# STATS205 Autumn 2019 Homework 1

SUNet ID: raunakbh
Name: Raunak Bhattacharyya
Collaborators: [list all the people you worked with]

By turning in this assignment, I agree by the Stanford honor code and declare that all of this is my own work.

## Problem a.

### Glass dataset

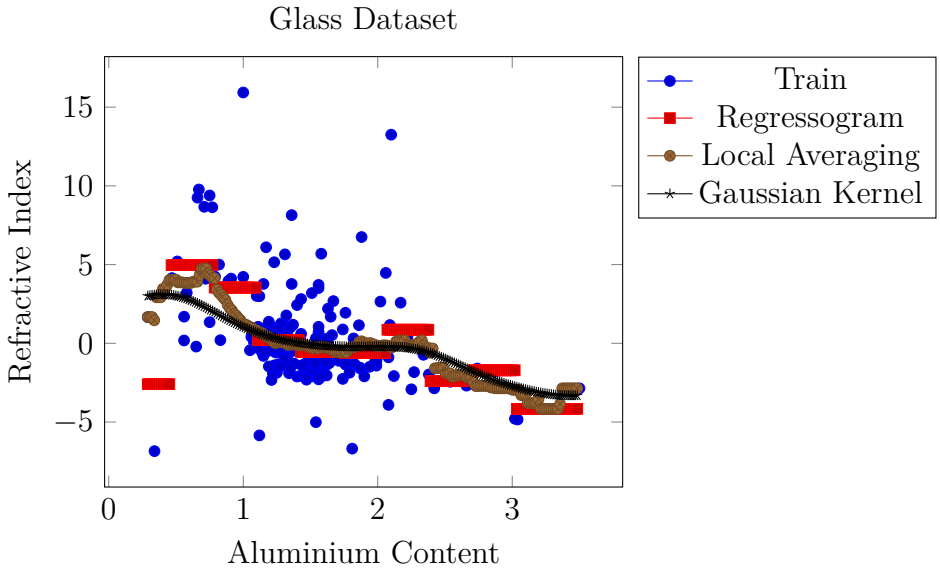



Figure 1: Glass dataset

## Motorcycle dataset

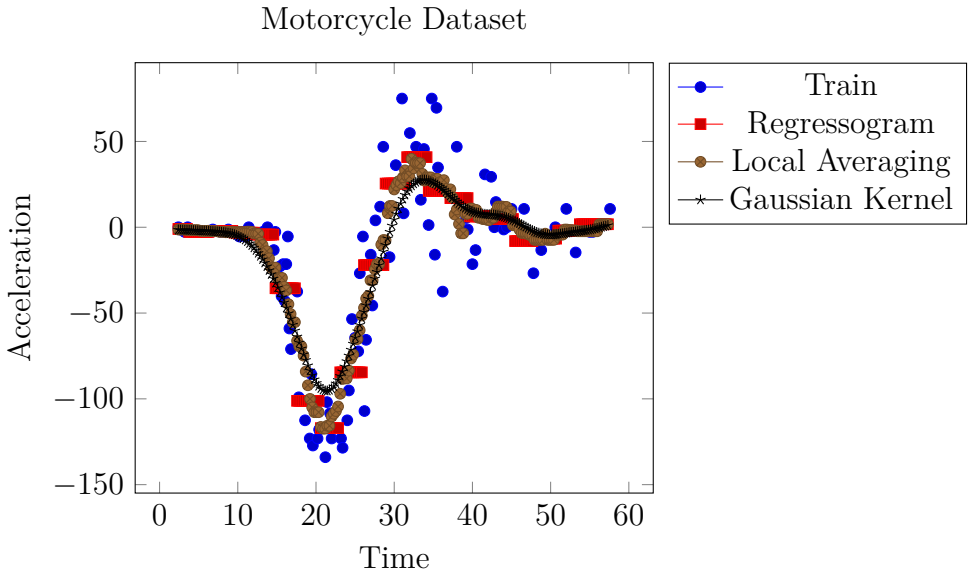



Figure 2: Motorcycle dataset

# Problem b. Understand bias-variance tradeoff

## Part 1

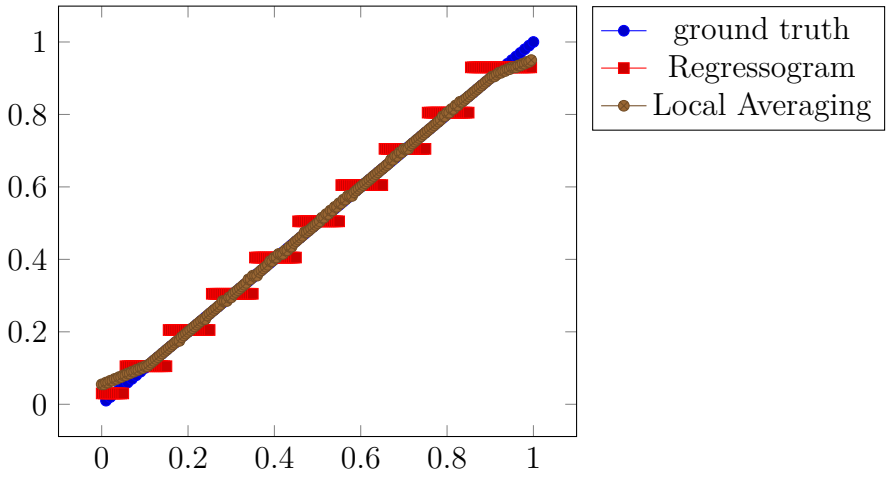



