5-a)

The result in the tables are based on my experiments. The result can be modified by changing the values of learning rate and number of iterations.

IRIS Data Set				
Classifiers	**Learning Rate**	**Number of Iterations**	**Accuracy %**	**Error Rate %**
Percpetron	**.01**	**10**	**100**	**0**
AdaLine	**0.00001**	**100**	**100**	**0**
SGD	**.01**	**100**	**100**	**0**
OneVsAll with SGD	**.01**	**100**	**Max Accuracy 100**	**0**

Phishing Data Set				
Classifiers	**Learning Rate**	**Number of Iteration**	**Accuracy %**	**Error Rate %**
Percpetron	**.0001**	**100**	**93**	**7**
AdaLine	**.001**	**100**	**93**	**7**
SGD	**.001**	**50**	**93**	**7**
OneVsAll with SGD	**.01**	**50**	**Max Accuracy 93**	**7**

5-b)

Applying perceptron, Adaline and SGD on IRIS dataset:

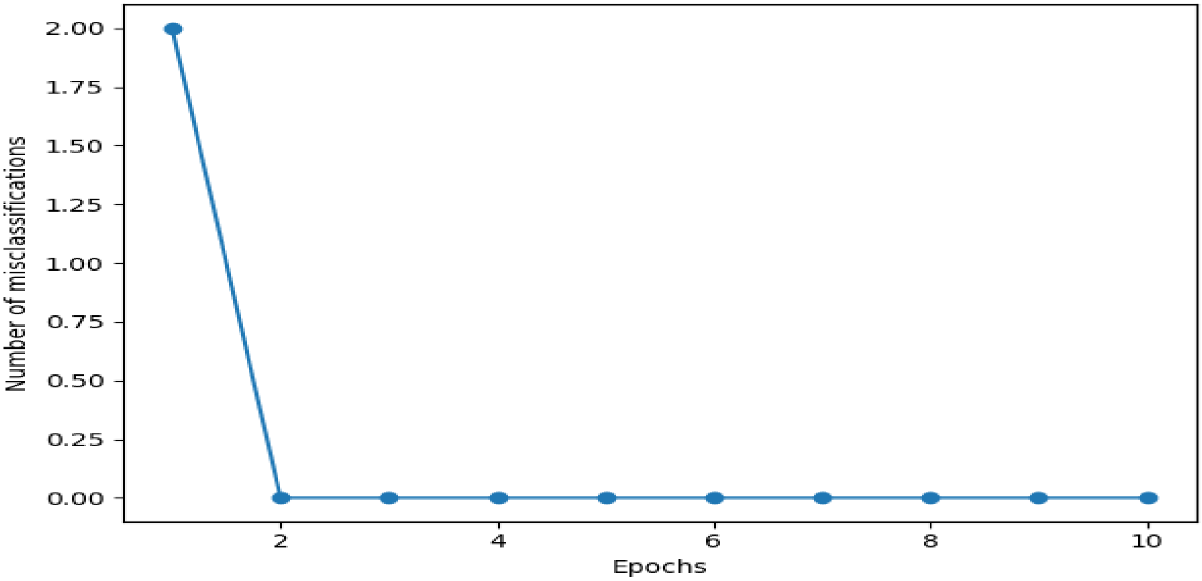

Figure 1: Number of misclassifications in each iteration

by applying perceptron classifier on IRIS data set

The cost of each iteration by using perceptron classifier

[2.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0]

