Machine Learning Assignment 1

BHARAT KASHYAP KARRI *2020AAPS0319H*
SRIRAM KASHYAP KURILLA *2020AAPS0359H*
MUSKAAN KUMAR *2020AAPS2188H*

Part A- Perceptron

The perceptron algorithm is as follows:

Initialize $\vec{w} = \vec{0}$
while TRUE **do**
  $m = 0$
  for $(x_i, y_i) \in D$ **do**
    if $y_i(\vec{w}^T \cdot \vec{x_i}) \leq 0$ **then**
      $\vec{w} \leftarrow \vec{w} + y\vec{x}$
      $m \leftarrow m + 1$
    end if
  end for
  if $m = 0$ **then**
    break
  end if
end while

Learning Task 1

The accuracy of the dataset to detect whether the digitized images show benign or malignant cells is

Number of iterations	Learning Rate	Training accuracy	Testing Accuracy
5000	0.01	91.08	93.09
5000	0.001	90.81	94.15
10000	0.01	87.14	82.98
10000	0.001	90.55	90.96

By changing the order of the training examples, the following accuracy is found for Iterations-5000 and Learning Rate -0.001is 90.25(training) and 90.43(testing)
Variance of training accuracy: 0.000342
Variance of testing accuracy: 0.002554

Learning Task 2
On the normalized data for Iterations-5000 and Learning rate- 0.001 we get training accuracy of 90.81 and testing accuracy of 86.70.

Learning Task 3
After taking permutations of the features of the data we are getting the same accuracy of 90.5(training) and 90.43(testing). This shows that after shuffling the accuracy will not change.

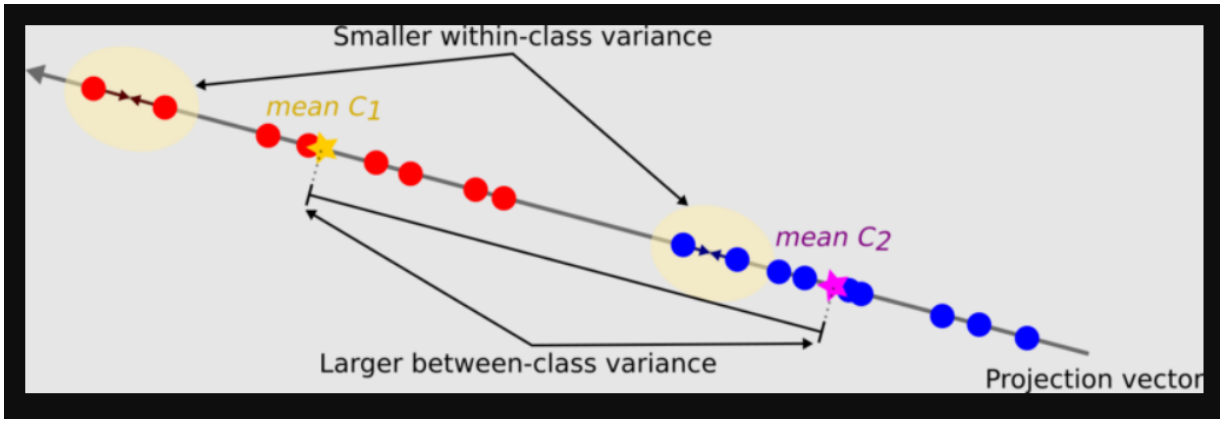
Following is the accuracy for 10 different train-test splits

No.	Number of Iterations	Learning Rate	Training Accuracy	Testing Accuracy
1	5000	0.001	89.24	86.17
2	5000	0.001	92.39	90.96
3	5000	0.001	92.39	89.36
4	5000	0.001	61.55	71.28
5	5000	0.001	91.60	89.89
6	5000	0.001	91.34	91.49
7	5000	0.001	81.63	84.04
8	5000	0.001	93.18	87.77
9	5000	0.001	92.39	90.43
10	5000	0.001	91.60	93.62

As observed we are getting close to the same accuracy in PM4 compared to PM1.
PM2(normalized) gives better results compared to non-normalized.
Our model for perceptron achieves the highest accuracy of 0.9362.
Average training accuracy was obtained as 0.8773 and the variance of accuracy as 0.009586.