Figure 3: Regressogram vs local averaging for b.1

We see that the local averaging has a lower bias as shown by closeness of the predicted output to the ground truth. The reason is that local averaging does not force all covariates in the same bin to output the same value and allows individual points to have individual outputs.

## Part 2

| Analytical | Sampling |
|---|---|
| 0.059 | 0.053 |

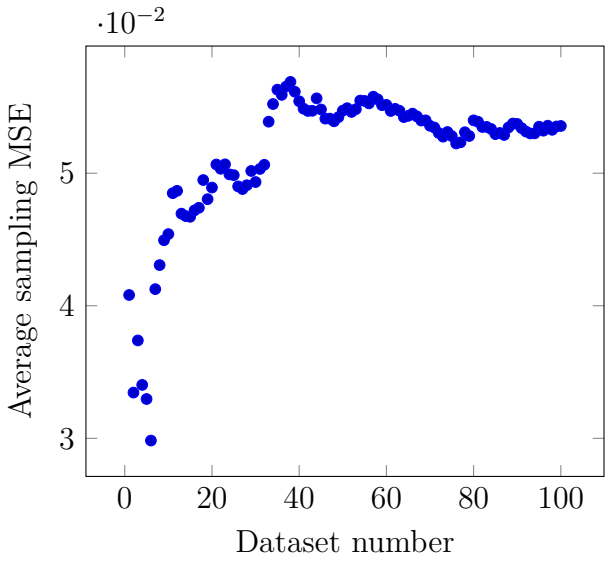
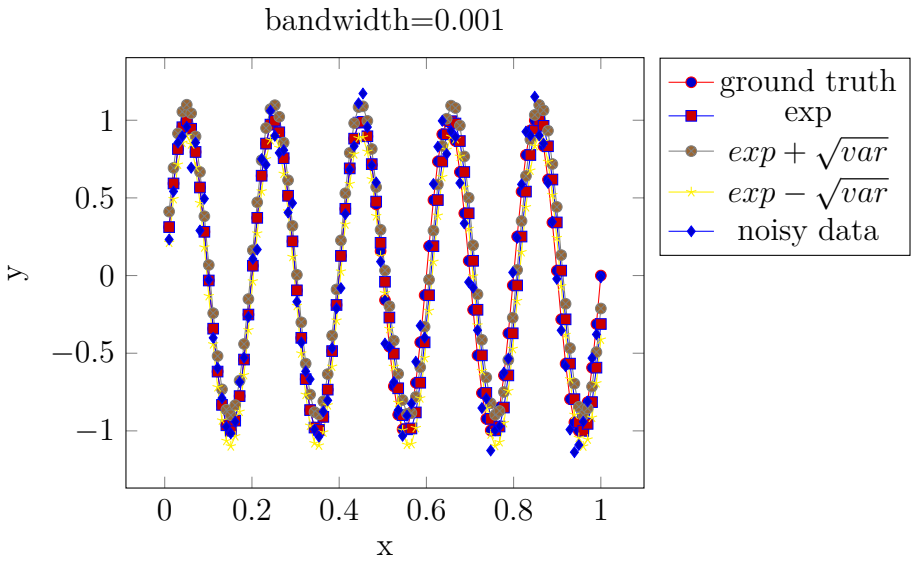
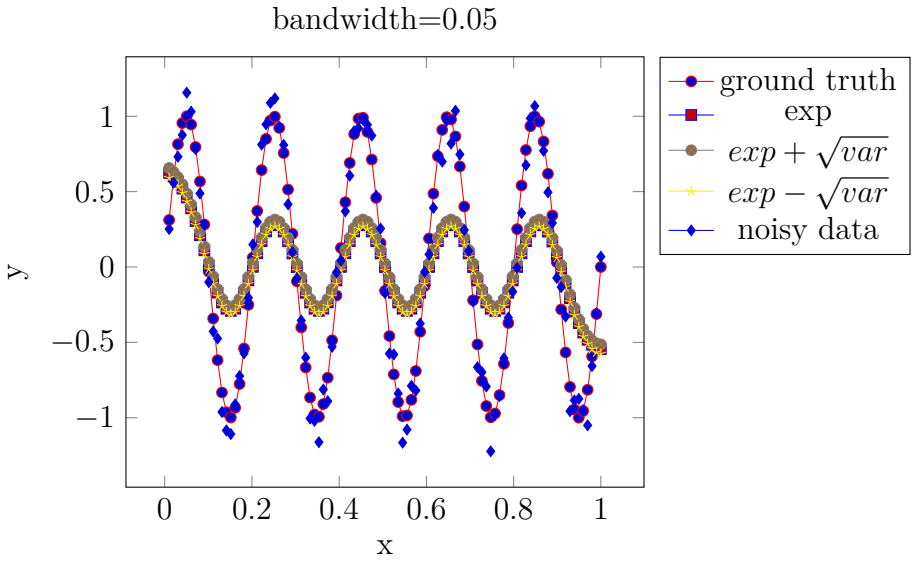