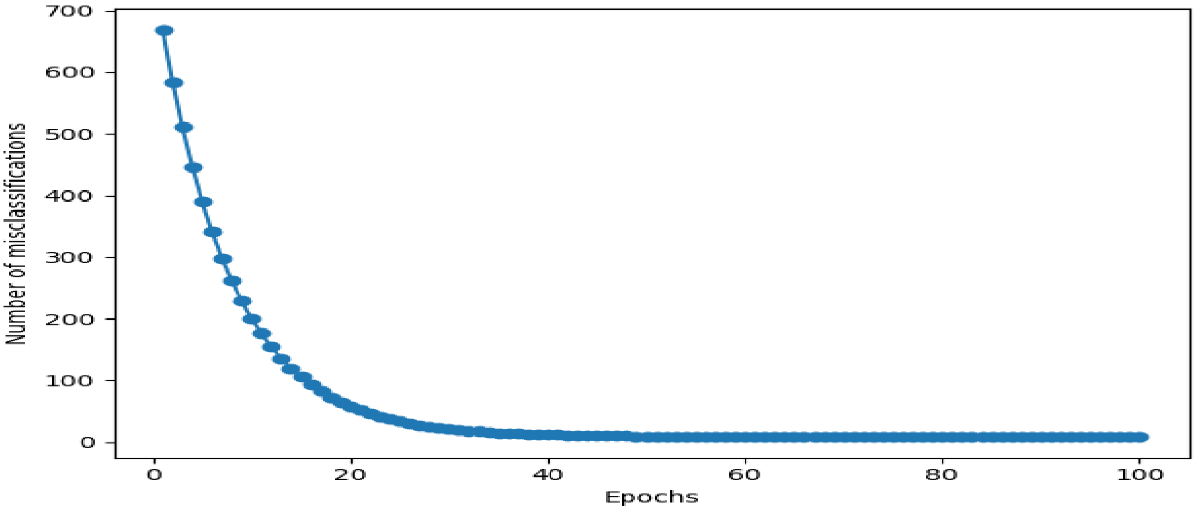

Figure 2: Number of misclassifications in each iteration

by applying AdaLine classifier on IRIS data set

The cost of each iteration by using Adaline classifier

```
 [669.271758883794,         584.447762307731,         510.52063024565336,         446.0902029192458,
389.9362312556302,        340.99525961781615,        298.3404789429188,        261.16416861148133,
228.7623944140241,        200.5216727220766,         175.9073482201405,        154.45346501796595,
135.75393925419533,        119.45486595881836,        105.24781443009897,        92.86398510876535,
82.06911725307741,         72.65905094210031,         64.45585933057758,         57.30447788205286,
51.06976672200536,         45.63395045810021,         40.89438696566421,         36.76162286865239,
33.15769887773547,        30.014672880612572,        27.273332804908343,        24.882074869201944,
22.795925970967488,        20.975691690838893,        19.387213772378477,        18.000723010511138,
16.79027528927651,        15.733260084792493,        14.80997212215584,        14.003238071440157,
13.29809121065134,        12.681487892221213,        12.142060441576644,        11.669901806516597,
11.256377877640048,         10.8939639242879,         10.57610204732363,        10.297076948235599,
10.051907661041643,         9.836253195887133,         9.646330306781207,         9.478841825603809,
9.330914204691759,         9.200043084765497,         9.084045856996442,         8.981020320516587,
8.88930865214863,         8.807466005773325,         8.73423314645811,         8.668512600908205,
8.609347872417395,          8.555905326551652,          8.507458404395086,          8.4633738642826 7,
8.423099791373447,         8.386155147908958,         8.352120666189645,         8.320630911739391,
8.291367366297095,          8.264052399594709,          8.238444015718878,          8.21433127452754,
8.191530301381713,         8.169880809598054,         8.149243069741129,         8.129495268339625,
8.110531205988307,         8.092258291226914,         8.074595792190815,         8.057473312911545,
8.040829465401245,          8.02461071236417,         8.008770358610882,          7.993267672067893,
7.978067117730581,         7.9631376900470725,         7.948452331085281,         7.933987423460667,
7.919722348418432,         7.905639100698297,         7.891721952885678,         7.877957162890617,
7.864332719012841,         7.850838117763453,         7.837464170234163,         7.824202833346005,
7.811047062780624,         7.797990684808136,         7.785028284583431,         7.772155108794905,
7.759366980821385, 7.746660226790062, 7.734031611134714, 7.721478280433537]
```

