

Assignment 4 and 5 Report

Ayush Chaurasia

2016MT10617

Part 1: Support Vector Machines

a. Linear kernel

- a. $c=0.01$ – train acc. =97.64% ,test acc. =97.16%
- b. $c=0.1$ – train acc. =60.82% ,test acc. =59.83%
- c. $c=1$ – train acc. =99.82% ,test acc. =95.66%
- d. $c=10$ – train acc. =100% ,test acc. =95.5%
- e. $c=100$ – train acc. =100% ,test acc. =95.5%

b. RBF kernel

- a. $c=0.01$, $\gamma=0.01$ – train acc. =50% ,test acc. =50%
- b. $c=0.1$, $\gamma=0.1$ – train acc. =59.29% ,test acc. =60.5%
- c. $c=1$, $\gamma=1$ – train acc. =100% ,test acc. =65.1%
- d. $c=10$, $\gamma=10$ – train acc. =100% ,test acc. =50%
- e. $c=100$, $\gamma=100$ – train acc. =100% ,test acc. =50%
- f. $c=1$, $\gamma=10$ – train acc. =100% ,test acc. =50%

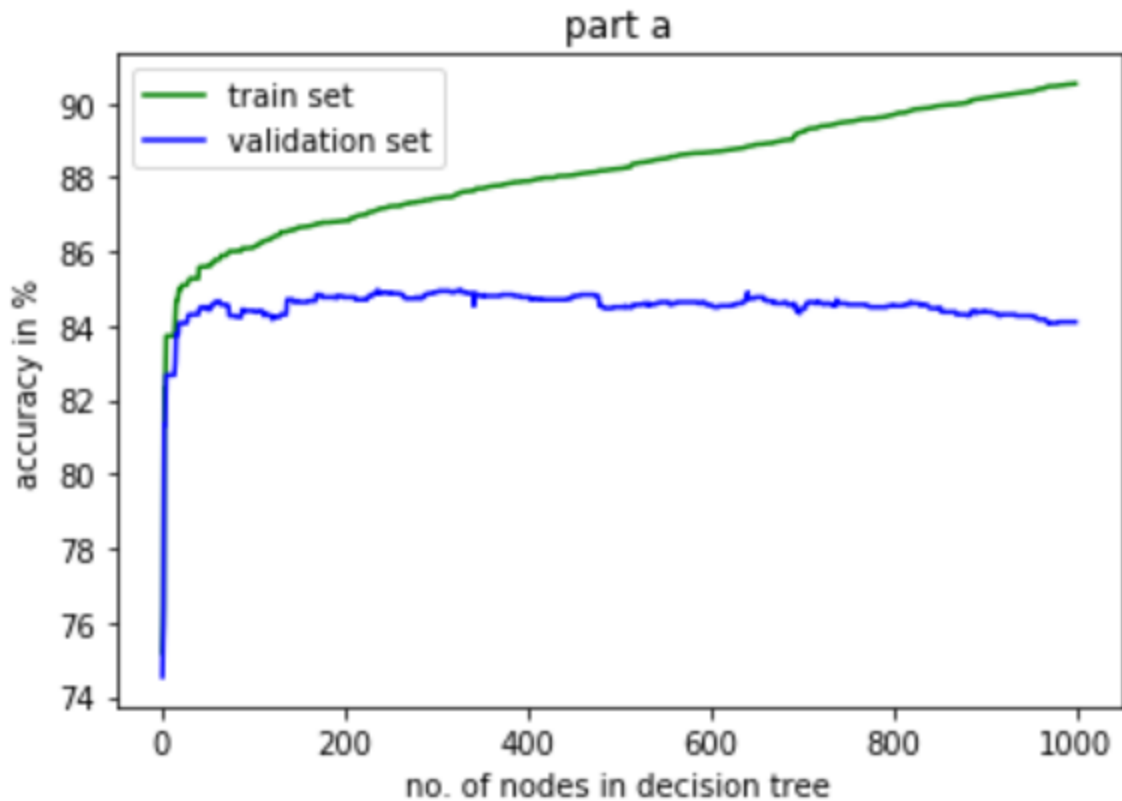
- g. $c=0.1$, $\gamma=10$ – train acc. = 50% ,test acc. = 50%
- h. $c=10$, $\gamma=1$ – train acc. = 100% ,test acc. = 65%
- i. $c=10$, $\gamma=0.1$ – train acc. = 54% ,test acc. = 56%
- j. $c=100$, $\gamma=0.1$ – train acc. = 54.79% ,test acc. = 56.83%