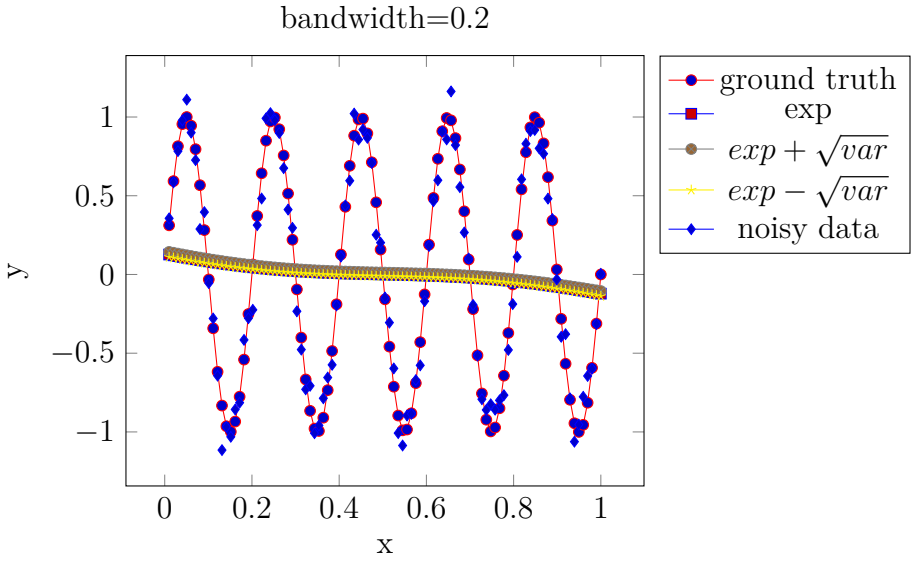

Figure 4: Average expected square loss with increasing number of training datasets.

