

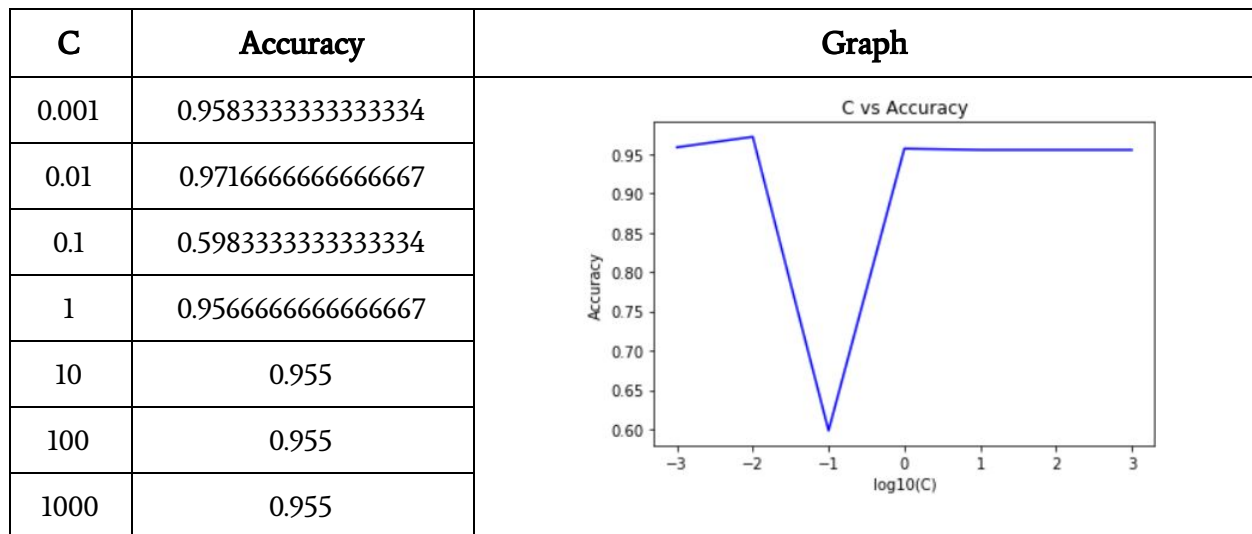
# Support Vector Machines | Decision Tree | K-means

Priyank | 2016MT10628

## SVM:

(a) Linear Kernel:

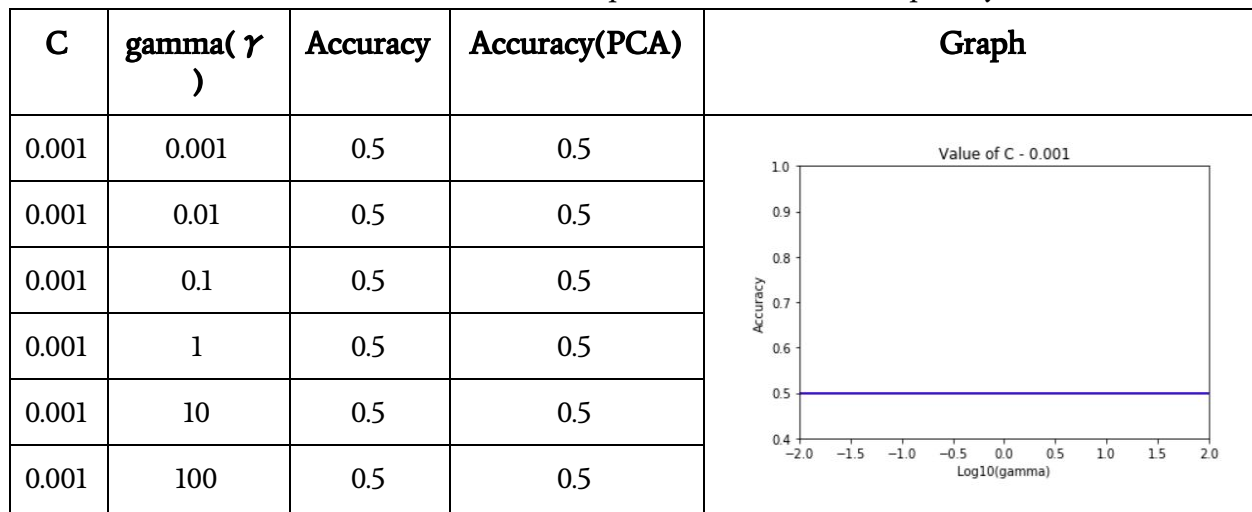
The following is the C vs Accuracy table and plot.