Part B- Fisher's Linear Discriminant Analysis

The Fisher Linear Discriminant Analysis maximizes the mean between classes and minimizes the in-class variance of the projection of data onto lower dimensions.



Both feature engineering tasks were done prior to performing Fisher's LDA. 10 different seeds were used to split randomize the shuffling.

Learning Task 1 - Without shuffling

```
[2832511248, 922759433, 2885193187, 971167687, 3748846250, 1301998448, 3306749904, 1464561512, 650496267, 2848102087]

Accuracy for seed 0: 0.612
Accuracy for seed 1: 0.654
Accuracy for seed 2: 0.819
Accuracy for seed 3: 0.670
Accuracy for seed 4: 0.771
Accuracy for seed 5: 0.867
Accuracy for seed 6: 0.846
Accuracy for seed 7: 0.628
Accuracy for seed 8: 0.670
Accuracy for seed 9: 0.580
Task1 Average Accuracy: 0.712 Variance in accuracy: 0.00984
```

The seeds are mentioned at the top of the above image. Each run corresponds to different splits in the data. It can be seen that the accuracy peaks at around 80% with the average accuracy at 71.2%.

Learning Task 2 - With shuffling of columns

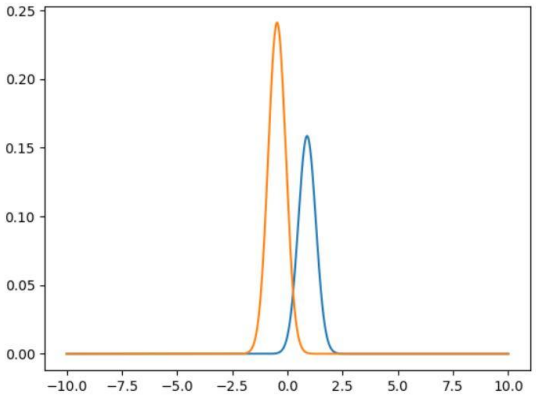
```
Accuracy for seed 0: 0.574
Accuracy for seed 1: 0.612
Accuracy for seed 2: 0.622
Accuracy for seed 3: 0.617
Accuracy for seed 4: 0.649
Accuracy for seed 5: 0.644
Accuracy for seed 6: 0.601
Accuracy for seed 7: 0.633
Accuracy for seed 8: 0.660
Accuracy for seed 9: 0.606
Task2 Average Accuracy: 0.622 Variance in accuracy: 0.00058
```

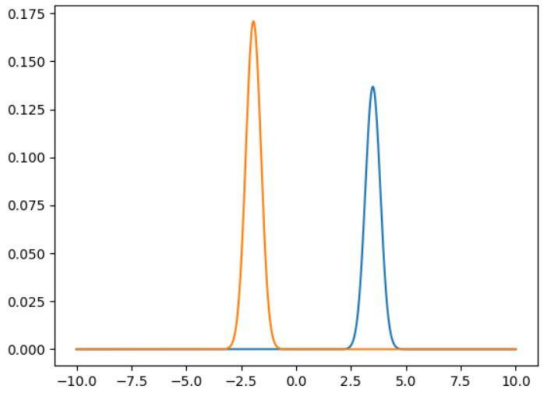
It can be seen that the average accuracy and the peak accuracy are different but close to the previous case. We believe that this is due to the stochastic nature of projection onto 1 dimensional data from the 32 original dimensions. It must be noted that the variance

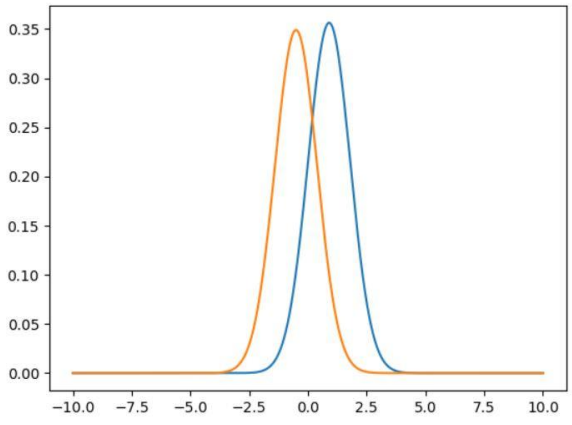
between the 2 cases is not the same. The second model's lower variance in accuracy indicates that it is a more robust and reliable model.