c. PCA accuracies

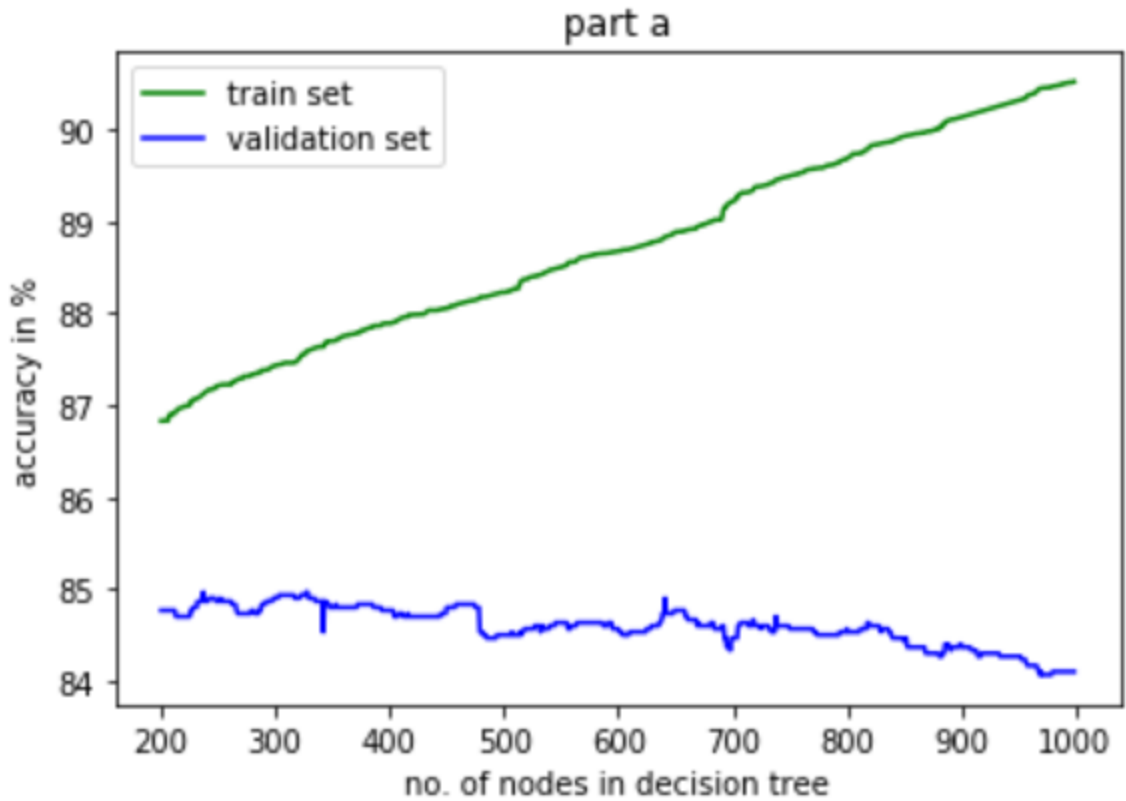
- a. $c=0.1$, $\gamma=0.1$ – test acc. = 50%
- b. $c=0.01$, $\gamma=0.01$ – test acc. = 50%
- c. $c=1$, $\gamma=1$ – test acc. = 71.5%
- d. $c=10$, $\gamma=10$ – test acc. = 71.3%
- e. $c=100$, $\gamma=100$ – test acc. = 71%
- f. $c=1$, $\gamma=0.01$ – test acc. = 71.5%
- g. $c=0.01$, $\gamma=0.01$ – test acc. = 71%

Part 2: Decision trees

- a. plot of accuracy vs number of nodes in tree from 2-1000. It is clear from here that as the number of nodes increase the tree starts overfitting the training data. And hence the train accuracy increases but the validation accuracy decreases.



b. While pruning this plot is observed. As the number of nodes decrease the validation accuracy increases and the train accuracy decreases.



Part 3: K-means

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

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