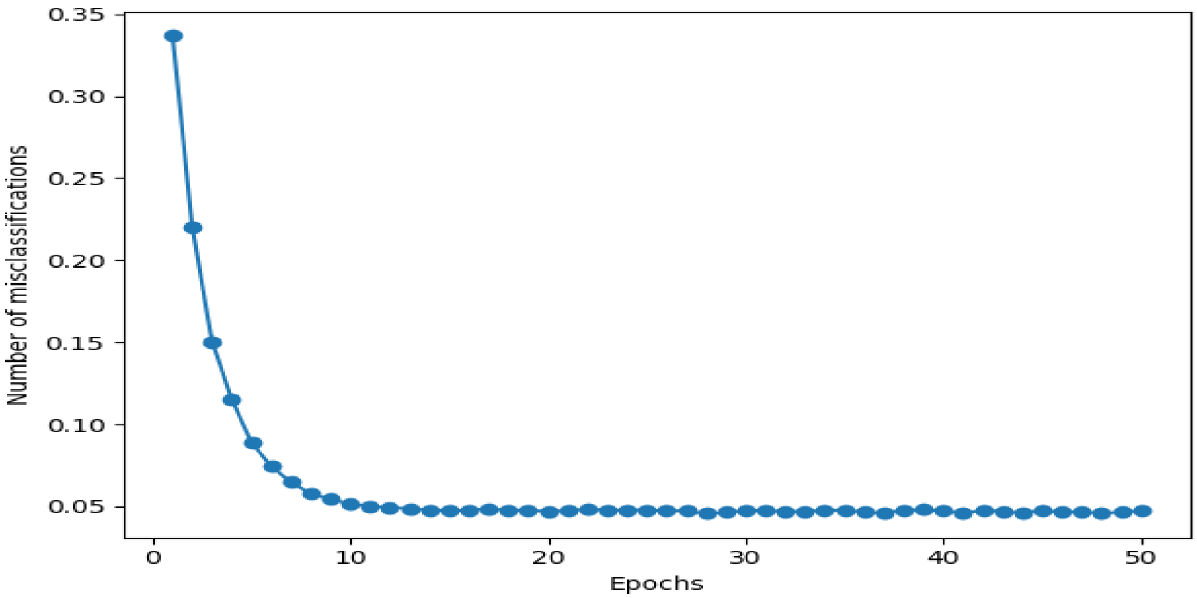

Figure 3: Number of misclassifications in each iteration

by applying SGD classifier on IRIS data set

The cost of each iteration by using SGD classifier

```
 [0.3369767987782166,                0.22030425999440512,                0.15045584621826436,
0.11523542728293923,                 0.08887530959008118,                0.07436177443732965,
0.06475056248307554,                 0.05774158334610318,                0.05440687738907124,
0.05127430614033048,                 0.05021197016596989,                0.04938277297916066,
0.04845034360543151,                0.047876225326993434,               0.047376571061800876,
0.047723944734017566,                 0.04860418158143559,                0.04761234409310669,
0.0476259244613643,                  0.04692751296718127,                0.04731909456885888,
0.048408764306554206,                 0.04767230255049808,                0.04782724388168933,
0.047805731958154775,                0.047331993365283785,               0.047282884116400005,
0.04560440977389131,                 0.04633581055393753,                0.04734067214881023,
0.04756318716325751,                 0.04667750892509006,               0.047038205995834984,
0.04737649783161046,                 0.04786627570093129,                0.04632985500587638,
0.046157600425647344,                0.04755697343244289,               0.047983414826286555,
0.04723445976824928,                 0.04589247590987908,                0.04747802533573104,
0.0469748196817522,                  0.04601054888588387,                0.04757368693308987,
0.046324195265644796,                 0.04693705242005904,                0.04570533010786647,
0.0466187647666242, 0.04717966184924072]
```

Applying perceptron, Adaline and SGD on Phishing dataset:

