Report :: TIPR Assignment – I

Mohit Gupta

15755

TASK1 – Done.

TASK2- Done.

Task3- Looking at the plots. Most of the time low-dimension gives us better results. Sometimes accuracy doesn't vary much from low-dim and high-dim. So, I'll be with Akbar.

Used skfold cross validation where k=10.

Plots are in the output_plots folder with task_3_<dataset>_<K>_<Model>.png

Task 4:- Overall (except some cases like for dolphins dataset which is very small), at low-dimensions, model performs better. So, Akbar.

Used skfold cross validation where k=10.

Plots are in the output_plots folder with task_4_<dataset>_<K>_<Model>.png

Task 5:-

In general, for **KNN** results are often similar. But mine approach had little more accuracy in some cases. Like for PUBMED dataset. But for dolphins, generally SKlearn performed better.

For Twitter dataset, I see the plot very same but I had a better accuracy by 2%.

In general, for **Naïve-bayes**, results were generally better in sklearn over mine except for few values of K, mine model got a better accuracy.

I see that for naïve bayes, there turns out to be underflow/overflow error for twitter dataset. Thus my results may not be as accurate as sk-learn. I tried likelihood then got underflow so I tried using log-likelihood but then got overflow error for twitter dataset. I tried Multinomial and gaussian naïve bayes but same results.

Task 6:- Not done.