1. Assume that data is linearly separable by a hyperplane through the origin. Run two, three and four passes of perceptron, voted perceptron, and averaged perceptron on the training dataset to find classifiers that separate the two classes. What are the training errors and the test errors of perceptron, voted perceptron and averaged perceptron after two, three and four passes?

After two passes:

	Perceptron	Voted Perceptron	Averaged Perceptron
Training Error	0.0404	0.0404	0.0541
Test Error	0.0584	0.0610	0.0822

After three passes:

	Perceptron	Voted Perceptron	Averaged Perceptron
Training Error	0.0211	0.0303	0.0376
Test Error	0.0451	0.0451	0.0610

After four passes:

	Perceptron	Voted Perceptron	Averaged Perceptron
Training Error	0.0183	0.0248	0.0339
Test Error	0.0477	0.0451	0.0504

2. Find the three coordinates in w_{avg} with the highest and lowest values. What are the words (from pa3dictionary.txt) that correspond to these coordinates? The three highest coordinates are those words whose presence indicates the positive class most strongly, and the three lowest coordinates are those words whose presence indicates the negative class most strongly.

Note: we map label 1 to +1 and label 2 to −1.

The words that correspond to the three highest coordinates (in decreasing order) are “file”, “program”, and “line” with “file” having the most positive coordinate value.

The words that correspond to the three lowest coordinates (in increasing order) are “he”, “team”, and “game” with “he” having the most negative coordinate value.

3. Write down the confusion matrix for the one-vs-all classifier on the training data in pa3train.txt based on the test data in pa3test.txt. Looking at the confusion matrix, what are the i and j in the following statements?

The confusion matrix:

```
[[0.71891892 0.00520833 0.03428571 0.02173913 0.         0.        ]
 [0.01081081 0.65625    0.03428571 0.02717391 0.01282051 0.01851852]
 [0.         0.015625   0.37142857 0.         0.         0.02777778]
 [0.01621622 0.00520833 0.         0.69021739 0.         0.        ]
 [0.01621622 0.03125    0.07428571 0.00543478 0.80128205 0.12037037]
 [0.00540541 0.01041667 0.03428571 0.         0.07051282 0.49074074]
 [0.23243243 0.27604167 0.45142857 0.25543478 0.11538462 0.34259259]]
```

(a). The perceptron classifier has the highest accuracy for examples that belong to class i.

According to the matrix, the perceptron classifier has the highest accuracy for examples that belong to class 5.

(b). The perceptron classifier has the least accuracy for examples that belong to class i.

According to the matrix, the perceptron classifier has the least accuracy for examples that belong to class 3.

(c). The perceptron classifier most often mistakenly classifies an example in class j as belonging to class i, for i, j ∈ {1, 2, 3, 4, 5, 6} (i.e., excluding Don't Know).

According to the matrix, the perceptron classifier **most** often mistakenly classifies an example in class 6 as belonging to class 5.

For examples in other classes:
An example in class 1 is often mistakenly classified as belonging to either class 4 or class 5.
An example in class 2 is often mistakenly classified as belonging to class 5.
An example in class 3 is often mistakenly classified as belonging to class 5.
An example in class 4 is often mistakenly classified as belonging to class 2.
An example in class 5 is often mistakenly classified as belonging to class 6.