1.Implementation

I've chosen to solve this homework in Python. I'm also using the libraries NumPy and MathPlotLib. I've decided to use NumPy because it makes the calculations so much easy to understand and read (the fact that you can do add a vector with a constant and behind the scene it adds the constant to all the elements of the vector is so helpful). I've used the MathPlotLib in order to create the graphics where I can plot the Dev set accuracy during each of the epochs.

I have created 4 perceptron classes, one for each type required by the homework: Simple Perceptron, Dynamic Learning Rate Perceptron, Margin Perceptron and Averaged Perceptron. Each of these classes has 3 methods:

- ***predict*** - static method for doing a prediction; it receives a record and a weight vector and returns 1 or -1;
- ***train_one_epoch*** - this method was created only because we need to report some information after each epoch (development set accuracy after each epoch and the number of updates the learning algorhithm performs); depending on the perceptron type this method returns an array having varible length; but for all the perceptron types it will return at least the number of updates, and the weight vector determined during that epoch.
- ***train*** - returns the weight vector; this method is used when we are determing the best hyper-parameter; it received the feature vectors, the lables and the number of epochs that it needs to run (for determining the hyper-parameters I used 10 epochs); behind the scenes, it uses ***train_one_epoch*** method.

The heavy lift is done by the ***CrossValidatorTester*** class. It receives:

- A perceptron class type (SimplePerceptron, DynamicLearningRatePerceptron, MarginPerceptron and AveragedPerceptron);
- A dictionary containing the hyper parameters. For the MarginPerceptron this dictionary looks like: `{'Learning rate': [1, 0.1, 0.01], 'Margin': [1, 0.1, 0.01]}`. When determining the best hyper-parameter combination, the ***CrossValidatorTester*** class will do a cartesian product in order to get all the possible combinations.
- An array containing the paths to the cross validation training set.
- The path to the development set.
- The path to the training set.

CrossValidatorTester class is using the ***DataSetLoader*** class in order to read the data from the files. If we have a row like:

```
-1 1:0.18257 2:0.18257 4:0.18257 6:0.18257 7:0.18257
```

the label is -1, and the other values are features values. I decided to add the bias value inside the vector so the final vector returned by ***DataSetLoader*** will look like:

[1, 0.18257, 0.18257, 0, 0.18257, 0, 0.18257, 0.18257].

DataSetLoader is returning a list of vectors and a list of labels.

CrossValidatorTester class is following the steps that were listed in the homework description document:

- Determines the best hyper-parameter using the cross validation data sets;
- Starts training the perceptron using the best hyper-parameters;
- For each epoch, is using the weights vector trained and checks the accuracy of it on the development set;
- After all the epochs were executed, it will choose the weights vector that had the best accuracy and calculates the accuracy on the testing set.

2. Majority Baseline

Majority Baseline accuracy for the development set: **54.92%**.
Majority Baseline accuracy for the test set: **57.31%**.

3. Experiments

3.1 Simple perceptron

A. The best hyper-parameters

Learning Rate	Error rate when **training00 .data** was used as test set	Error rate when **training01 .data** was used as test set	Error rate when **training02 .data** was used as test set	Error rate when **training03 .data** was used as test set	Error rate when **training04 .data** was used as test set	Average error rate
1.00	8.50%	8.50%	5.85%	8.62%	7.53%	7.80%
0.10	8.81%	5.91%	6.39%	9.17%	6.45%	7.35%
0.01	23.28%	8.08%	6.15%	6.33%	8.92%	10.55%

Best Hyper-parameter: Learning rate: **0.10**

B. The cross-validation accuracy for the best hyperparameter

Cross validation accuracy: **92.65%** (100% - 7.35%)

C. The total number of updates the learning algorithm performs on the training set

Updates performed during training: **12767**

D. Development set accuracy

Epoch 1	Epoch 2	Epoch 3	Epoch 4	Epoch 5	Epoch 6	Epoch 7	Epoch 8	Epoch 9	Epoch 10
9.19%	7.74%	7.96%	9.12%	8.83%	11.43 %	7.74%	10.27 %	13.68 %	8.76%

