kNN Report

Student: Anrui Wang #U75971461

For this part, I implement 2 knn algorithm to predict the label of the test sets.

To process the data, I treat them in two ways. For those numerical(real) value, I calculate the mean of this column and first replace the "?" with np.nan(that's numpy's missing number) and then replace the np.nan with the mean I computed. For those values that are not numerical, I replace them with the value which appear the most(mode).

kNN is very simple, I just calculate the L2 norm distance between every pair of training value and testing value and sort them, and find the smallest k th distance. Then I record the label in this k th set and relabel the testing value's label with a label which appear the most in k th smallest set. Finally I calculate the accuracy using (number of value that truly labeled)/(number of all data).(I did not use accuracy.pl but calculate it myself)

And the accuracy is just as below:

```
For crx data:
When k = 3, the accuracy is 0.674
When k = 4, the accuracy is 0.659
When k = 5, the accuracy is 0.659
When k = 6, the accuracy is 0.674
When k = 7, the accuracy is 0.63
For lense data:
When k = 1 ,the accuracy is 0.833
When k = 2 ,the accuracy is 0.833
When k = 3 ,the accuracy is 0.833
When k = 4 ,the accuracy is 0.833

Process finished with exit code 0
```