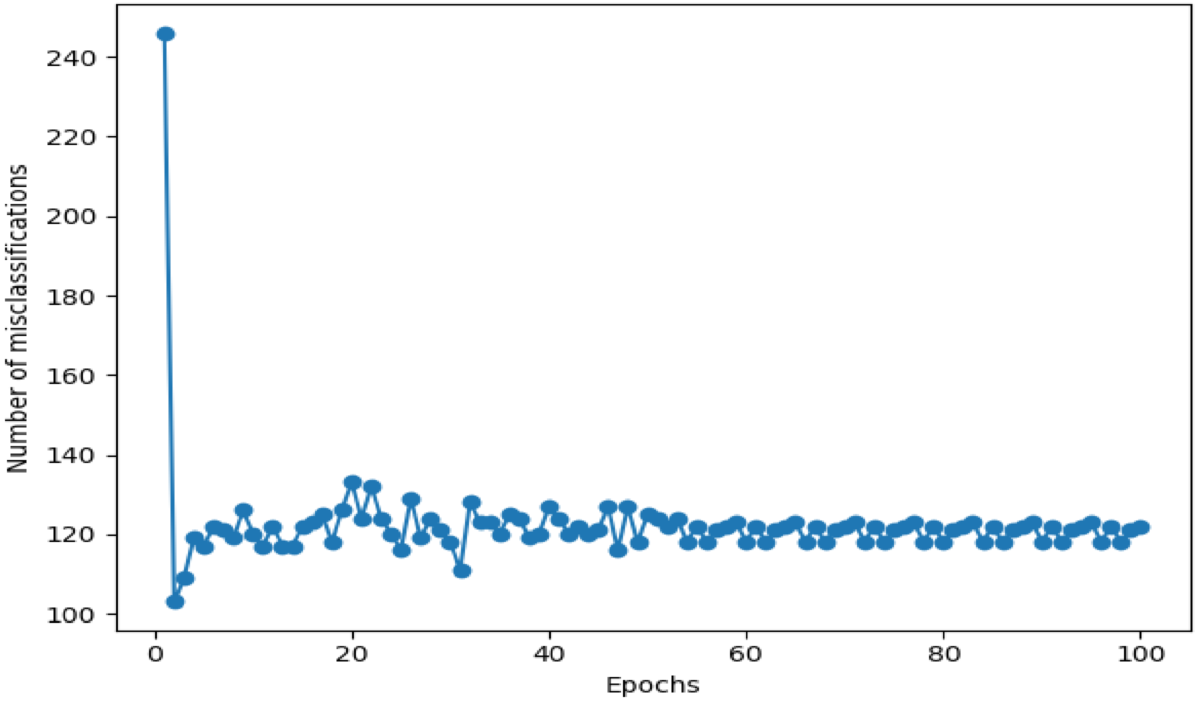

Figure 4: Number of misclassifications in each iteration

by applying perceptron classifier on Phishing data set

The cost of each iteration by using perceptron classifier

[246.0, 103.0, 109.0, 119.0, 117.0, 122.0, 121.0, 119.0, 126.0, 120.0, 117.0, 122.0, 117.0, 117.0, 122.0, 123.0, 125.0, 118.0, 126.0, 133.0, 124.0, 132.0, 124.0, 120.0, 116.0, 129.0, 119.0, 124.0, 121.0, 118.0, 111.0, 128.0, 123.0, 123.0, 120.0, 125.0, 124.0, 119.0, 120.0, 127.0, 124.0, 120.0, 122.0, 120.0, 121.0, 127.0, 116.0, 127.0, 118.0, 125.0, 124.0, 122.0, 124.0, 118.0, 122.0, 118.0, 121.0, 122.0, 123.0, 118.0, 122.0, 118.0, 121.0, 122.0, 123.0, 118.0, 122.0, 118.0, 121.0, 122.0, 123.0, 118.0, 122.0, 118.0, 121.0, 122.0, 123.0, 118.0, 122.0, 118.0, 121.0, 122.0, 123.0, 118.0, 122.0, 118.0, 121.0, 122.0, 123.0, 118.0, 122.0, 118.0, 121.0, 122.0, 123.0, 118.0, 122.0, 118.0, 121.0, 122.0]