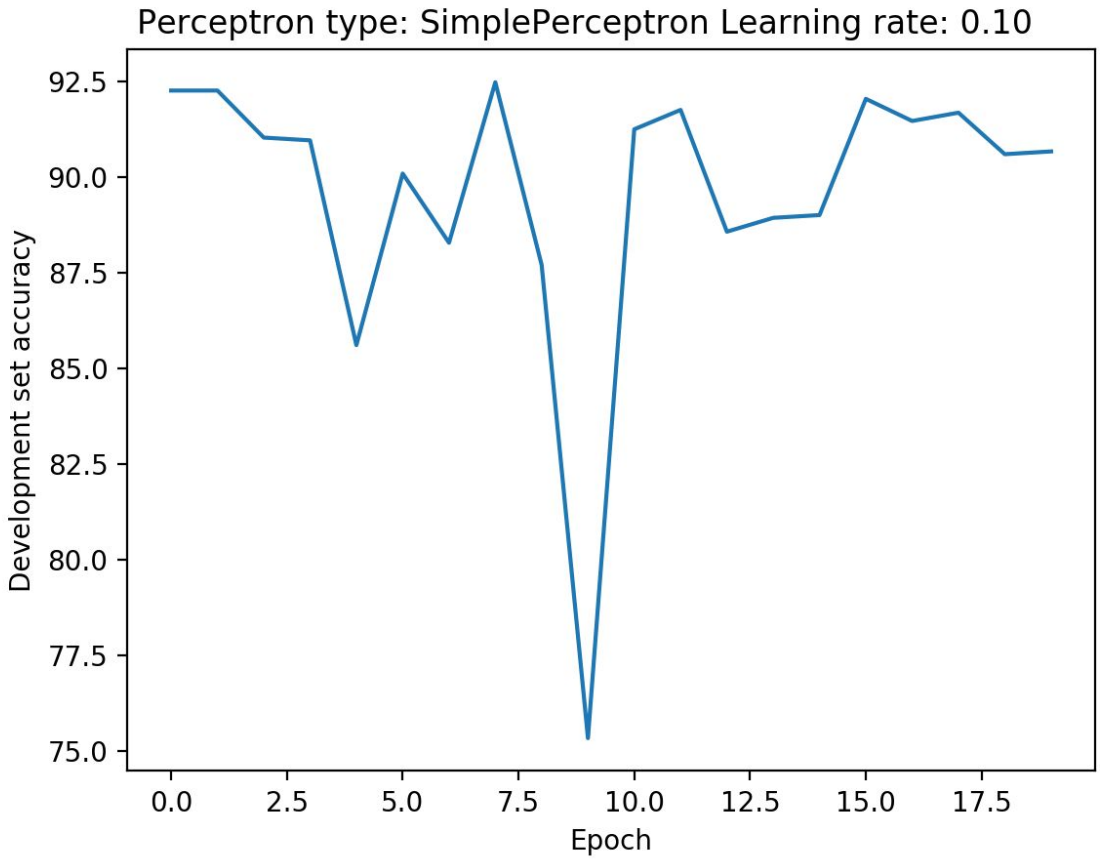
Epoch 11	Epoch 12	Epoch 13	Epoch 14	Epoch 15	Epoch 16	Epoch 17	Epoch 18	Epoch 19	Epoch 20
9.62%	8.90%	8.54%	18.81 %	9.99%	7.74%	9.84%	7.81%	9.33%	10.85 %

Minimum error rate: **7.74%**
Epoch: **1**
Development set accuracy: **92.26%**

E. Test set accuracy

Testing data error rate: **8.25%**
Test set accuracy: **91.75%**

F. Plot



3.2 Dynamic learning rate perceptron

A. The best hyper-parameters

Learning Rate	Error rate when **training00 .data** was used as test set	Error rate when **training01 .data** was used as test set	Error rate when **training02 .data** was used as test set	Error rate when **training03 .data** was used as test set	Error rate when **training04 .data** was used as test set	Average error rate
1.00	10.86%	6.03%	9.89%	6.63%	11.15%	8.91%
0.10	18.40%	6.27%	11.34%	6.57%	7.41%	10.00%
0.01	8.26%	6.15%	6.09%	8.02%	7.05%	7.11%

