Date: / /

ASSIGNMENT-2

180050009
Anjali Priya

1. Prior on weight

$$P(w) = \frac{1}{2b} \exp\left(-\frac{|w-\mu|}{b}\right)$$

$$f(w) = \frac{1}{2}(y - Xw)^T(y - Xw) + \lambda \|w\|_1$$

so $P(w|D) = \frac{P(D|w)\,P(w)}{c}$ where c is constant

posterior will also follow laplace distribution

$$P(D|\mu) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^N \exp\left(\frac{-1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2 + \frac{|\mu-\mu_0|(\sqrt{2})}{\sqrt{\sigma_0^2}}\right)$$

where σ is from gaussian distribution &
σ_0 is from laplace distribution

$$\frac{\sqrt{2}}{\sqrt{\sigma_0^2}}(\mu-\mu_0) + \left(\frac{-1}{2\sigma^2}\right)\sum_{i=1}^{m}(x_i-\mu)^2 = 0$$

$$\mu_0 = \frac{\sigma^2\sqrt{2}}{\sigma_0(m)} + \frac{1}{m}\sum_{i=1}^{m} x_i$$

for LASSO, MAP estimate

$$\sum_{i=1}^{N}(x_i-\mu)^2 + \frac{\lambda|\mu|}{N}$$

$$= \sum_{i=1}^{m}(x_i-\mu)(-1) + \frac{\lambda}{m} = 0, \quad \text{if } \mu > 0$$

so $\mu = \frac{1}{m}\sum_{i=1}^{N} x_i - \frac{\lambda}{m}$

both have same format where $\lambda = \frac{\sigma^2\sqrt{2}}{\sigma_0}$

— —

2. One vs rest for multi class classification is a method for using binary classification algorithms for multi-class.
While One-vs-One splits a multi class classification dataset into binary classification problems
For ex- if we have problem with three classes:
red, blue, yellow
then acc to one vs rest, we will have classification problem as -

-> blue vs [red, yellow]
-> red vs [blue, yellow]
-> yellow vs [blue, red] but acc to one vs one we have 3 classification problems as.
-> red vs yellow -> red vs blue
-> yellow vs blue

One vs rest approach takes one class as +ve & rest all as negative while one vs one considers each binary pair of classes.

One-vs-rest-

- Advantage - It ~~ba~~ trains less no. of classifier and hence is faster overall & is usually preferred as it is faster so it is generally used for predicting scores or numerical class
- Disadvantage - for using this approach, we should create a model for each class and it can be challenging when dealing with large sets of data with million rows.

One vs. One approach

- Advantage - This model is best to support SVMs & associated kernel based classification algos because these methods don't scale in proportion to the size of the training dataset.
- Disadvantage - this approach is less prone to imbalance in dataset due to dominance of particular classes.

3.a) Increasing the value of λ in lasso regression-

⇒ bias will increase & variance will decrease

Reason - increasing λ will make several w components shrink so this will increase bias while variance decreases as model won't show much fluctuation by flattening out of w's

b) Adding higher number of training example for perceptron

⇒ bias increases & variance will decrease.

On increasing no. of training example, sensitivity to small fluctuation will decrease & but it may lead to miss relevant relation between feture & outputs.

c) variance will increase, bias will decrease.