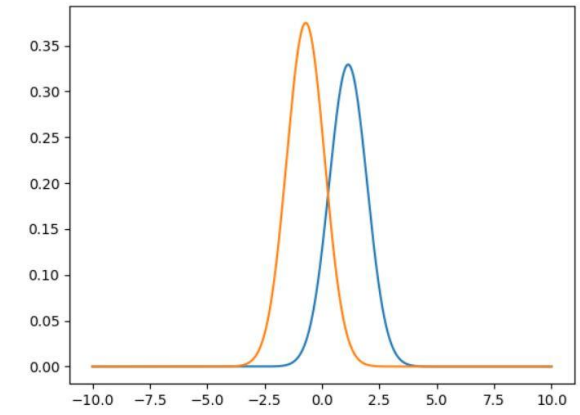
Shuffling of the features of the data thus is expected not to change the results.

Following are a few graphs that depict the distribution of the projected one dimensional data separated between the two classes. Different seeds provide different boundaries and different means and variances. One with the largest difference between the means and the lowest in-class variance is considered to be the best model.









Part C- Logistic Regression

Logistic Regression uses the sigmoid function to transform the output of a linear regression into a probability distribution, which allows us to predict the likelihood of an event occurring based on values of other variables

$$L_{BCE} = -\frac{1}{n}\sum_{i=1}^{n}(Y_i \cdot \log \hat{Y}_i + (1 - Y_i) \cdot \log (1 - \hat{Y}_i))$$

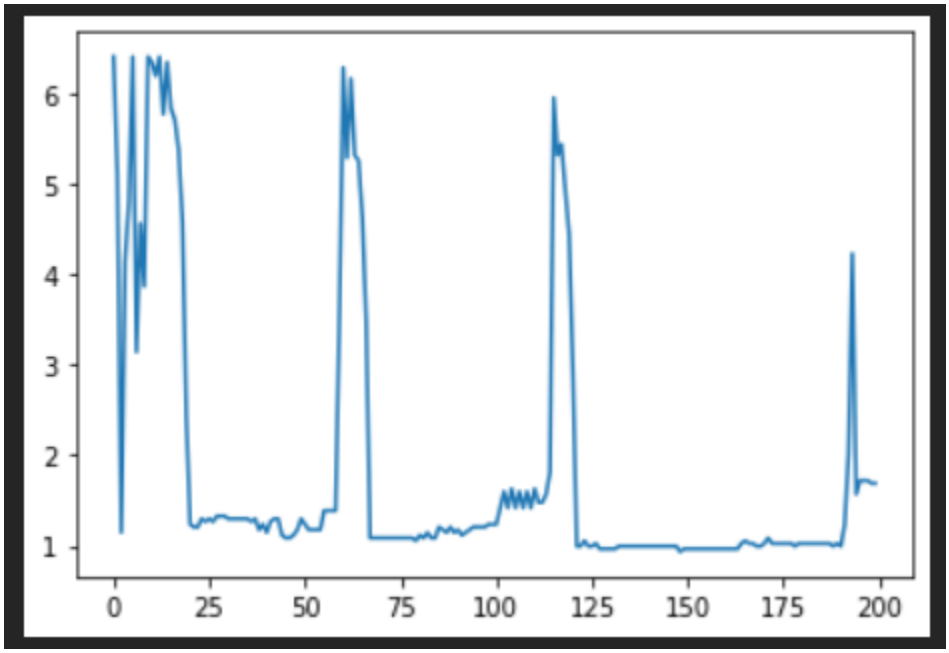
Learning Task 1 with lr=0.01

Model	Threshold	Accuracy
GD	0.3	0.65
	0.4	0.65
	0.5	0.65
	0.6	0.65
	0.7	0.65
SGD	0.3	0.81
	0.4	0.74
	0.5	0.91
	0.6	0.95
	0.7	0.65
MBSGD	0.3	0.95
	0.4	0.73
	0.5	0.87