(b) RBF Kernel

(c) RBF Kernel with PCA

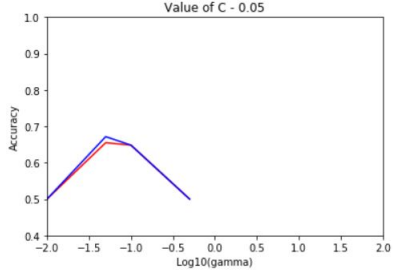
The following is the gamma(  $\gamma$  ) vs Accuracy table and plot with C = 0.001. The blue line (PCA) and the red line(without PCA) overlap with each other completely



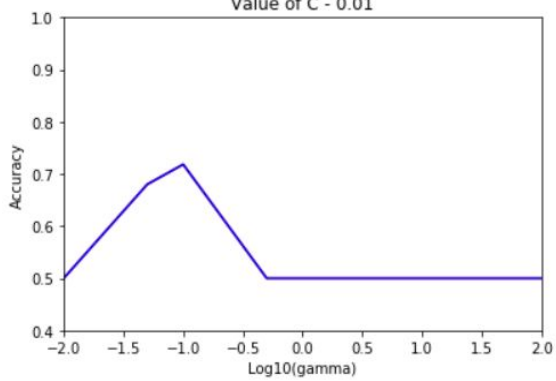
The following is the gamma(  $\gamma$  ) vs Accuracy table and plot with C = 0.05. The blue line (PCA) and the red line(without PCA) doesn't overlap completely, PCA works better.

C	gamma( $\gamma$ )	Accuracy	Accuracy(PCA)	Graph
---	-------------------	----------	---------------	-------

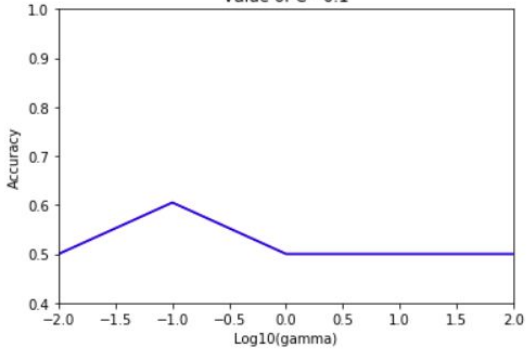
0.05	0.01	0.5	0.5
0.05	0.05	0.655	0.671666
0.05	0.1	0.648333	0.648333
0.05	0.5	0.5	0.5



The following is the gamma(  $\gamma$  ) vs Accuracy table and plot with C = 0.01. The blue line (PCA) and the red line(without PCA) overlap with each other completely

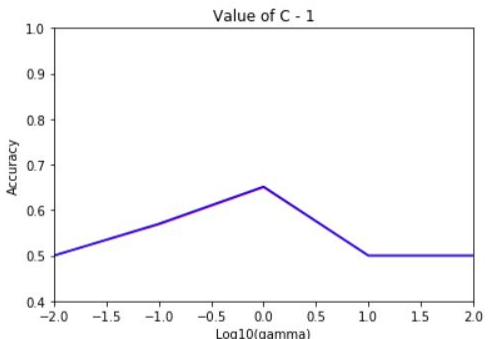
C	gamma( $\gamma$ )	Accuracy	Accuracy(PCA)	Graph
0.01	0.001	0.5	0.5	
0.01	0.01	0.5	0.5	
0.01	0.05	0.679999	0.679999	
0.01	0.1	0.718333	0.718333	
0.01	0.5	0.5	0.5	
0.01	1	0.5	0.5	
0.01	10	0.5	0.5	
0.01	100	0.5	0.5	

The following is the gamma(  $\gamma$  ) vs Accuracy table and plot with C = 0.1. The blue line (PCA) and the red line(without PCA) overlap with each other completely

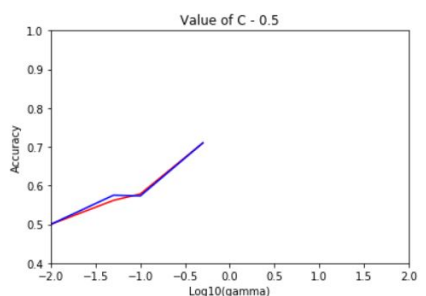
C	gamma( $\gamma$ )	Accuracy	Accuracy(PCA)	Graph
0.1	0.001	0.5	0.5	
0.1	0.01	0.5	0.5	
0.1	0.1	0.605	0.605	
0.1	1	0.5	0.5	
0.1	10	0.5	0.5	
0.1	100	0.5	0.5	

