Assignment - 3

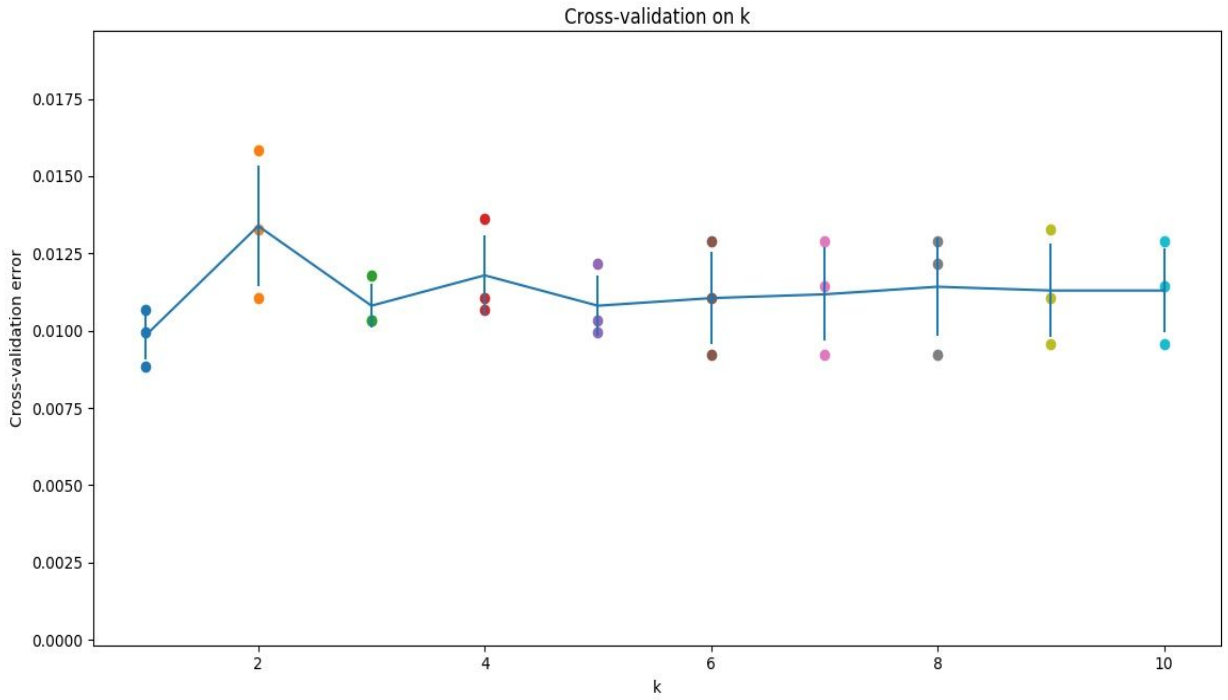
Group : **M.Krishna Chaithanya (IS201401032)**
I.Sandeep Naga Varma (IS201401023)
D.Harish (IS201401019)

Error Matrix for KNN problem :

	Fold - 1	Fold - 2	Fold - 3
k = 1	0.008843036109064112	0.010681399631675874	0.009948415622697125
k = 2	0.01105379513633014	0.013259668508287293	0.015843773028739867
k = 3	0.010316875460574797	0.011786372007366482	0.010316875460574797
k = 4	0.010685335298452468	0.011049723756906077	0.013633014001473839
k = 5	0.009948415622697125	0.012154696132596685	0.010316875460574797
k = 6	0.009211495946941784	0.01289134438305709	0.01105379513633014
k = 7	0.009211495946941784	0.01289134438305709	0.011422254974207811
k = 8	0.009211495946941784	0.01289134438305709	0.012159174649963155
k = 9	0.009579955784819455	0.013259668508287293	0.01105379513633014
k = 10	0.009579955784819455	0.01289134438305709	0.011422254974207811

Therefore, from the above table, optimal value of k = 1 since the average error is minimum for k = 1.

Note : Here, error = (No.of misclassified patterns/ Total no.of patterns)



The mean and standard deviation for the above errors is as follows :

	K = 1	K = 2	K = 3	K = 4	K = 5
mean	0.0098243	0.01338575	0.010806708	0.01178936	0.010806662
std	0.0007556	0.00195753	0.000692727	0.00131212	0.000964999

	K = 6	K = 7	K = 8	K = 9	K = 10
mean	0.0110522	0.01117503	0.011420672	0.01129781	0.011297852
std	0.0015023	0.00151243	0.001590463	0.00151211	0.001354728

With the optimal value of k = 1,

Test accuracy for ‘datatest.txt’ = 93.6585365854

Test accuracy for ‘datatest2.txt’ = 95.0266611977

Error Matrix for Modified-KNN problem :

