Bagging Application Methodology

1. Bagging is applied on Perceptron learning algorithm that uses gradient descent.
2. Training data is sampled and the model is trained on it which is then used to generate predictions on the test data.
3. Based on the number of bags specified this process is repeated and the class labels for each test data are stored.
4. Majority count on the test data are taken with respect to results stored by bagging and final class label on the test data are assigned accordingly.

Experiments:

The experiments are done manually specifying k and training & test file along with learning rate and number of iterations at each step.

Results:

General Perceptron Learning Algorithm (HW3)

```
#---------------Accuracy------------------------------------------------------

# Learning Rate = 0.5 Iterations= 1000

                               No Bagging    K=50      K=100     K=200

# Training file   Testing file  Accuracy    Accuracy  Accuracy  Accuracy

# train.dat      train.dat      93.33%     93.33%    93.33%    93.33%

# train.dat      test.dat       85.71%     85.71%    85.71%    85.71%

# train2.dat     train2.dat     85.75%     86.00%    85.75%    85.75%

# train2.dat     test2.dat      80.00%     81.00%    79.00%    79.00%

# train3.dat     train3.dat     100.00%   100.00%   100.00%  100.00%

# train3.dat     test3.dat      100.00%   100.00%   100.00%  100.00%

# train4.dat     train4.dat      87.00%    89.00%    91.00%    90.00%

# train4.dat     test4.dat       70.00%    68.50%   68.50%    69.25%

#------------------------------------------------------------------------------
```

Discussion:

1. As seen from the above accuracy readings it can be clearly understood that Bagging rarely reduces the accuracy.
2. Some datasets do show an increase in accuracy such as train2 on training & test data, train4 on training data. While train 4 gives reduction in accuracy on the test data. (Due to small dataset size)
3. Other datasets such as train3(small size) maintain their perfect accuracies.
4. Even dataset train maintains it’s accuracy over training and test set.
5. It can be inferred that we can apply bagging and then select k which gives best accuracy for any dataset by mapping k versus accuracy plots.
6. Hence, bagging is an effective procedure to increase the accuracy of an unstable algorithm.