

## CS246: Mining Massive Data Sets

**Assignment number: 3**

Fill in and include this cover sheet with each of your assignments. It is an honor code violation to write down the wrong time. Assignments and code are due at 5:00 PM on Scoryst and SNAP respectively. Failure to include the coversheet with your assignment will be penalized by 2 points. Each student will have a total of *two* free late periods. *One late period expires at the start of each class.* (Assignments are due on Thursdays, which means the first late period expires on the following Tuesday at 5:00 PM.) Once these late periods are exhausted, any assignments turned in late will be penalized 50% per late period. However, no assignment will be accepted more than one late period after its due date. (If an assignment is due to Thursday then we will not accept it after the following Thursday.)

**Your name:** Priyank Mathur

**Email:** priyankm@stanford.edu **SUNet ID:** priyankm

Collaborators: Emrah Budur, Shundan Xiao, Dibyajyoti Ghosh, Arkajyoti Misra

I acknowledge and accept the Honor Code.

(Signed) Priyank Mathur

### Answer to Question 1.a

$$\epsilon_{iu} = 2 \times (R_{iu} - q_i \cdot p_u^T)$$

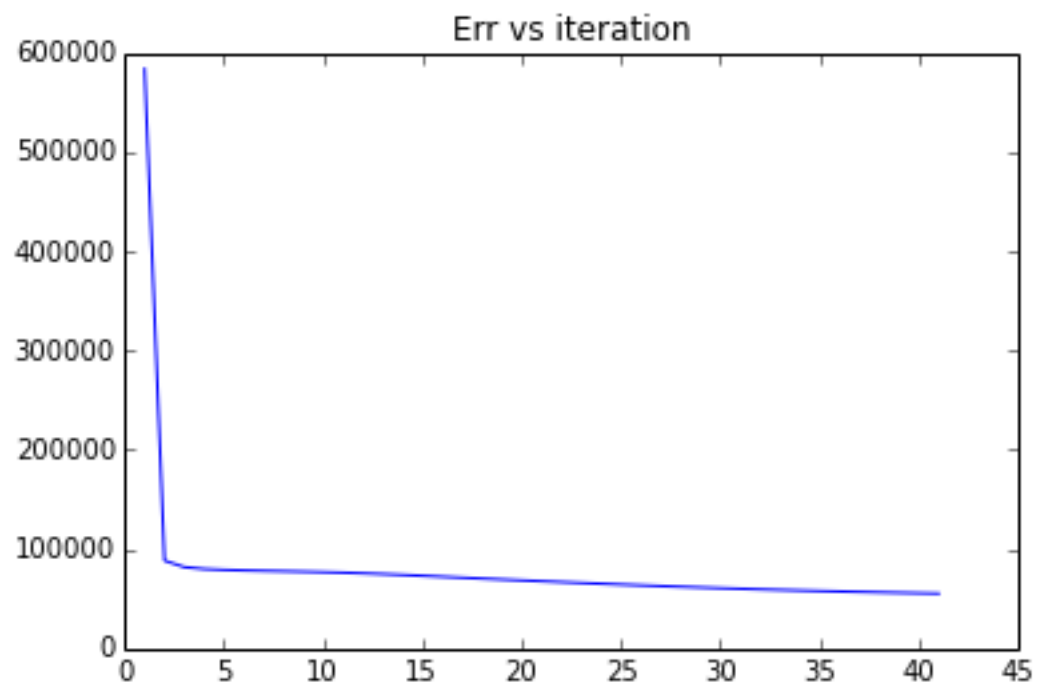
$$q_i = q_i + \eta(\epsilon_{iu} p_u - \lambda q_i)$$

$$p_u = p_u + \eta(\epsilon_{iu} q_i - \lambda p_u)$$

Note that the factor 2 obtained during differentiation w.r.t  $p_u$  and  $q_i$  has been consumed within the learning rate  $\eta$ .