Best Hyper-parameter: Learning rate: **0.01**

B. The cross-validation accuracy for the best hyperparameter

Cross validation accuracy: **92.89%** (100% - 7.11%)

C. The total number of updates the learning algorithm performs on the training set

Updates performed during training: **12746**

D. Development set accuracy

Epoch 1	Epoch 2	Epoch 3	Epoch 4	Epoch 5	Epoch 6	Epoch 7	Epoch 8	Epoch 9	Epoch 10
10.71 %	12.01 %	7.96%	7.45%	13.24 %	11.79 %	7.67%	11.58 %	11.22 %	7.81%

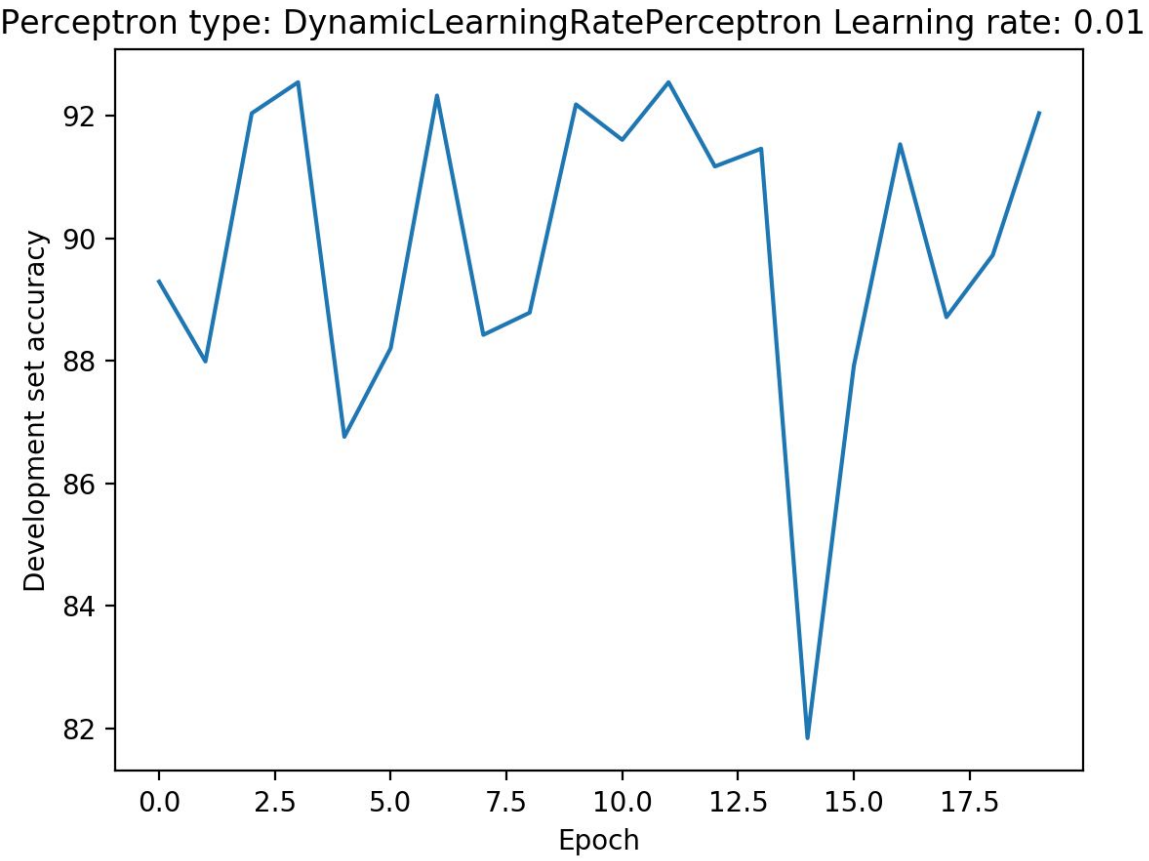
Epoch 11	Epoch 12	Epoch 13	Epoch 14	Epoch 15	Epoch 16	Epoch 17	Epoch 18	Epoch 19	Epoch 20
8.39%	7.45%	8.83%	8.54%	18.16 %	12.08 %	8.47%	11.29 %	10.27 %	7.96%

