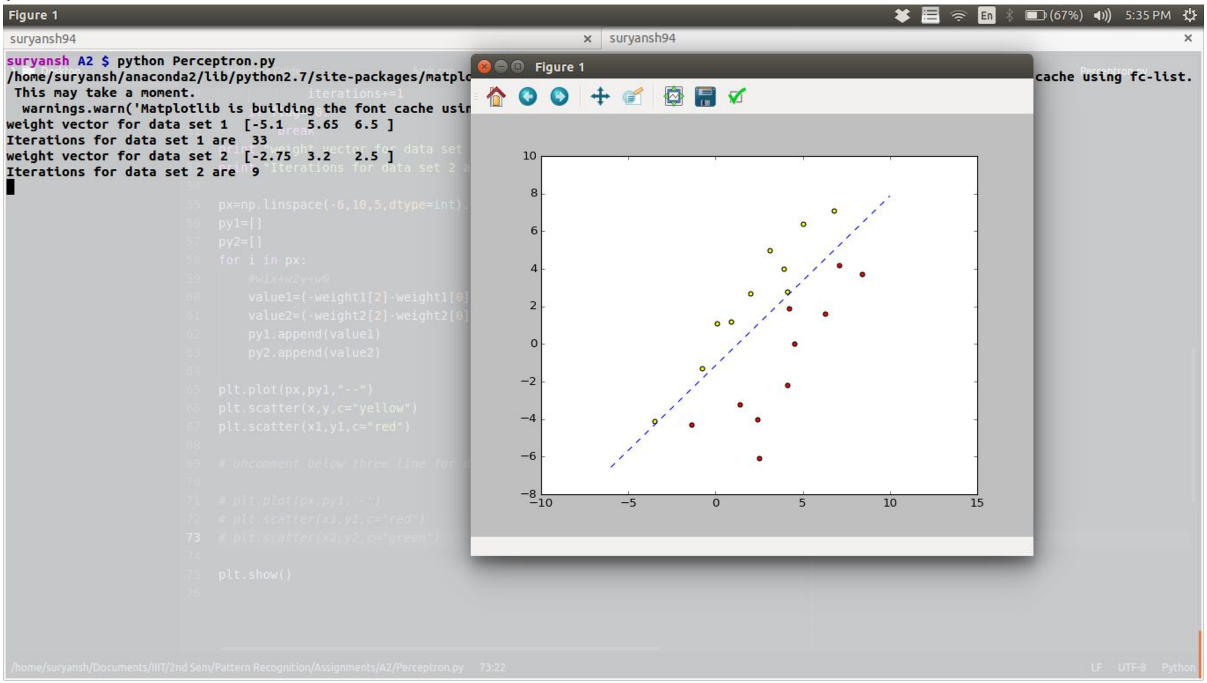
SMAI ASSIGNMENT 2

Submitted By:
Suryansh Agnihotri
20162037

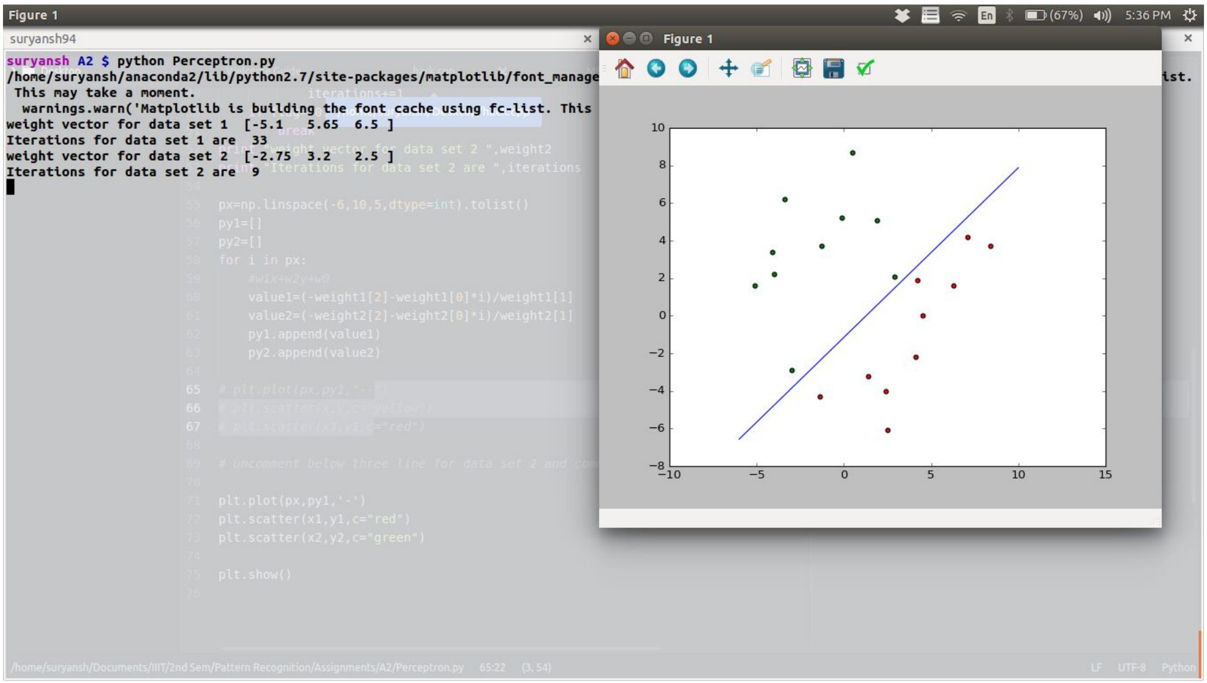
1) Perceptron

i) Number of iterations for data set 1 are 33.



ii)

Number of iteration for data set 2 are 9.



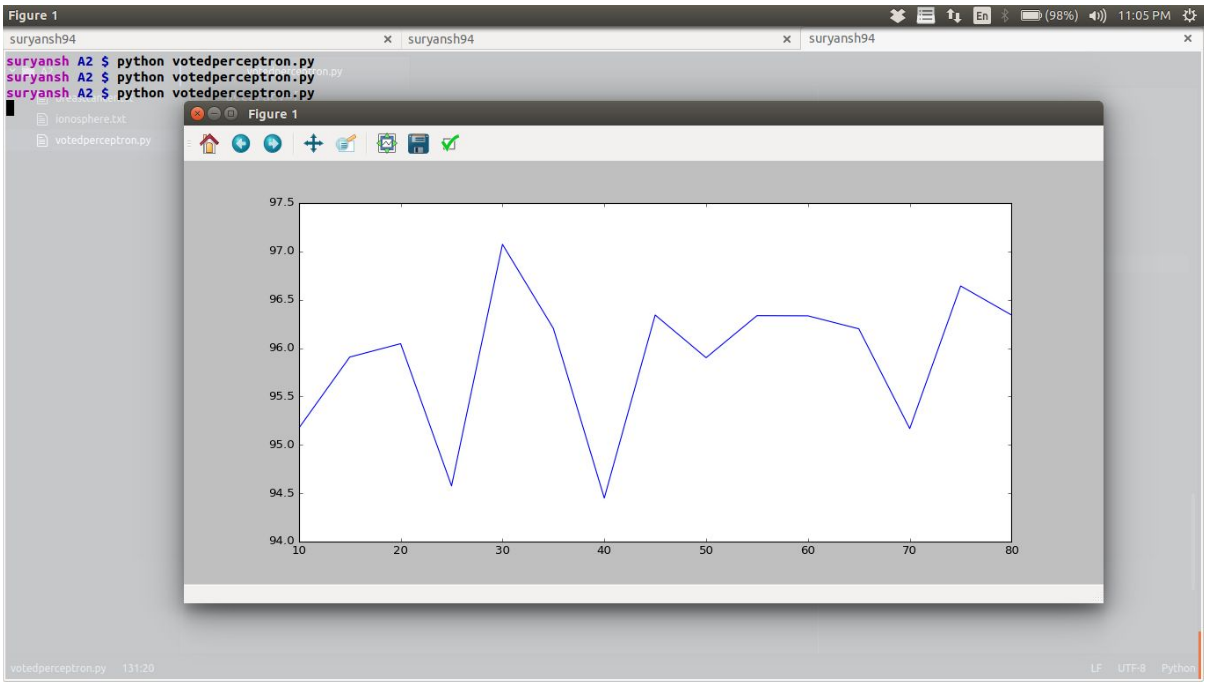
iii)

No of iterations required in the first case is 33 iterations whereas in the second case, 9 iterations are required. Learning rate is both in the same case. The main reason for the difference between number of iterations is because of the data points. In first case the data points are close and that why weight vector(if the point is misclassified) is updated by a small margin and so moving the weight vector in the solution region takes more time as compared to second case where the data is scattered far away and so the weight vector is updated by more margin.