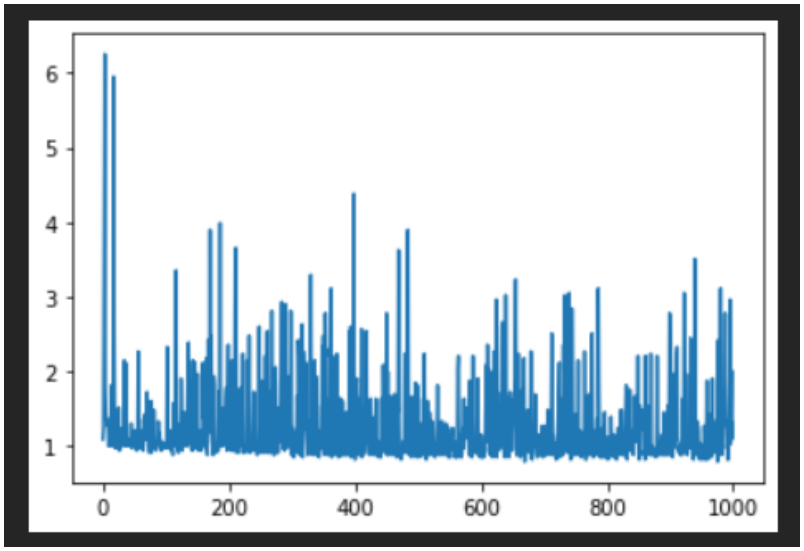
	0.6	**0.85**
	0.7	**0.80**

Average accuracy: 0.7677
Variance of accuracy: 0.0137

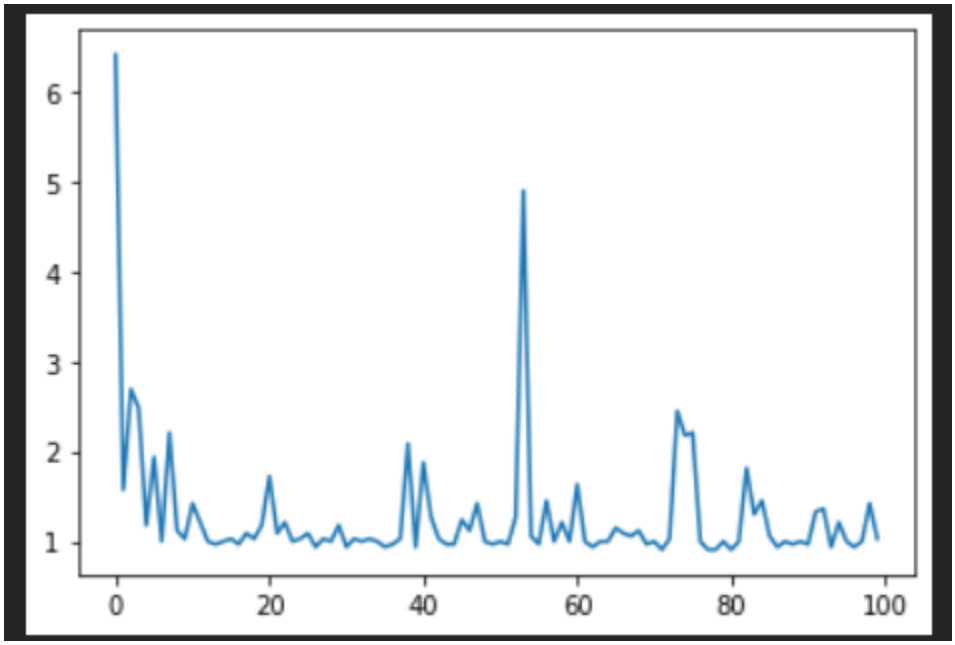
GD :



SGD :



MBSGD :