Minimum error rate: **7.45%**
Epoch: **4**
Development set accuracy: **92.55%**

E. Test set accuracy

Testing data error rate: **6.80%**
Test set accuracy: **93.20%**

F. Plot



3.3 Margin perceptron

A. The best hyper-parameters

Learning Rate	Margin	Error rate when **training00.data** was used as test set	Error rate when **training01.data** was used as test set	Error rate when **training02.data** was used as test set	Error rate when **training03.data** was used as test set	Error rate when **training04.data** was used as test set	Average error rate
1.00	1.00	8.20%	6.15%	5.37%	6.33%	6.39	6.49%
1.00	0.10	8.75%	5.91%	6.21%	7.36%	7.23	7.09%
1.00	0.01	11.28%	7.72%	6.69%	6.63%	7.78	8.02%
0.10	1.00	7.66%	5.67%	5.91%	6.63%	6.51	6.48%
0.10	0.10	8.20%	5.49%	5.31%	6.09%	6.69	6.36%
0.10	0.01	8.38%	8.56%	6.82%	6.82%	7.47	7.61%
0.01	1.00	8.26%	7.48%	6.94%	8.20%	7.05	7.59%
0.01	0.10	7.72%	5.25%	5.55%	6.63%	6.75	6.38%
0.01	0.01	8.44%	5.55%	5.37%	6.57%	7.41	6.67%

Best Hyper-parameter: Learning rate: **0.10** Margin: **0.10**

B. The cross-validation accuracy for the best hyperparameter

Cross validation accuracy: **93.64%** (100% - 6.36%)