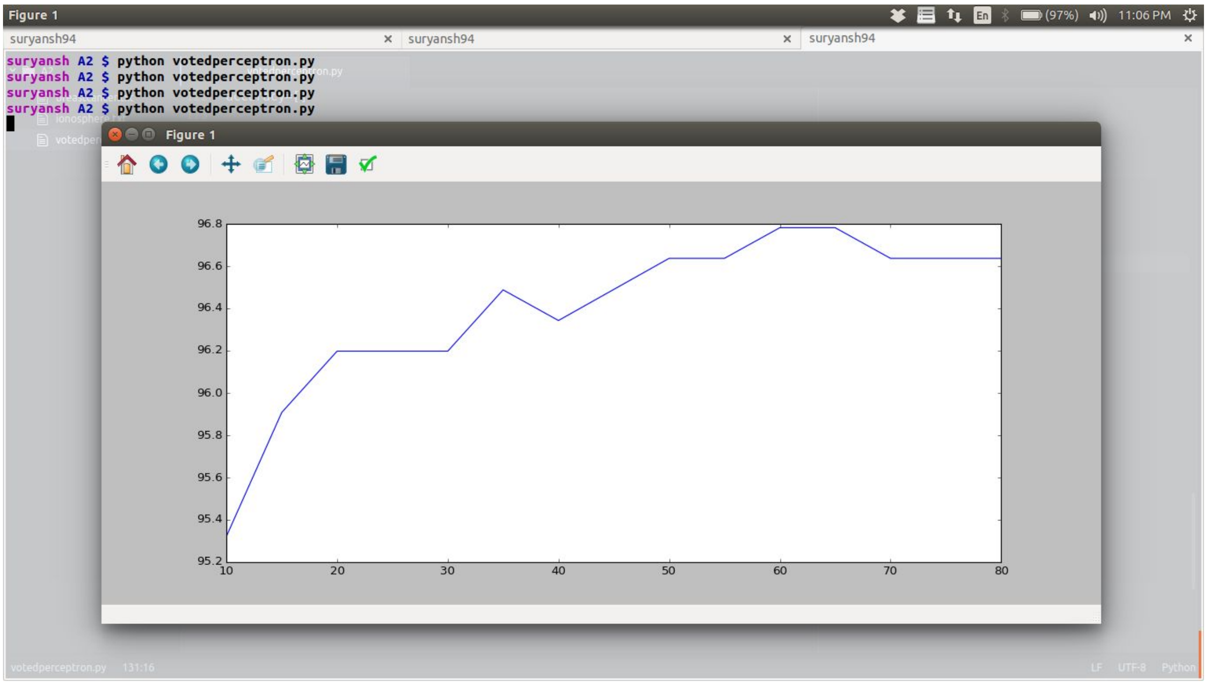
2)

Data Set of Breast Cancer

Epoch vs Accuracy (vanilla)

