jxu72364@usc.edu

1.

(a)

I am using scikit-learn and python.

(b)

Mean and standard deviation are showed below:

```
mean:
[1.29653933e+01 2.27000000e+00 2.37629213e+00 1.96494382e+01
 9.89101124e+01 2.27235955e+00 2.02943820e+00 3.60674157e-01
 1.57617978e+00 5.09123596e+00 9.53213483e-01 2.56033708e+00
 7.29707865e+02]


standard deviation:
[8.19560112e-01 1.10306417e+00 2.73004021e-01 3.46517583e+00
 1.15589748e+01 6.14548382e-01 9.19718276e-01 1.20241309e-01
 5.41470129e-01 2.40437795e+00 2.29907577e-01 7.23461482e-01
 3.07221225e+02]
```

the mean and standard deviation are showed above

The reason why normalizing factors should be calculated from the training data only is that the test data set is supposed to simulate data set that the algorithm did not see before. In other words, the algorithm never saw test data before, so we cannot use the mean and deviation of testing data, but we can use the mean and deviation of training data to simulate testing data. Therefore, we need to use the mean and deviation of training data to standardize testing data.

(C)

(i) The default initial weight vector is 0

(ii) The default halting condition is when the maximum iterations reaches to 1000 or tol is less than 1e-3. tol is the stopping criterion. If it is not None, the iterations will stop when (loss > previous_loss - tol).

If the solution weight vector is not reached, then the backup halting condition is iterate 1000 times.

(d)

```
 resulting 3 weight vectors:
[[-1.          2.39810751 -2.10323213]
 [-1.         -3.65391958 -0.82497467]
 [ 0.          1.4642001  -0.39888885]]


classification accuracy:
0.797752808988764
```

the resulting 3 weight vectors and the classification accuracy of 2 features training set are showed above. The result is in augmented feature space.

```
 resulting 3 weight vectors:
[[-1.          2.39810751 -2.10323213]
 [-1.         -3.65391958 -0.82497467]
 [ 0.          1.4642001  -0.39888885]]


classification accuracy:
0.7752808988764045
```

the resulting 3 weight vectors and the classification accuracy of 2 features testing set are showed above. The result is in augmented feature space.

```
 resulting 3 weight vectors:
[[-5.          4.14486533  1.79499983  2.45952668 -3.21854057  1.1722979
  -1.08730617  3.73722161 -0.13829867  0.3894931  -3.03355811  2.81011799
   2.61477002  4.98188016]
 [-2.         -4.95291733 -3.07325729 -1.41908434  3.11640071 -2.69745163
   1.09791047  0.22710912  1.09237262 -0.66049895 -5.36418499  2.83777932
  -0.13729251 -3.86231215]
 [-5.          0.31354173  1.88565639  0.40827484  1.22567836  3.24860661
  -0.89204083 -4.01515234 -1.21945789 -0.46004592  4.78135303 -3.80123438
  -2.15397039 -0.07961909]]


classification accuracy:
1.0
```

the resulting 3 weight vectors and the classification accuracy of 13 features training set are showed above. The result is in augmented feature space.

```
 resulting 3 weight vectors:
[[-5.          4.14486533  1.79499983  2.45952668 -3.21854057  1.1722979
  -1.08730617  3.73722161 -0.13829867  0.3894931  -3.03355811  2.81011799
   2.61477002  4.98188016]
 [-2.         -4.95291733 -3.07325729 -1.41908434  3.11640071 -2.69745163
   1.09791047  0.22710912  1.09237262 -0.66049895 -5.36418499  2.83777932
  -0.13729251 -3.86231215]
 [-5.          0.31354173  1.88565639  0.40827484  1.22567836  3.24860661
  -0.89204083 -4.01515234 -1.21945789 -0.46004592  4.78135303 -3.80123438
  -2.15397039 -0.07961909]]


classification accuracy:
0.9325842696629213
```

the resulting 3 weight vectors and the classification accuracy of 13 features testing set are showed above. The result is in augmented feature space.