The following is the gamma(  $\gamma$  ) vs Accuracy table and plot with C = 1. The blue line (PCA) and the red line(without PCA) overlap with each other completely

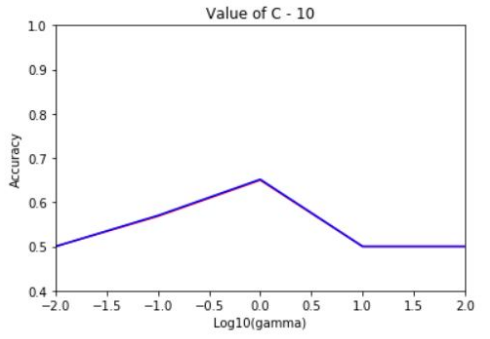
C	gamma( $\gamma$ )	Accuracy	Accuracy(PCA)	Graph
---	-------------------	----------	---------------	-------

	$\gamma$			
1	0.001	0.5	0.5	
1	0.01	0.5	0.5	
1	0.1	0.568333	0.570000	
1	1	0.65	0.651666	
1	10	0.5	0.5	
1	100	0.5	0.5	

The following is the gamma(  $\gamma$  ) vs Accuracy table and plot with C = 0.5. The blue line (PCA) and the red line(without PCA) doesn't overlap completely, PCA works better

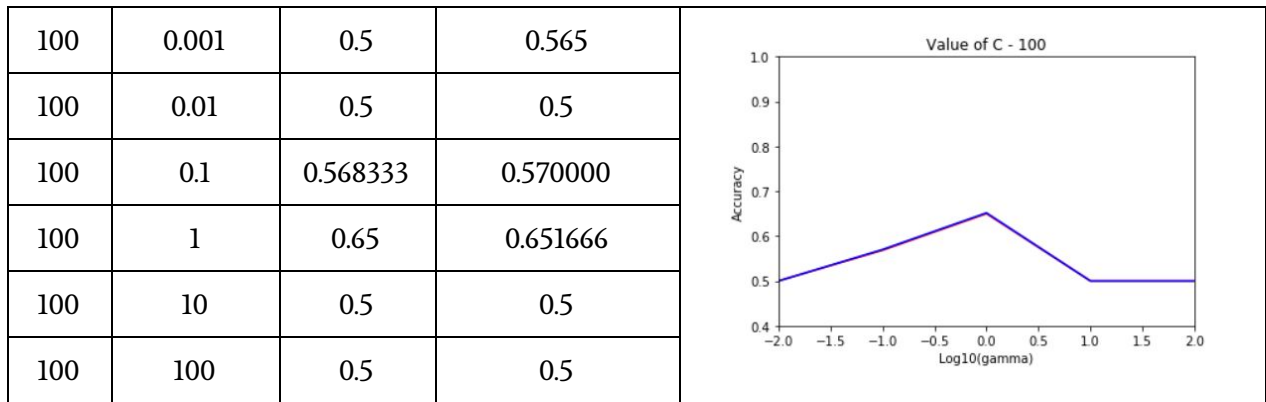
C	gamma( $\gamma$ )	Accuracy	Accuracy(PCA)	Graph
0.5	0.01	0.5	0.5	
0.5	0.05	0.561666	0.575	
0.5	0.1	0.578333	0.573333	
0.5	0.5	0.71	0.71	

The following is the gamma(  $\gamma$  ) vs Accuracy table and plot with C = 10. The blue line (PCA) and the red line(without PCA) overlap with each other completely

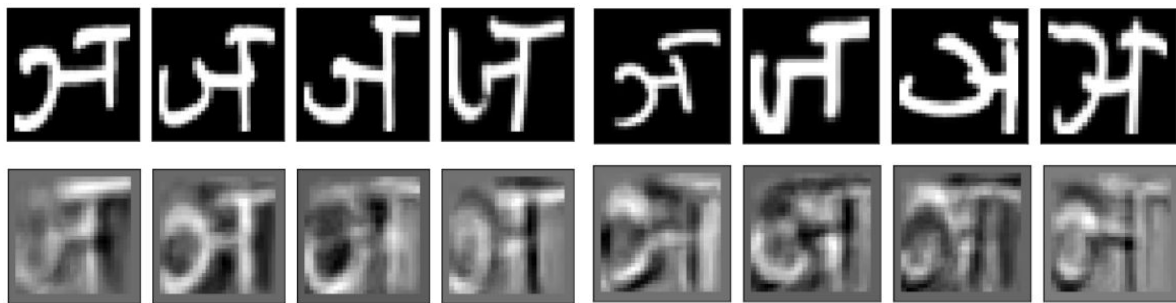
C	gamma( $\gamma$ )	Accuracy	Accuracy(PCA)	Graph
10	0.001	0.5	0.5	
10	0.01	0.5	0.5	
10	0.1	0.568333	0.570000	
10	1	0.65	0.651666	
10	10	0.5	0.5	
10	100	0.5	0.5	