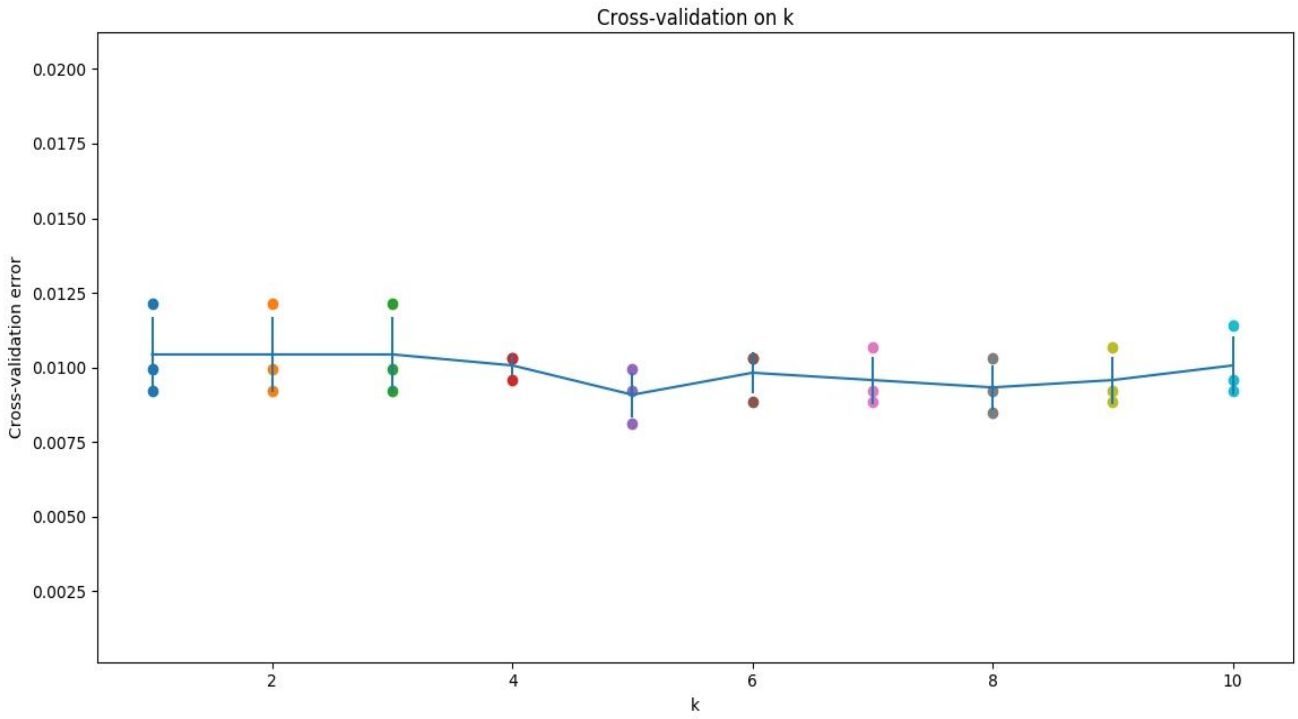
	Fold - 1	Fold - 2	Fold - 3
k = 1	0.012159174649963155	0.009208103130755065	0.009948415622697125
k = 2	0.012159174649963155	0.009208103130755065	0.009948415622697125
k = 3	0.012159174649963155	0.009208103130755065	0.009948415622697125
k = 4	0.010316875460574797	0.009576427255985266	0.010316875460574797
k = 5	0.008106116433308769	0.009208103130755065	0.009948415622697125
k = 6	0.008843036109064112	0.010313075506445672	0.010316875460574797
k = 7	0.008843036109064112	0.010681399631675874	0.009211495946941784
k = 8	0.00847457627118644	0.010313075506445672	0.009211495946941784
k = 9	0.008843036109064112	0.010681399631675874	0.009211495946941784
k = 10	0.009579955784819455	0.01141804788213628	0.009211495946941784

Therefore, from the above table, optimal value of k = 5 since the average error is minimum for k = 5.

The mean and standard deviation for the above errors is as follows :

	K = 1	K = 2	K = 3	K = 4	K = 5
mean	0.0104386	0.01043856	0.010438564	0.01007006	0.009087545
std	0.0012536	0.00125363	0.001253632	0.00034905	0.000756931

	K = 6	K = 7	K = 8	K = 9	K = 10
mean	0.0098243	0.00957864	0.009333049	0.00957864	0.010069833
std	0.0006939	0.00079414	0.000755469	0.00079414	0.000965126



With the optimal value of k = 5,

Test accuracy for ‘datatest.txt’ = 93.4333958724

Test accuracy for ‘datatest2.txt’ = 95.3240360952