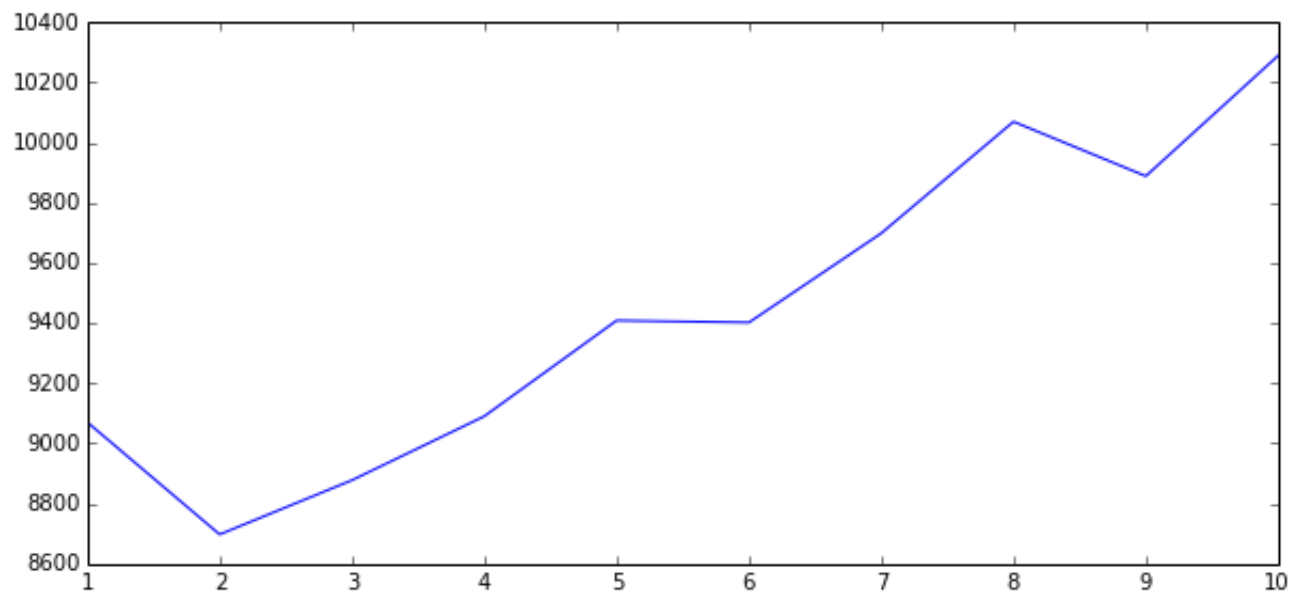
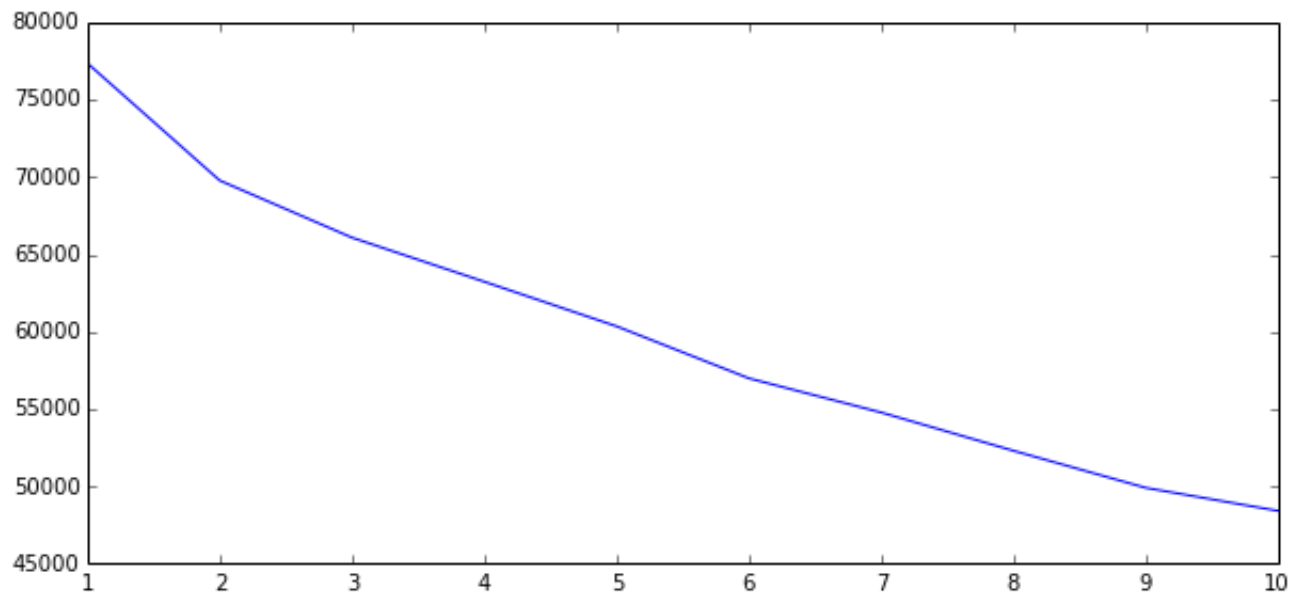
## Answer to Question 1.b