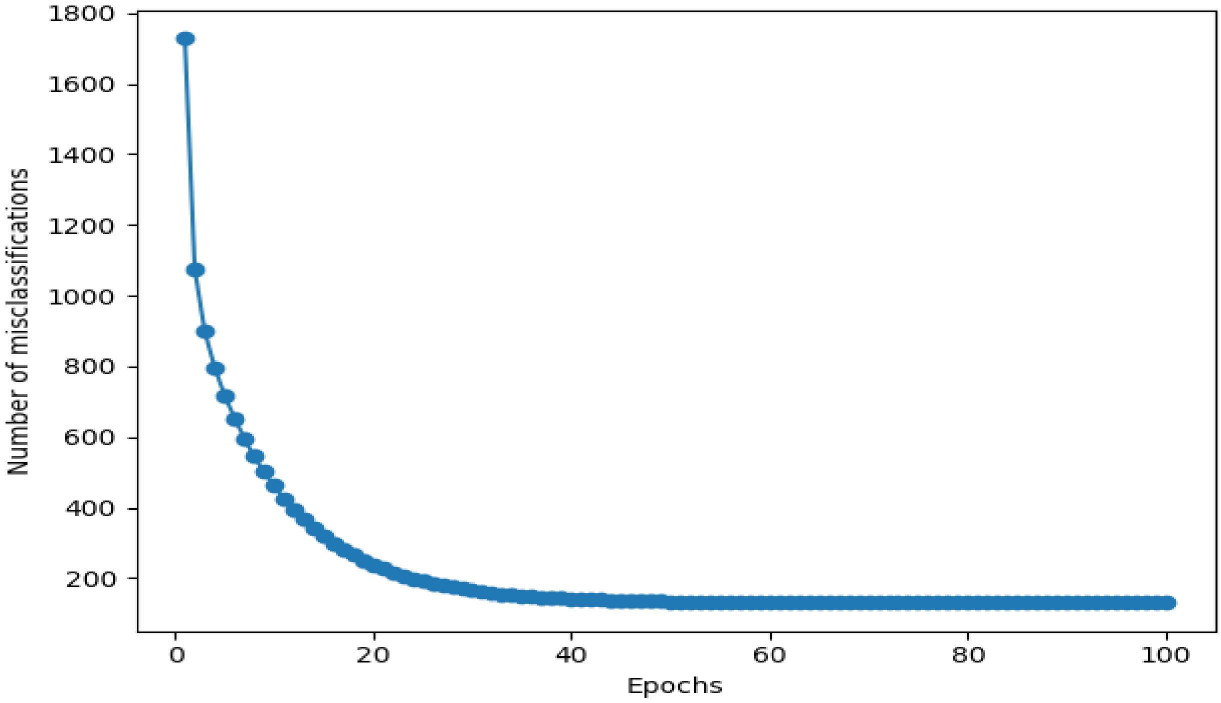

Figure 5: Number of misclassifications in each iteration

by applying Adaline classifier on Phishing data set

The cost of each iteration by using Adaline classifier

```
 [1729.0126033988993,    1073.7965358177973,    898.8011274929022,    796.0126843002786,
717.8440692414247,    652.3590799720398,    595.4688760241999,    545.3185365507327,
500.8273280997659,    461.2373712457337,    425.95366035423785,    394.4799118496191,
366.38922035182213,    341.3085297973437,    318.90912141330045,    298.900011469178,
281.02292893900176,    265.0482582641509,    250.77163802486533,    238.01104524843038,
226.6042623474186,    216.40665846714475,    207.28923631233684,    199.1369070452798,
191.84696327476948,    185.32772530537932,    179.49733962439993,    174.2827115682753 7,
169.6185565088632,    165.44655589403294,    161.71460616630324,    158.37615003209854,
155.38958080751627,    152.71771165690654,    150.32730249310862,    148.1886381427587,
146.27515211292837,    144.56309094002012,    143.03121466978112,    141.66052951825583,
140.43404920587514,    139.33658184790613,    138.3545396304402,    137.47576880742542,
136.68939782569663,    135.98570162567182,    135.35598037898143,    134.79245111395608,
134.28815084839138,    133.8368499987858,    133.43297496844832,    133.0715389353907,
132.74807996641727,    132.45860567777177,    132.19954374639286,    131.96769765041196,
131.76020708401254,    131.57451255105687,    131.408323694766,    131.25959096792005,
131.1264802901478,    131.00735037645913,    130.90073245472288,    130.80531211975838,
130.71991309846368,    130.6434827243092,    130.5750789408786,    130.51385867321994,
130.45906742282176,    130.41002995726765,    130.36614197924564,    130.32686267176402,
```