(e)

```
classification accuracy
0.797752808988764


the final 3 weight vectors
[[-1.          2.39810751 -2.10323213]
 [-1.         -3.65391958 -0.82497467]
 [ 0.          1.4642001  -0.39888885]]
```

the resulting 3 weight vectors and the classification accuracy of 2 features training set are showed above. The result is in augmented feature space.

```
classification accuracy
0.7752808988764045


the final 3 weight vectors
[[-1.          2.39810751 -2.10323213]
 [-1.         -3.65391958 -0.82497467]
 [ 0.          1.4642001  -0.39888885]]
```

the resulting 3 weight vectors and the classification accuracy of 2 features testing set are showed above. The result is in augmented feature space.

```
classification accuracy
1.0


the final 3 weight vectors
[[-5.          4.14486533  1.79499983  2.45952668 -3.21854057  1.1722979
  -1.08730617  3.73722161 -0.13829867  0.3894931  -3.03355811  2.81011799
   2.61477002  4.98188016]
 [-2.         -4.95291733 -3.07325729 -1.41908434  3.11640071 -2.69745163
   1.09791047  0.22710912  1.09237262 -0.66049895 -5.36418499  2.83777932
  -0.13729251 -3.86231215]
 [-5.          0.31354173  1.88565639  0.40827484  1.22567836  3.24860661
  -0.89204083 -4.01515234 -1.21945789 -0.46004592  4.78135303 -3.80123438
  -2.15397039 -0.07961909]]
```

the resulting 3 weight vectors and the classification accuracy of 13 features training set are showed above. The result is in augmented feature space.

```
classification accuracy
0.9325842696629213


the final 3 weight vectors
[[-5.          4.14486533  1.79499983  2.45952668 -3.21854057  1.1722979
  -1.08730617  3.73722161 -0.13829867  0.3894931  -3.03355811  2.81011799
   2.61477002  4.98188016]
 [-2.         -4.95291733 -3.07325729 -1.41908434  3.11640071 -2.69745163
   1.09791047  0.22710912  1.09237262 -0.66049895 -5.36418499  2.83777932
  -0.13729251 -3.86231215]
 [-5.          0.31354173  1.88565639  0.40827484  1.22567836  3.24860661
  -0.89204083 -4.01515234 -1.21945789 -0.46004592  4.78135303 -3.80123438
  -2.15397039 -0.07961909]]
```

the resulting 3 weight vectors and the classification accuracy of 13 features testing set are showed above. The result is in augmented feature space.

(f)

Compare 2 features to 13 features in (d) for each dataset, we can know that for 13 features, we will have higher accuracy. Because it allow us to have more features to classify. If we only use the first 2 features, maybe we can not linearly classify very well. But if we have more features, probably, it allow us to linearly classify all most of the points.

Compare 2 features to 13 features in (e) for each dataset, although the accuracy is better than(d), but for the same reason as in(d). the accuracy of 13 features is higher than 2 features. Because probably with more features we are more likely to linearly classify all the datapoints.

Compare (d) to (e) for 2 features for each dataset we can find that in (e) we have same accuracy to (d), because the goal of perceptron is to optimize initial weight vector, we can find that after 100 times, we can get the same result as in (d). So, the model is always convergence with different initial weight vector, just the speed of convergence is different.

Compare compare (d) to (e) for 13 features for each dataset for the same reason, we can find that in (e) we have same accuracy to (d).The reason is just like what I explained above.