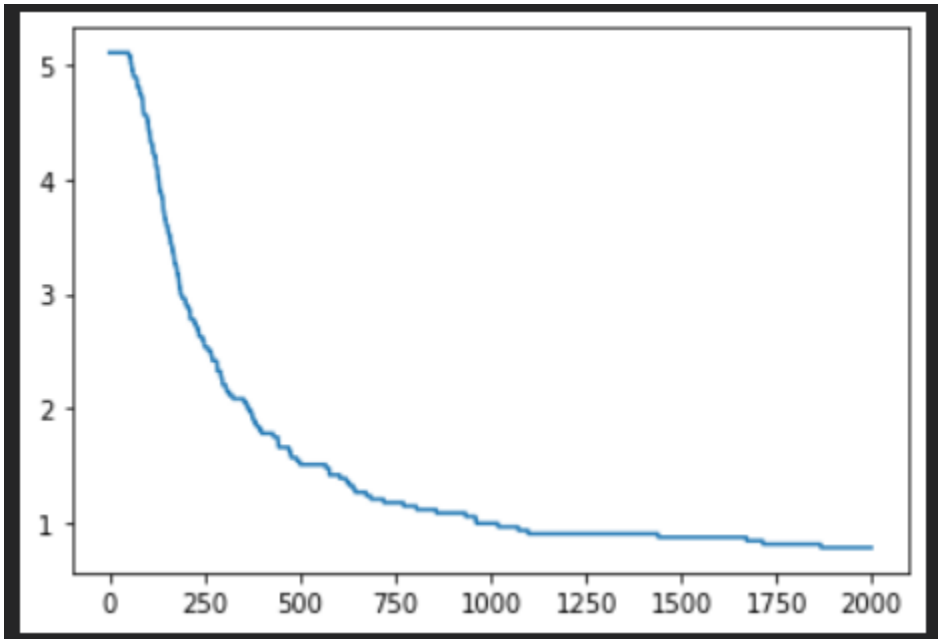
Learning Task 2
With lr = 0.01

Model	Threshold	Accuracy
GD	0.3	0.95
	0.4	0.967
	0.5	0.973
	0.6	0.983
	0.7	0.9839
SGD	0.3	0.9518
	0.4	0.9572
	0.5	0.9679
	0.6	0.9679
	0.7	0.9679
MBSGD	0.3	0.8823

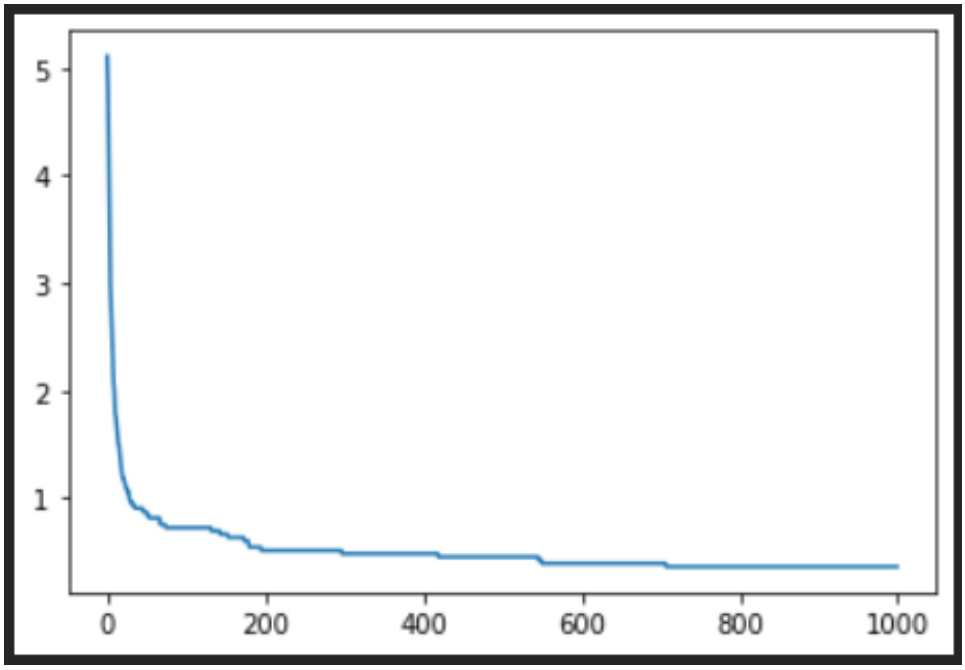
	0.4	**0.9358**
	0.5	**0.9358**
	0.6	**0.9465**
	0.7	**0.9358**