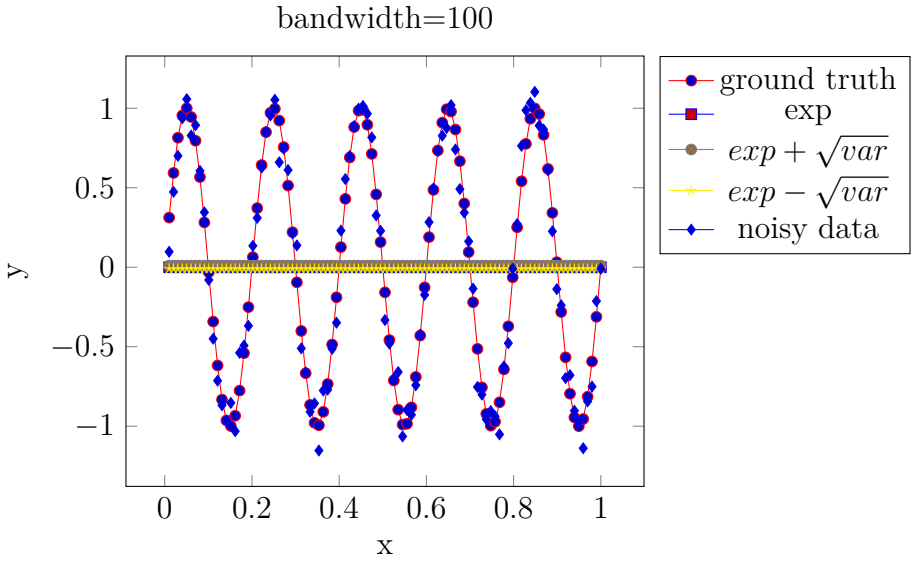
## Part 3



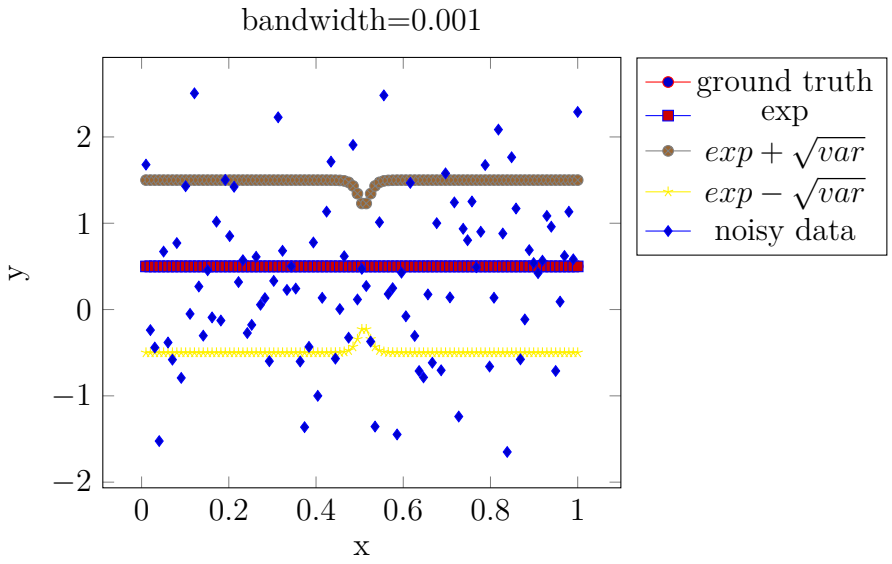



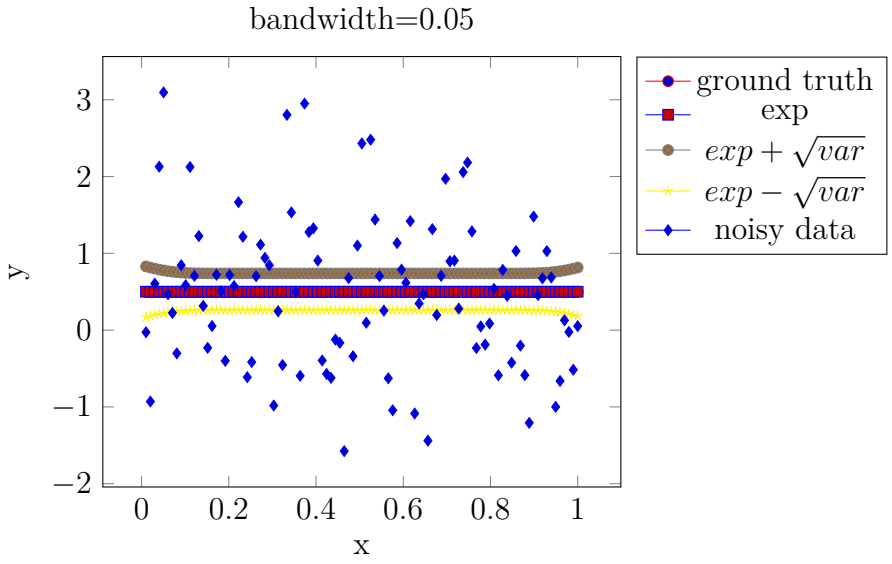

From the plots, we see that the lowest bandwidth, $h = 0.001$ yields the lowest bias.


bandwidth=0.2
ground truth
exp
$exp + \sqrt{var}$
$exp - \sqrt{var}$
noisy data
y
x
1
0.5
0
−0.5
−1
0
0.2
0.4
0.6
0.8
1