(g)

```
0.7528089887640449
```

Using unnormalized data, the classification accuracy of test data for the first 2 features is showed above

```
0.9775280898876404
```

Using unnormalized data, the classification accuracy of test data for the first 13 features is showed above

(h)

```
0.7528089887640449
```

Using standardized data, the classification accuracy of test data for the first 2 features is showed above

```
0.9775280898876404
```

Using standardized data, the classification accuracy of test data for the first 13 features is showed above

(i)

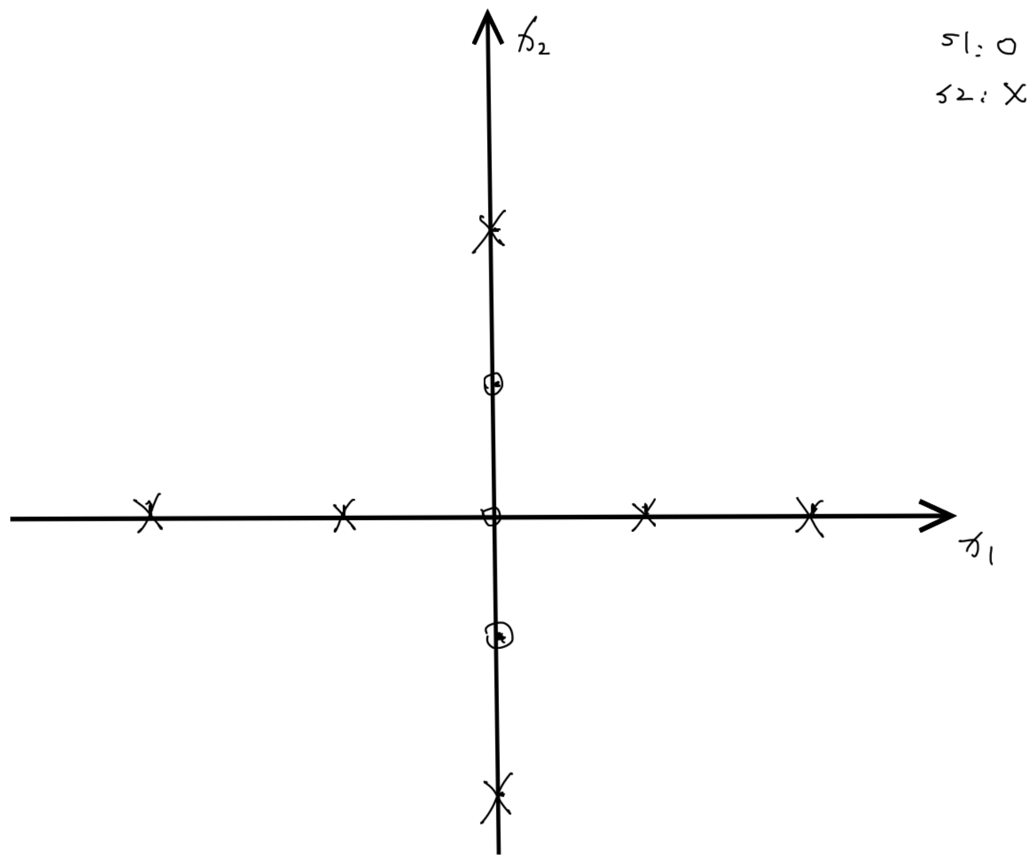
The results of (g) and (h), they are identical

(j)

The results of (h) and (e), they are similar but not identical, because unlike Perceptron, MSE will give no guarantee of 100% correct classification on a linearly separable training dataset. And the goal of MSE is not correctly classify all datapoints, it just use an ideal output bn for the discriminant function of the n^{th} data point Xn. Therefore, we can find that the accuracy in (e) is higher than the accuracy in (h)

Problem2

(a)



from the picture we can know that they are not linearly separable

$$\underline{u} = \underline{\phi}(\underline{x}) \stackrel{\Delta}{=} \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{bmatrix}_{[(D'+1)\times 1]} \qquad \underline{w} \stackrel{\Delta}{=} \begin{bmatrix} w_{00} \\ w_{01} \\ w_{02} \\ w_{11} \\ w_{12} \\ w_{22} \end{bmatrix}$$

