

Project 4 Report

Results

```
lda datasets/iris_shuffled.csv
Training Size (instances): 120
Testing Size (instances): 30
Run Time (s): 0.002611509000416845
Accuracy: 1.0
```

```
pca datasets/iris_shuffled.csv
Training Size (instances): 120
Testing Size (instances): 30
Run Time (s): 0.002102181955706328
Accuracy: 1.0
```

```
kpca datasets/iris_shuffled.csv
Training Size (instances): 120
Testing Size (instances): 30
Run Time (s): 0.012456855969503522
Accuracy: 1.0
```

```
-----
lda mnist
Training Size (instances): 1437
Testing Size (instances): 360
Run Time (s): 0.011285066953860223
Accuracy: 0.9055555555555556
```

```
pca mnist
Training Size (instances): 1437
Testing Size (instances): 360
Run Time (s): 0.0055523019982501864
Accuracy: 0.9055555555555556
```

```
kpca mnist
Training Size (instances): 1437
Testing Size (instances): 360
Run Time (s): 0.014835996960755438
Accuracy: 0.9055555555555556
```

Analysis

The dimension reduction methods are meant to improve our accuracy. The PCA method is unsupervised. The LDA method is linear and the kernel PCA method is made to handle non linear cases. My results all came out the same. I had trouble converting the `load_digits` return type of Bunch to a dataframe. Instead, I tried not converting it at all and just returning the data immediately, but that gave a warning that my data was 'collinear' on each run.