C. The total number of updates the learning algorithm performs on the training set

Updates performed during training: **20511**

D. Development set accuracy

Epoch 1	Epoch 2	Epoch 3	Epoch 4	Epoch 5	Epoch 6	Epoch 7	Epoch 8	Epoch 9	Epoch 10

7.45%	8.47%	7.02%	7.16%	8.10%	8.10%	7.96%	7.60%	8.54%	7.67%

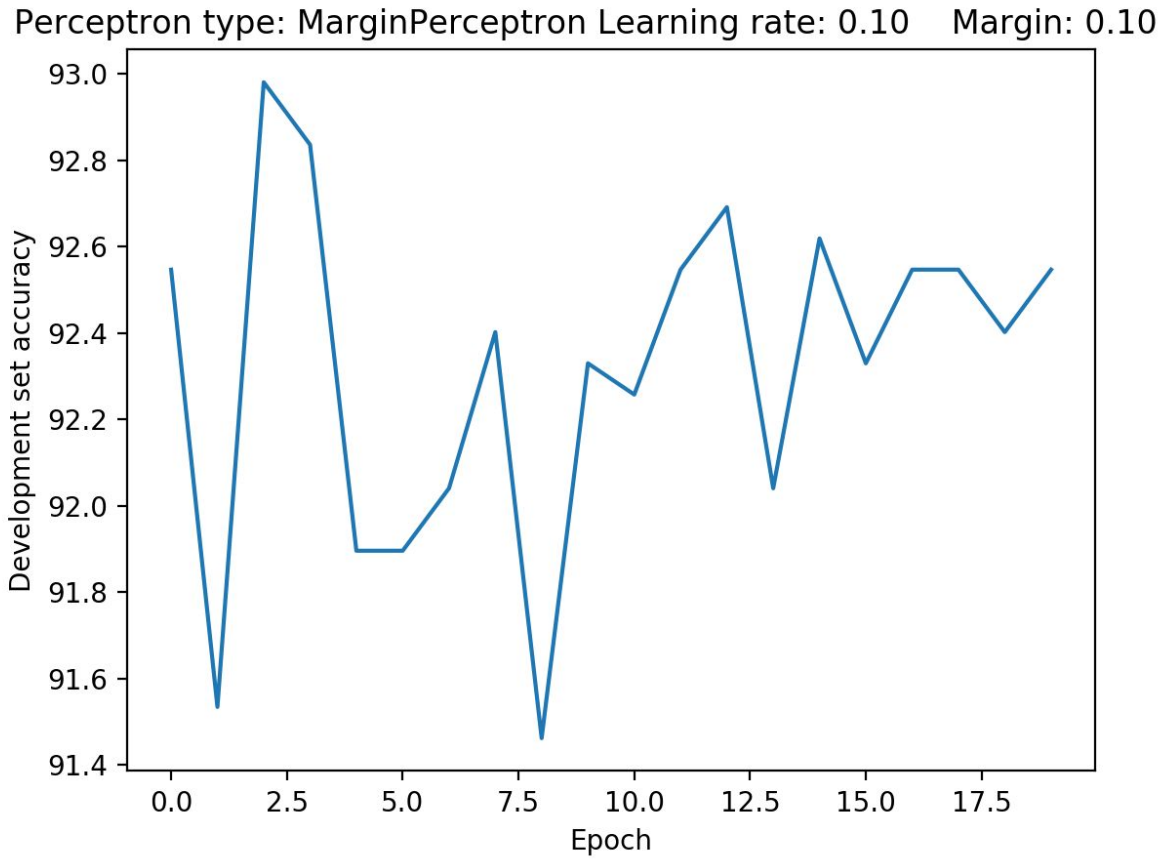
Epoch 11	Epoch 12	Epoch 13	Epoch 14	Epoch 15	Epoch 16	Epoch 17	Epoch 18	Epoch 19	Epoch 20
7.74%	7.45%	7.31%	7.96%	7.38%	7.67%	7.45%	7.45%	7.60%	7.45

Minimum error rate: **7.02%**
Epoch: **3**
Development set accuracy: **92.98%**

E. Test set accuracy

Testing data error rate: **6.95%**
Test set accuracy: **93.05%**

F. Plot



3.4 Averaged perceptron

A. The best hyper-parameters

Learning Rate	Error rate when **training00 .data** was used as test set	Error rate when **training01 .data** was used as test set	Error rate when **training02 .data** was used as test set	Error rate when **training03 .data** was used as test set	Error rate when **training04 .data** was used as test set	Average error rate
1.00	8.26%	5.55%	5.19%	6.51%	711.00%	6.52%
0.10	8.14%	5.73%	5.49%	6.33%	687.00%	6.51%
0.01	8.14%	5.37%	5.13%	6.45%	705.00%	6.43%

Best Hyper-parameter: Learning rate: **0.01**

B. The cross-validation accuracy for the best hyperparameter

Cross validation accuracy: **93.57%** (100% - 6.43%)

C. The total number of updates the learning algorithm performs on the training set

Updates performed during training: **14672**

D. Development set accuracy

Epoch 1	Epoch 2	Epoch 3	Epoch 4	Epoch 5	Epoch 6	Epoch 7	Epoch 8	Epoch 9	Epoch 10
7.45%	7.74%	7.67%	7.74%	7.60%	7.67%	7.60%	7.45%	7.45%	7.38%

Epoch 11	Epoch 12	Epoch 13	Epoch 14	Epoch 15	Epoch 16	Epoch 17	Epoch 18	Epoch 19	Epoch 20
7.38%	7.38%	7.38%	7.38%	7.38%	7.38%	7.38%	7.38%	7.38%	7.38

Minimum error rate: **7.38%**

Epoch: **10**
Development set accuracy: **92.62%**

E. Test set accuracy

Testing data error rate: **7.16%**
Test set accuracy: **92.84%**

F. Plot