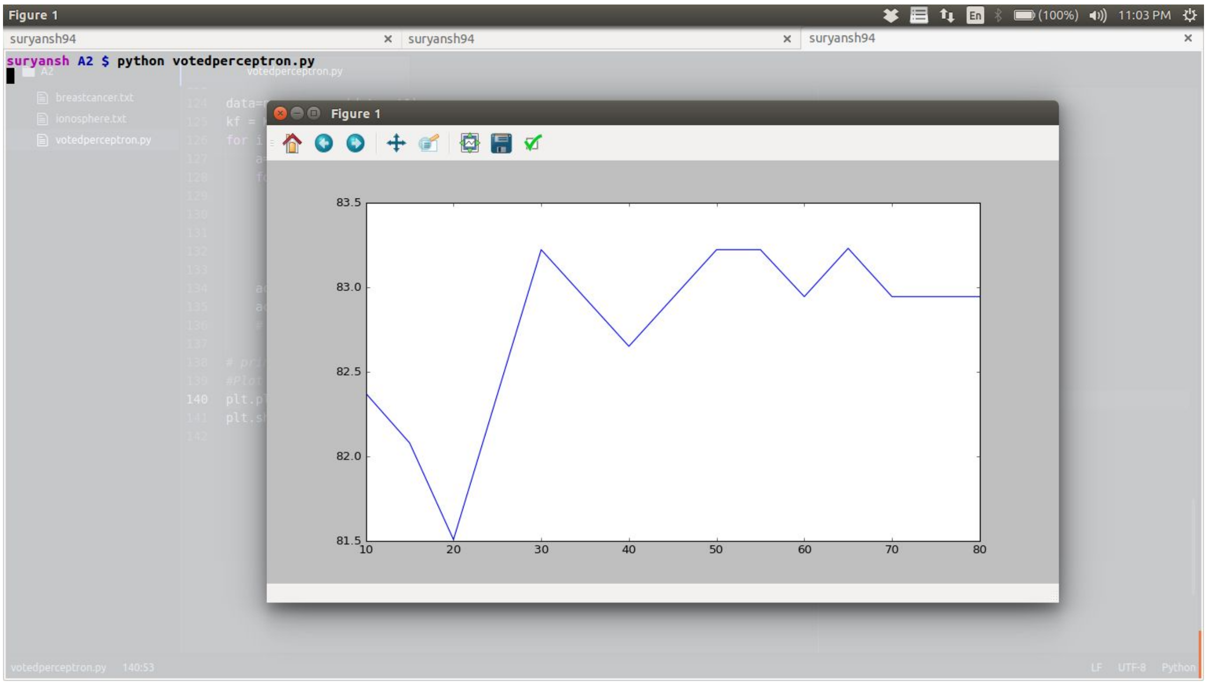


Epoch vs Accuracy (Voted Perceptron)

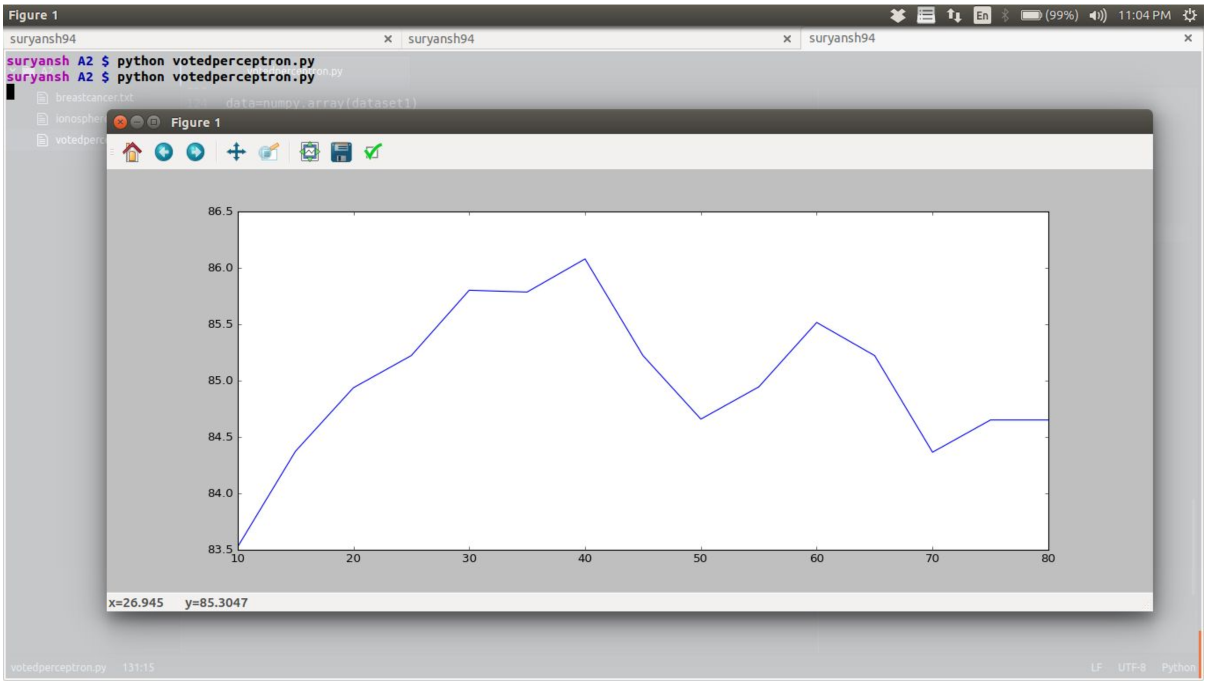


Data set of Ionosphere

Epoch vs Accuracy (Voted Perceptron)



Epoch vs Accuracy (vanilla)



Voted Perceptron takes into account the cumulative effect of all the weight vectors generated during the learning phase whereas Vanila Perceptron runs until all the samples are classified by one weight vector.