130.29170802731352, 130.26024487844745, 130.2320855559558, 130.20688310859146, 130.18432702526712, 130.16413940686562, 130.1460715403747, 130.12990083303544, 130.11542806865003, 130.10247495217644, 130.09088191230452, 130.08050613489496, 130.07121980301616, 130.06290852186618, 130.05546990915025, 130.0488123335276, 130.04285378556955, 130.0375208673054, 130.03274788789795, 130.0284760542983, 130.0246527469027, 130.02123087128183, 130.0181682779918, 130.0154272433159, 130.01297400453674, 130.01077834401156, 130.0088132169252, 130.0070544181331]

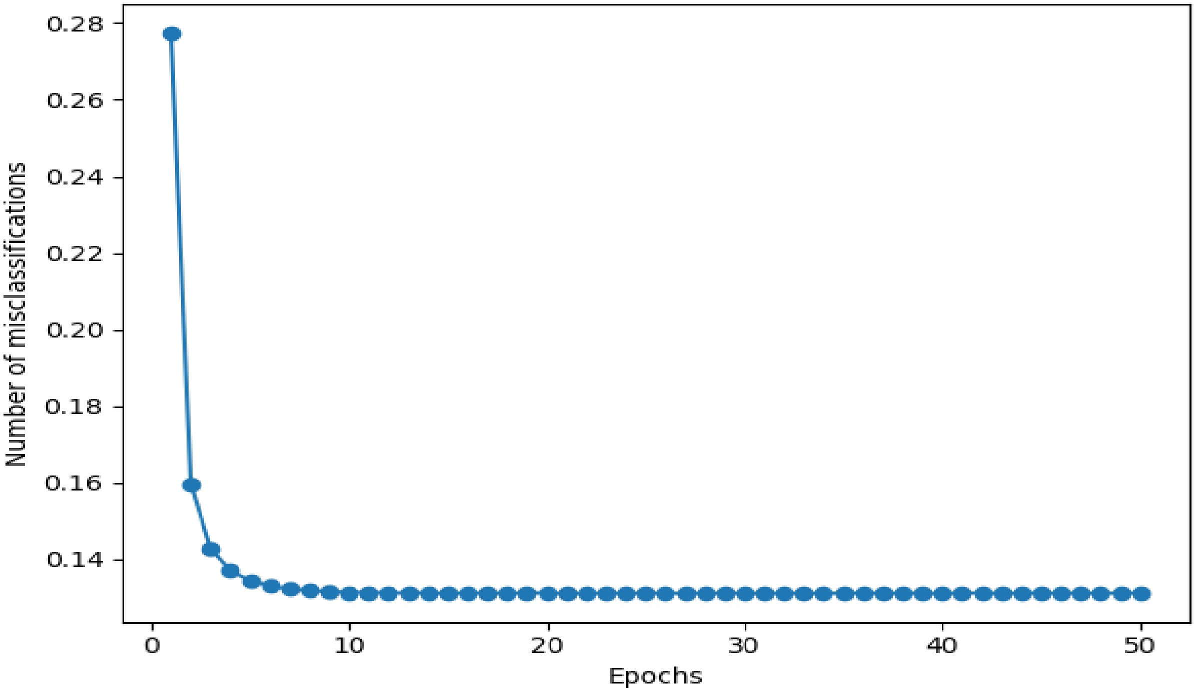

Figure 6: Number of misclassifications in each iteration

by applying SGD classifier on Phishing data set

The cost of each iteration by using SGD classifier

[0.27744126926816165, 0.15962732132693402, 0.1427910820510001, 0.13702747490209705, 0.1344754987907668, 0.13301309894278945, 0.13214145411002545, 0.13184484259099302, 0.1315146786142308, 0.1313143330531532, 0.13132463668299893, 0.13119712116965923, 0.1312150817848623, 0.1311521675526749, 0.13109965080157313, 0.1311126585984166, 0.13118746765851425, 0.13105875761497818, 0.13118829499637233, 0.13116789344795585, 0.13100774593549508, 0.13108551834695822, 0.1310718143167682, 0.13110278549898832, 0.13109583157763702, 0.131078215305363, 0.13107025238447143, 0.13108801958418573, 0.13110253999958218, 0.13107684611566822, 0.1310272070175422, 0.13107582773962045, 0.13110335640720525, 0.131042213290902, 0.13112699339254966, 0.1311248929665534, 0.13108649838902423,

```
0.13109092443745676,          0.13108535557760576,          0.13097025760814962,
0.13115044535212972,          0.13114242167189583,          0.13099180138339275,
0.1311679130105009,           0.13113393832422185,          0.13108055696002838,
0.13113383387322183,          0.13104675611132632,          0.13114754086295202,
0.13114059313581208]
```

5-c

I use fishing dataset from UCI machine Leaning repository (https://archive.ics.uci.edu/ml/machine-learning-databases/00379/) dataset contains 9 attributes ,1353 samples and 3 class labels {0,1,-1}. Also, I used iris data set which contains 4 features and three class labels { Iris-setosa, Iris-versicolor

, Iris-virginica}

5-d

Applying perceptron classifier on IRIS dataset shows that the number of misclassifications during each iteration fix after the second iteration. Based on that the weights keep unchanged during the rest of iteration. Therefore, we can say that the perceptron guarantees the two class are linearly separable. On other hand, applying perceptron on fishing dataset shows that the number of misclassifications during each iteration keep change. Based on that the weights keep changed. Therefore, the perceptron does not guarantee the two class are liner separately.

