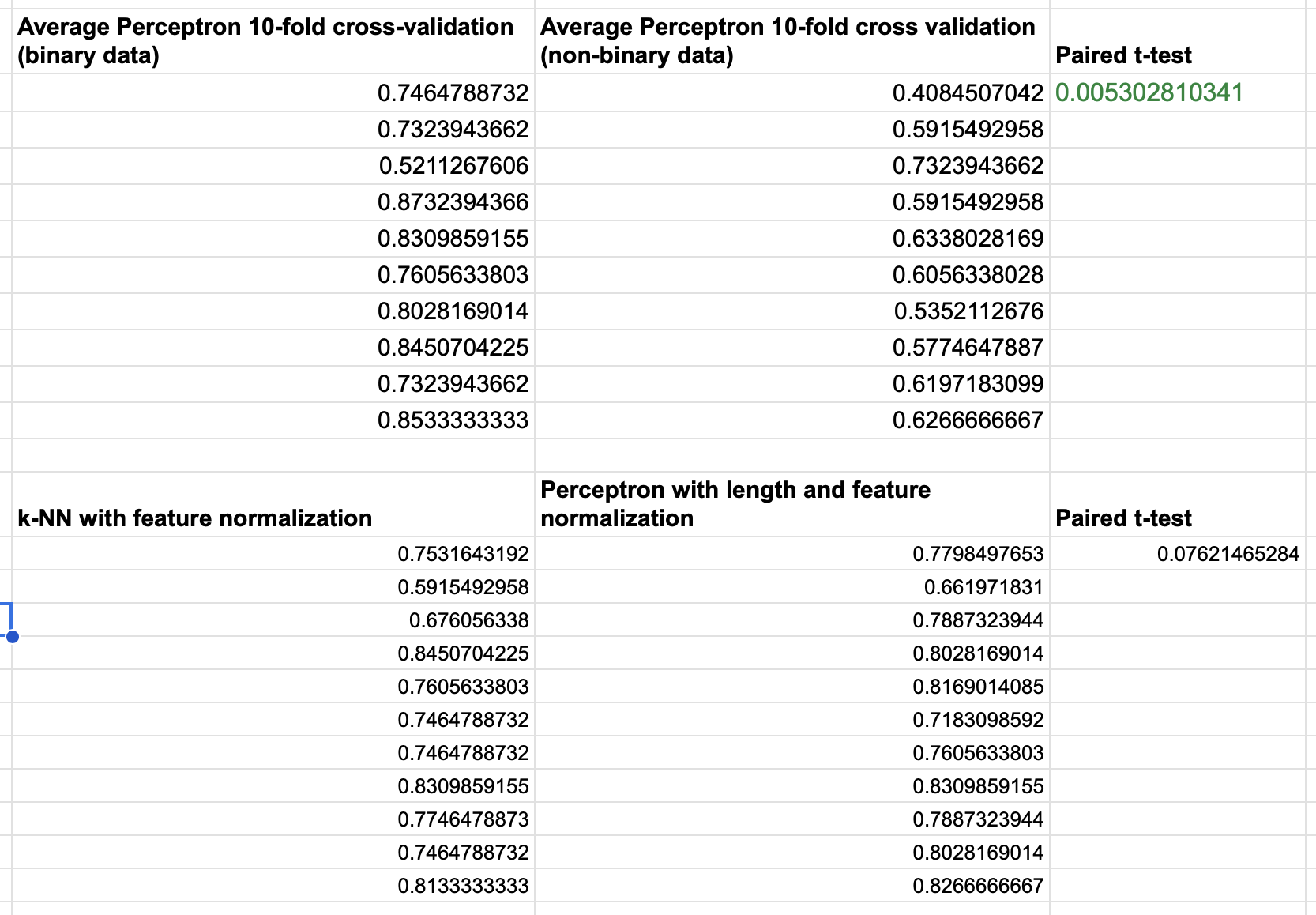


5. For this experiment, we shall pick:

* Average Perceptron 10-fold cross validation (on the old binary data).
* Average Perceptron 10-fold cross validation (on the non-binary data).
* k-NN with feature normalization.
* Perceptron with length and feature normalization.

The averages of the first two are far apart while the the averages of the latter two are close.

The t-test scores:



*6. Interpretation*:

In the first table, the t-test value of 0.0053 is significantly below the common significance level of 0.05. This suggests that the average perceptron runs on binary and non-binary data have similar means, with the small difference likely arising from random chance. In summary, the choice between binary and non-binary data doesn't significantly affect the results.

However, in the second table comparing k-NN with feature normalization and the perceptron with length and feature normalization, the t-test value is higher, indicating that the means of these two data sets are statistically different, and the difference is unlikely to be due to random chance.