$D'+1 = 6$

(b)

So, $\underline{u} = \underline{\phi}(\underline{x}) \triangleq \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{bmatrix}_{[(D'+1)\times 1]}$ $\quad \underline{w} \triangleq \begin{bmatrix} w_{00} \\ w_{01} \\ w_{02} \\ w_{11} \\ w_{12} \\ w_{22} \end{bmatrix}$

$D'+1 = 6$

So, S_1: $\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$

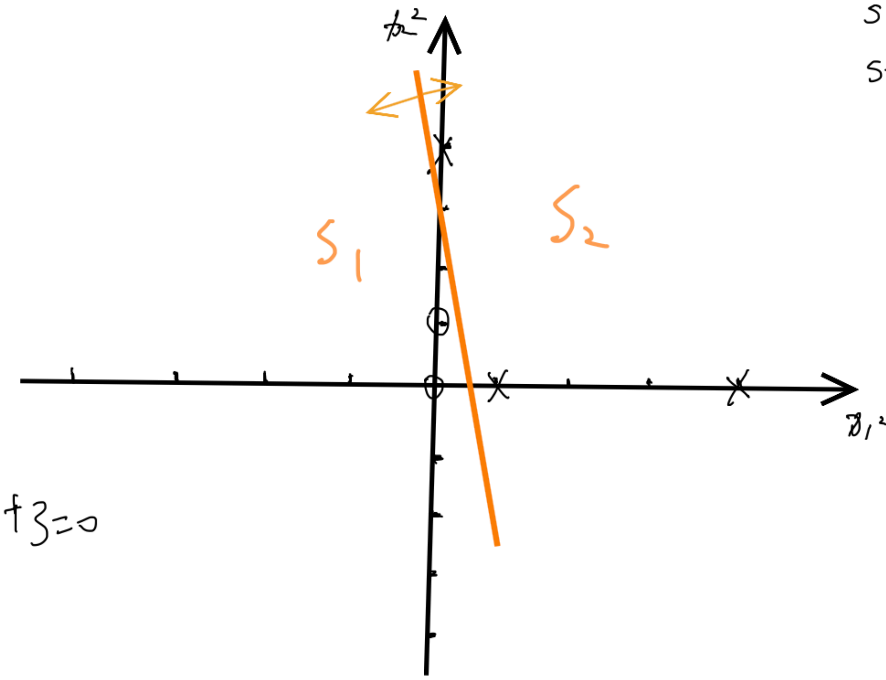
S_2: $\begin{bmatrix} 1 \\ -2 \\ 0 \\ 4 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ -2 \\ 0 \\ 0 \\ 4 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \\ 0 \\ 4 \\ 0 \\ 0 \end{bmatrix}$

$g(x) = w_{11} x_1^2 + w_{12} x_1 x_2 + w_{22} x_2^2 + w_{01} x_1 + w_{02} x_2 + w_{00}$

(c)

S_1: o

S_2: x



$-6x_1^2 - x_2^2 + 3 = 0$

So, $\underline{w}'$ that correctly separates the prototypes in the expanded feature will be

$$\underline{w}' = \begin{bmatrix} 3 \\ 0 \\ 0 \\ -6 \\ 0 \\ -1 \end{bmatrix} \qquad \underline{u} \triangleq \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{bmatrix}_{[(D'+1)\times 1]}$$

$g(\underline{u}) = \underline{w}'^T \underline{u}$

$g(\underline{u}) = -6u_4 - u_6 + 3 \underset{S_1}{\overset{S_2}{\gtrless}} 0$

(d) we can know that $\underline{u} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ x_1^2 \\ x_1 x_2 \\ x_2^2 \end{bmatrix}$

so $g(\underline{x}) = -6x_1^2 - x_2^2 + 3$

$$g(\underline{x}) = -6x_1^2 - x_2^2 + 3 \underset{S_2}{\overset{S_1}{\gtrless}} 0$$