Applying Adaline and SGD classifiers on IRIS dataset shows that the cost values keep unchanged after some of iterations. Based on that the Adaline classifier guarantee the two class are liner separable. On other hand, applying Adaline and SGD classifiers on phishing dataset. The cost result show that the Adaline and SGD classifiers does not guarantee the two classes are linearly separate.

Feature scaling is one of common preprocessing step in machine learning. The idea of feature scaling is that make all of the features on same scale (standardization method). Perceptron, SGD and Adaline are sensitive to feature scaling. It helps to update the weights faster.

5-c

The difference between perceptron classifier and Adaline classifier is that weights are updated based on liner activation method rather than unit step function. Moreover, the weight is updated by minimizing the cost function using gradient descent. In addition, the perceptron updates the weight after evaluating each individual instance. In Adaline, updating the weight based on the whole train dataset for that both of them are not good with huge dataset. Therefore, the alternative for both is SGD classifier updates the weight incrementally for each training instance and it applicable for large dataset

One VS all with SGD classifier:

Every time we choose one of the class and turn its values into positive and other classes values into negative.

One VS all with SGD classifier on IRIS Data set:

```
----------------------------------------------------------------------

The number of instances which classified correctly 45

The accuracy 1.0

The error rate is 0.0

-----------------------------------------------------------------------

The number of instances which classified correctly 29

The accuracy is 0.6444444444444445

The error rate is 0.3555555555555555

--------------------------------------------------------------------------

The number of instances which classified correctly 14

The accuracy is 0.3111111111111111

The error rate is 0.6888888888888889
```

One VS all with SGD classifier on Phishing Data set:

```
The number of instances which classified correctly 329

The accuracy 0.9293785310734464

The error rate is 0.07062146892655363

-----------------------------------------------------------------------

The number of instances which classified correctly 190

The accuracy is 0.536723163841808

The error rate is 0.46327683615819204

--------------------------------------------------------------------------

The number of instances which classified correctly 168

The accuracy is 0.4745762711864407

The error rate is 0.5254237288135593
```