ASSIGNMENT 5 REPORT

GROUP: Neha Pulipati, Rakesh Katkam, Saiteja Sirikonda

QUESTION 1(K-MEANS CLUSTERING):

For the Seeds dataset,
number of Samples – 210

#clusters	Sum of Squared Errors (five runs)				
3	587.3186	588.7820	587.3186	588.7820	588.7820
5	386.1576	398.8630	408.4678	386.1073	385.8605
7	302.8154	279.8279	314.8005	302.9313	319.1958

As shown in the table above the sum of squared error decreases with the increase in the number of clusters. The average number of iterations taken to converge is as follows-

For k=3, #iterations = 7
For k=5, #iterations = 14
For k=7, #iterations = 13

QUESTION 2(ENSEMBLE)

For the 'Handwriting Digits' data,
Training Samples – 500; Testing Samples – 3251

Classification method	Accuracies				
KNN	92.43%	92.43%	92.43%	92.43%	92.43%
SVM	88.31%	88.31%	88.31%	88.31%	88.31%
Artificial neural network	92.16%	89.85%	91.36%	91.94%	92.83%
Ensemble	94.03%	93.97%	94.19%	94.37%	94.31%

Because of the majority votes taken in the ensemble, the accuracy of ensemble is slightly higher even though the accuracies of individual classifiers are lower.