Voted Perceptron has more accuracy and efficiency due to the reason stated above .Although it is efficient, Voted Perceptron has more space complexity than Vanila Perceptron as we have to store all the weight vectors and their counts generated during the running of algorithm. As we can see from the graphs that the plot is a stabilized one for Voted after a certain number of epochs and not so for Vanila Perceptron. This is because Voted takes into account all the W's generated and after certain number of epochs the W changes rarely.

3)

Plot for data set 1 , dotted line represent Fishers Classfiers and other represents LMS.
Both lines are almost overlapping with each other

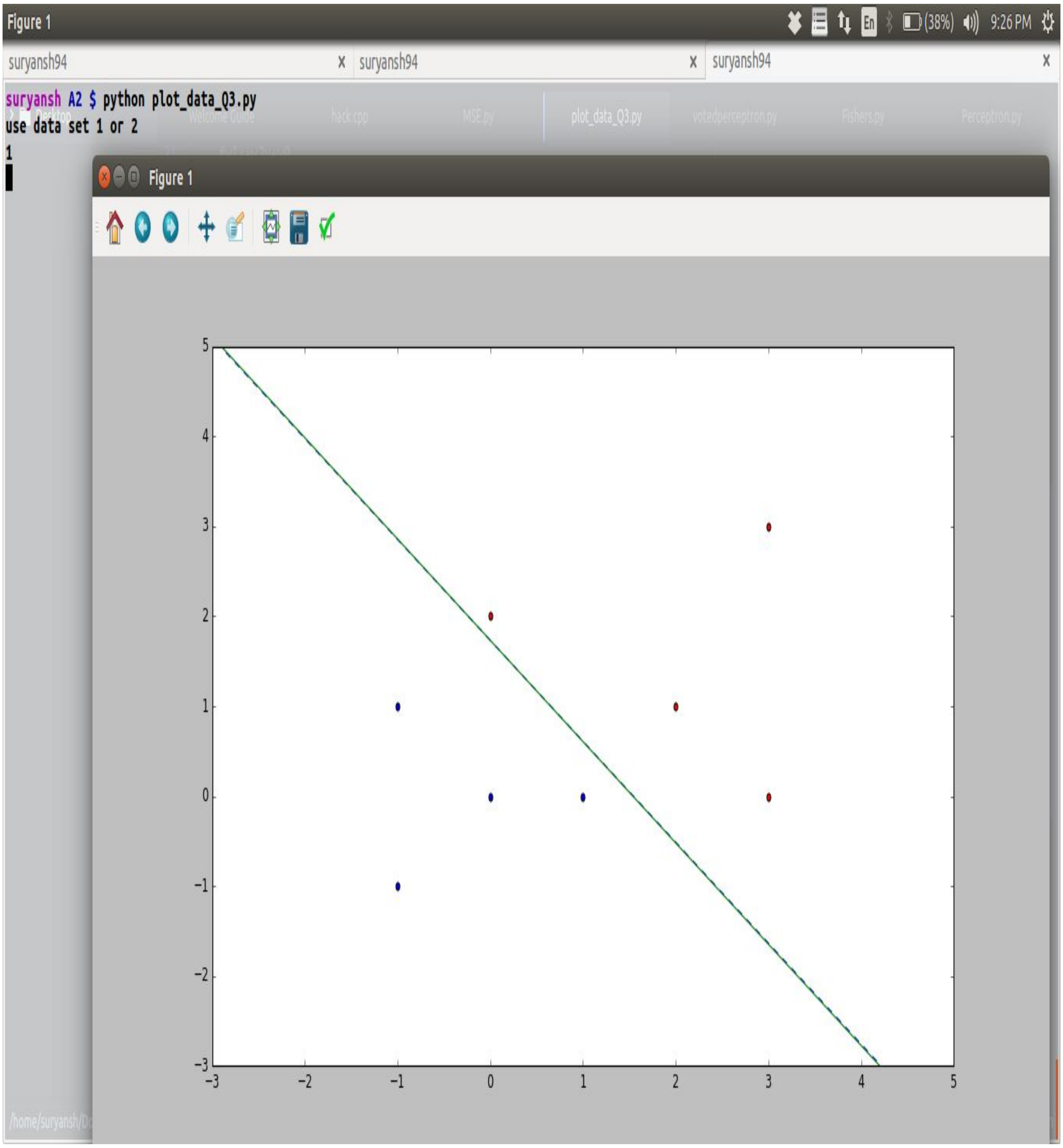
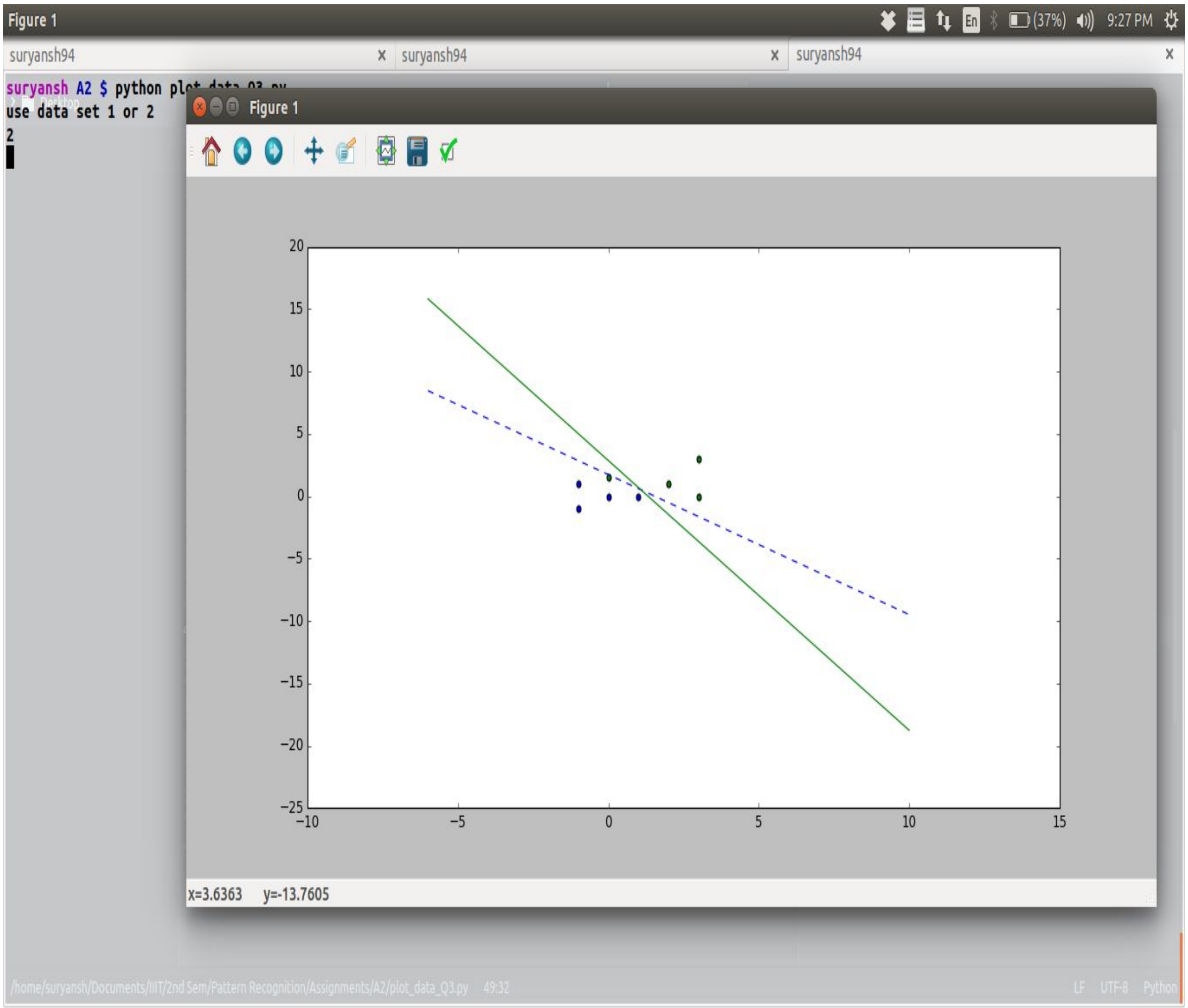


Figure 1
suryansh94
suryansh94
suryansh94
suryansh A2 $ python plot_data_Q3.py
use data set 1 or 2
1
Figure 1
5
4
3
2
1
0
−1
−2
−3
−3
−2
−1
0
1
2
3
4
5

Plot for data set 2 , dotted line represent Fishers Classfiers and other represents LMS



As it an be seen Fisher has correctly classified the sample in both the cases but the least square clasifier has not correctly classified data points in second case. The reason behind this is that least square classifies the point on the basis of the distance between the point and the plane and it try to minimize the sum of square errors. But the fisher discriminant focuses on finding weight vector that minimizes the misclassification errors. This is achieved by projecting the input vector onto a new subspace , where the classes are best separated.