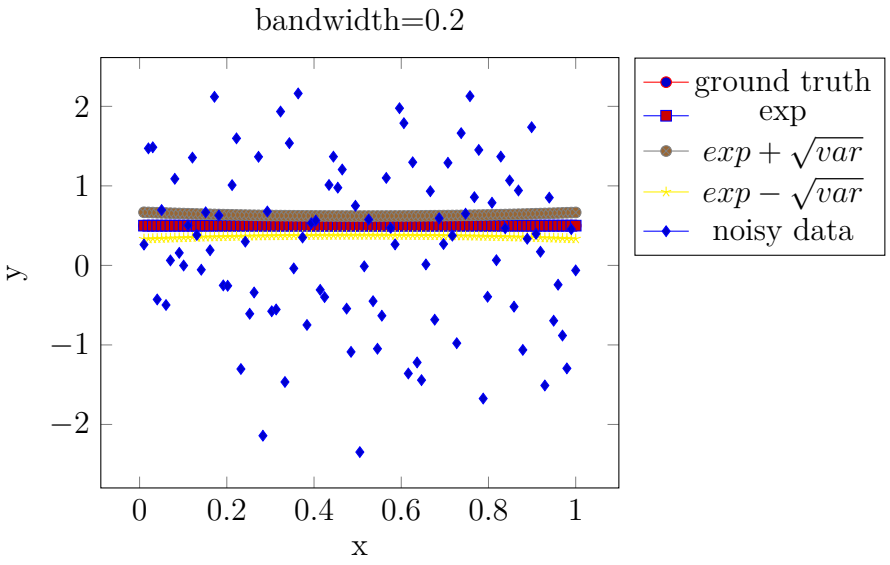

bandwidth=100
ground truth
exp
$exp + \sqrt{var}$
$exp - \sqrt{var}$
noisy data
y
x
1
0.5
0
−0.5
−1
0
0.2
0.4
0.6
0.8
1

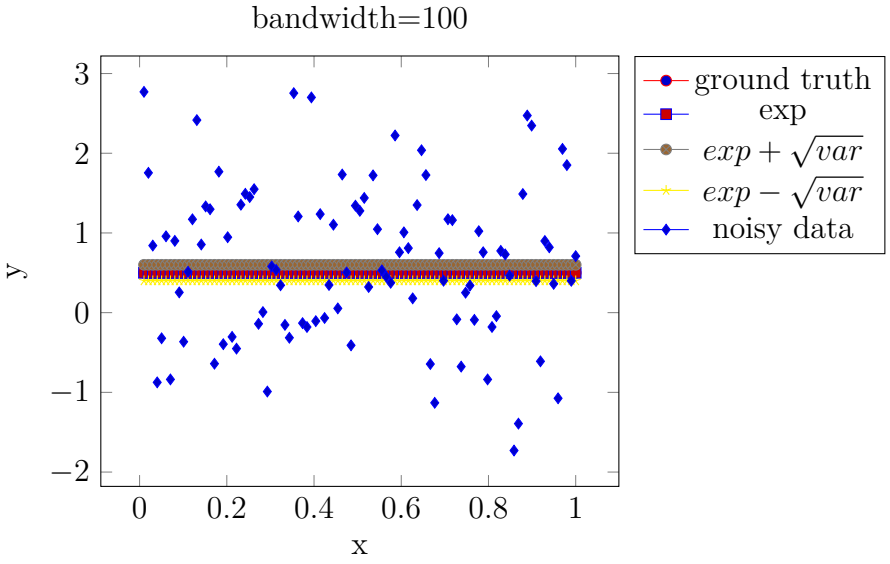
## Part 4




From the plots, we see that the highest bandwidth, $h = 100$ yields the lowest bias.


bandwidth=0.2
ground truth
exp
$exp + \sqrt{var}$
$exp - \sqrt{var}$
noisy data
y
x
2
1
0
−1
−2
0
0.2
0.4
0.6
0.8
1


bandwidth=100
ground truth
exp
$exp + \sqrt{var}$
$exp - \sqrt{var}$
noisy data
y
x
3
2
1
0
−1
−2
0
0.2
0.4
0.6
0.8
1

## Part 5

| $n$ | $h_n$ |
|---|---|
| 5 | 0.19 |
| 20 | 0.12 |
| 80 | 0.08 |
| 320 | 0.063 |
| 1280 | 0.048 |