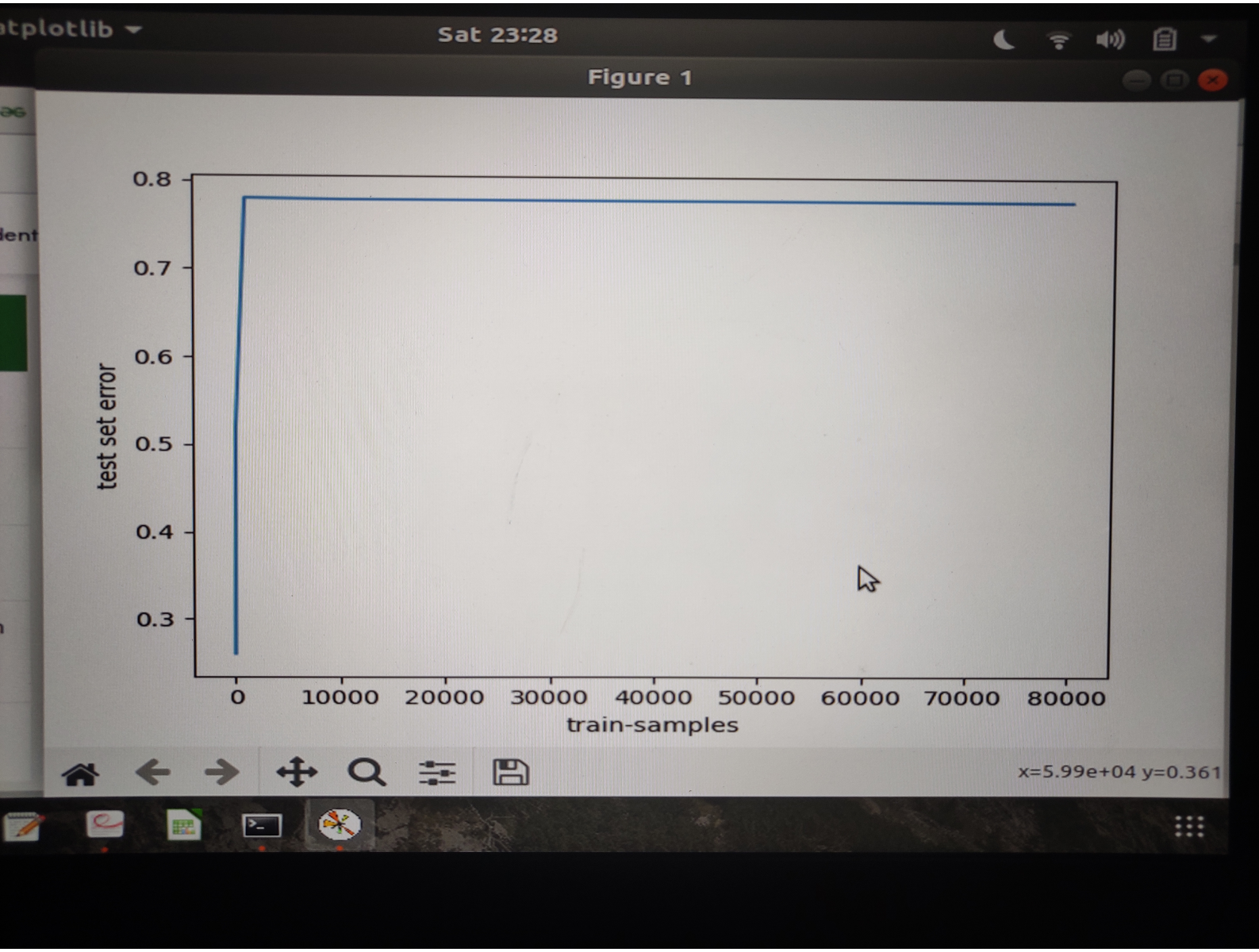
4 test error using the following samples.