4)

(4) Relation b/w Least Square & Fishers Descriminant

Let $S = \{(x_1,y_1),(x_2,y_2) \ldots (x_m,y_m)\}$ let m_1 be ~~m~~ points in class c1 & m2 be the No of points in class c2. m1 of which are in subset D_1 labelled ω_1 & m2 of which are in subset D_2 labelled as ω_2

We assume that sample z_i is formed from x_i by adding a threshold component $x_0 = 1$ to make an augmented pattern vector with no loss in generality we can assume that the first m1 sample are labelled ω_1 & second m2 are ω_2, The data is normalised if it $\in \omega_2$.

$$z = \begin{bmatrix} 1_1 & x_1 \\ 1_1 & -x_2 \end{bmatrix}$$

where 1_i is column vector of m_i ones & x_i is $m_i \times d$ matrix whose rows are the sample labelled ω_i

We partition a & y_i correspondingly

$$a = \begin{bmatrix} w_0 \\ w \end{bmatrix}$$

$$y_i = \begin{cases} m/m_1 & \text{if } x_i \in \omega_1 \\ -m/m_2 & \text{if } x_i \in \omega_2 \end{cases} \quad -(1)$$

Choosing the label class as defined in eq (1) links the Least Square to fisher descriminant.

For Least Square error method we know

$$z^t z a = z^t y \quad -(2)$$ where y is label class

writing eq ① in term of partitioned matrix

$$\begin{bmatrix} 1_1^t & 1_2^t \\ x_1^t & -x_2^t \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ -1_2 & -x_2 \end{bmatrix} \begin{bmatrix} w_0 \\ w \end{bmatrix} = \begin{bmatrix} 1_1^t & -1_2^t \\ x_1^t & -x_2^t \end{bmatrix} \begin{bmatrix} \frac{m}{m_1} & 1_1 \\ -\frac{m}{m_2} & 1_2 \end{bmatrix} \quad (3)$$

we know sample mean u_i & the pooled scatter matrix S_w are given as

$$u_i = \frac{1}{m_i} \sum_{x \in D_i} x \qquad i = 1, 2 \ldots$$

$$\& \; S_w = \sum_{i=1}^{2} \sum_{x \in D_i} (x - u_i)(x - u_i)^t$$

we can multiple eq ③ matrices

$$\begin{bmatrix} m & (m_1 u_1 + m_2 u_2)^t \\ (m_1 u_1 + m_2 u_2) & (S_w + m_1 u_1 u_1^t + m_2 u_2 u_2^t) \end{bmatrix} \begin{bmatrix} w_0 \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ m(u_1 - u_2) \end{bmatrix}$$

This can be viewed as a pair of equation, the first of which can be solved for w_0 in term of w

$$w_0 = -u^t w \quad -(4)$$

$u \rightarrow$ mean of all sample

Substituting in eq ② obtained from matrix & few algebraic manipulation

$$\left[\frac{1}{m} S_w + \frac{m_1 m_2}{m^2} (u_1 - u_2)(u_1 - u_2)^t \right] w = u_1 - u_2 \quad -(5)$$

because the vector $(u_1 - u_2)(u_1 - u_2)^t$ is the direction of $(u_1 - u_2)$ for any value of w_1, we can write

$$\frac{m_1 m_2}{m^2} (u_1 - u_2)(u_1 - u_2)^t w = (1 - \alpha)(u_1 - u_2)$$

where α is some scalar

then eq ⑤ yields

$$w \; \alpha \; m \, S_w^{-1} (u_1 - u_2) \leftarrow (6)$$

$S_n \ldots \quad J(w) \ldots$

eq ⑥ except for an ~~ot~~ unimportant scale factors is identical to the solution for fishers Linear Descriminant. In addition, we obtain the threshold weight ω_0 & decision required

$$\begin{cases} \omega_1, & \text{if } \omega^t (x - \mu) > 0 \\ \omega_2, & \text{otherwise.} \end{cases}$$

——————x——————x——————