The following is the gamma(  $\gamma$  ) vs Accuracy table and plot with C = 100. The blue line (PCA) and the red line(without PCA) overlap with each other completely

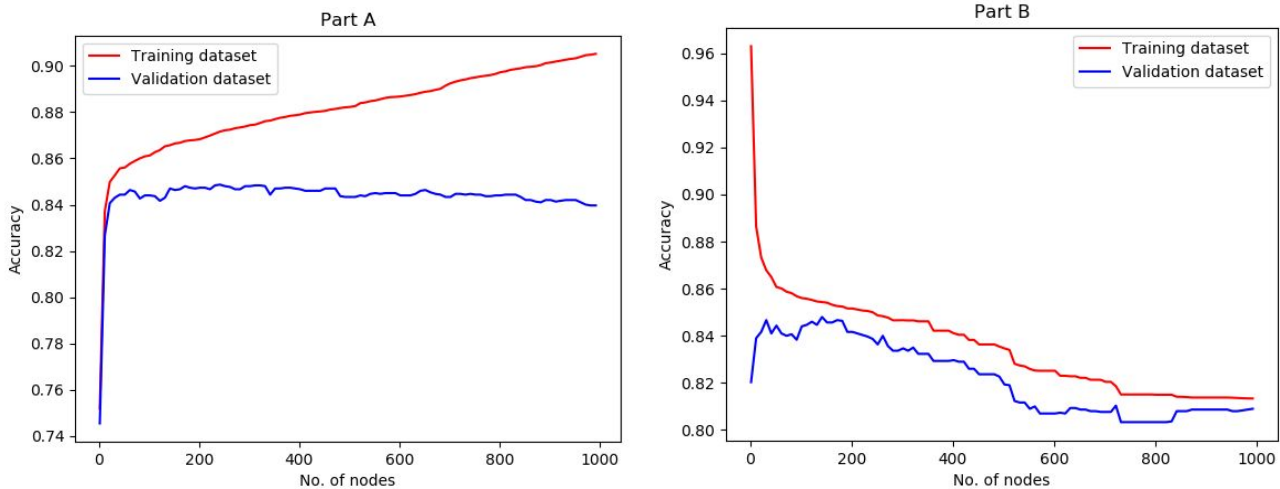
C	gamma( $\gamma$ )	Accuracy	Accuracy(PCA)	Graph
---	-------------------	----------	---------------	-------



The following are the eigenfaces corresponding the top five principal components. The top 5 values are scaled such that each pixel has at max 255 value. The following shows same alphabets without any compression and the next showing image by selecting top 5 eigenvalues/eigenvectors/feature.



## Decision Tree:



## K-means Clustering:

Initialization	Number of clusters	Purity	Entropy
Random (avg of 5)	46	0.30452685421994885	-5559.013399712942
	100	0.3949360613810742	-2285.1790750009145
	200	0.49475703324808185	-1025.540770432073
	255	0.5214450127877238	-770.951029748004

	300	0.553222506393862	-637.3225207958767
	340	0.5665728900255754	-549.5278314058057
Kmeans++	46	0.3105987236452467	-5558.98566432589
	200	0.52585527812469724	-1024.540770432073
	340	0.57134975621863149	-548.225064052506

PCA:

Number of eigenvalues	Number of clusters	Accuracy
10	2	0.5716666666666667
	5	0.6033333333333334
	10	0.7916666666666666
50	2	0.5716666666666667
	5	0.6383333333333333
	10	0.7566666666666667
100	2	0.5616666666666666
	5	0.6633333333333333
	10	0.7616666666666667
200	2	0.575
	5	0.6633333333333333
	10	0.7616666666666667
All	2	0.5633333333333334
	5	0.665
	10	0.765

<https://stats.stackexchange.com/questions/260917/stopping-condition-of-k-means>

The above link was used for various stopping criteria. The stopping criteria used were:

- Fixed number of iterations
- Maximum shift in any cluster center should be within a threshold