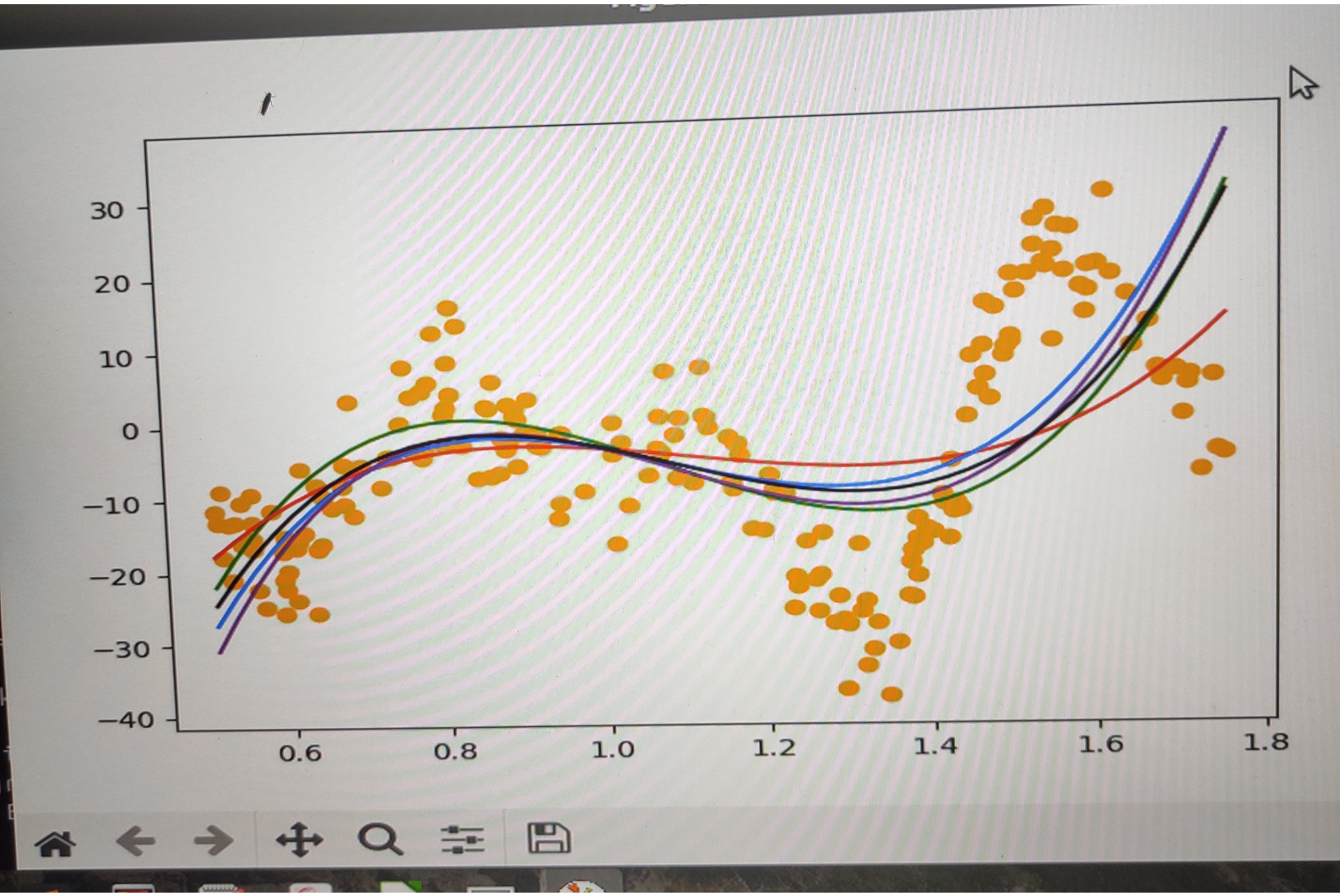
we need to plot test set error vs no of train samples.

$X = [10, 100, 1,000, 10,000, 50000, 80000]$

test error = [0.2633, 0.52175, 0.78025, 0.77875, 0.77905, 0.77905]

On increasing size of train samples test set error increases and plot is like first it increases drastically & then remains constant.


tplotlib
Sat 23:28
Figure 1
0.8
0.7
0.6
0.5
0.4
0.3
test set error
0
10000
20000
30000
40000
50000
60000
70000
80000
train-samples
x=5.99e+04 y=0.361


30
20
10
0
−10
−20
−30
−40
0.6
0.8
1.0
1.2
1.4
1.6
1.8