Average accuracy: 0.9537
Variance of accuracy: 0.00065

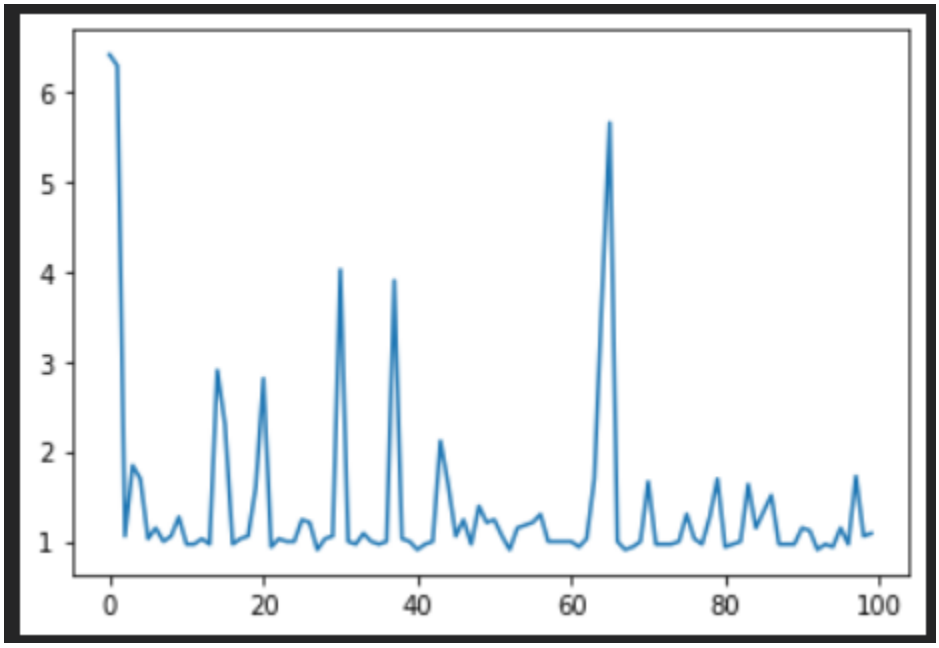
GD :



SGD :



MBSGD :



With lr = 0.001

Model	Threshold	Accuracy
GD	0.3	0.9197
	0.4	0.9518
	0.5	0.9786
	0.6	0.9839
	0.7	0.9839
SGD	0.3	0.9625
	0.4	0.9679
	0.5	0.9679
	0.6	0.9786
	0.7	0.9839
MBSGD	0.3	0.9090
	0.4	0.9090
	0.5	0.8128
	0.6	0.6844
	0.7	0.8128

Average accuracy: 0.9204
Variance of accuracy: 0.0075

With lr = 0.0001

Model	Threshold	Accuracy
GD	0.3	0.8342
	0.4	0.9037
	0.5	0.9732
	0.6	0.9732
	0.7	0.9732
SGD	0.3	0.9411
	0.4	0.9625
	0.5	0.9786
	0.6	0.9786
	0.7	0.9893
MBSGD	0.3	0.4598
	0.4	0.9518
	0.5	0.9251
	0.6	0.9465
	0.7	0.9358

Average accuracy: 0.9151
Variance of accuracy: 0.0174

Part D - Comparative Study

Accuracy	PM1	PM3	PM4	FLDM1	FLDM2	LR1	LR2
Mean	0.9025	0.9043	0.8773	0.712	0.622	0.7677	0.9537
Variance	0.0003	0.0026	0.0096	0.0098	0.0006	0.0137	0.00065

The maximum accuracy was achieved by the Logistic Regression model after normalization of the dataset. We believe this was the case as it is more capable of working with outliers and considering noise in the data. In an ideal scenario where the data is completely linearly separable with no noise present, the perceptron model would perform better. However as it can be seen from the data above that it did not perform perfectly. It is not possible to expect noiseless data from real world sources. Fisher's LDA did not perform well for the same reasons. In fact there is a larger impact as it is a pure mathematical model, that is, it considers every point to fit in the model and will continue to try finding parameters that would allow for the data to be as such.