$$\eta = 0.005$$



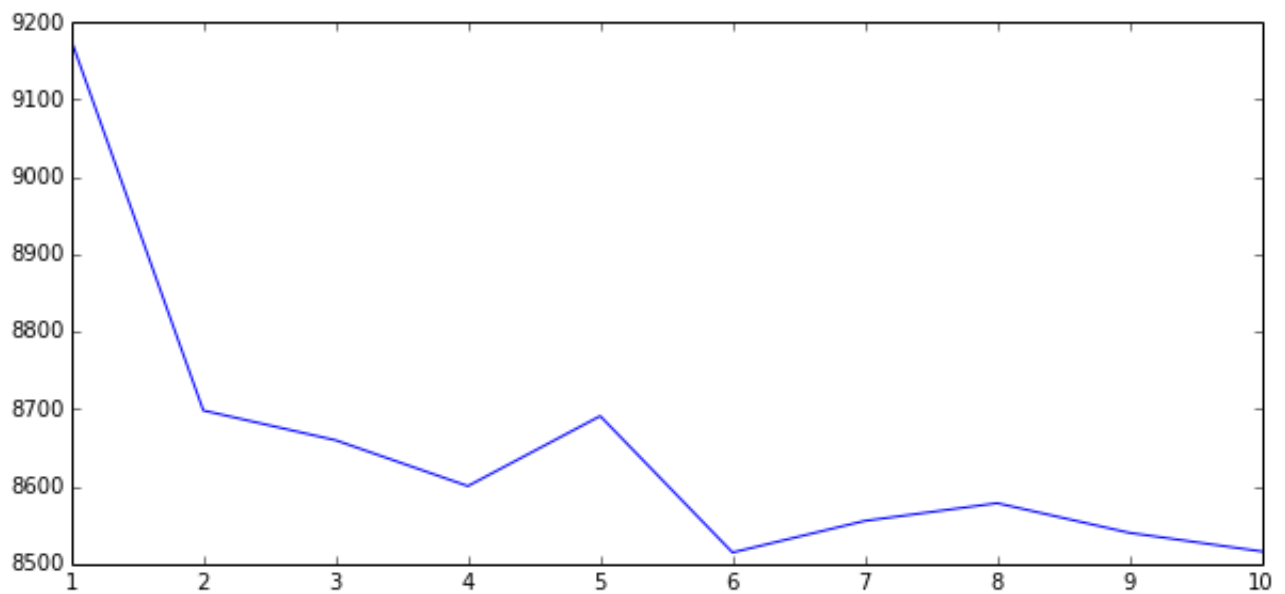
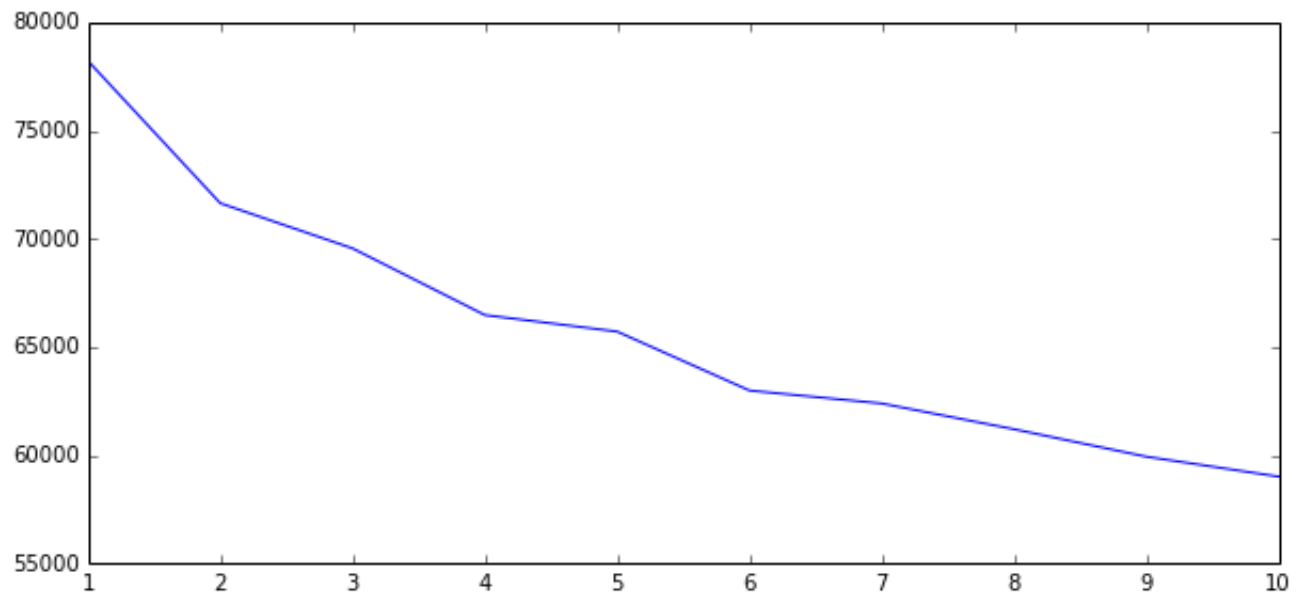
## Answer to Question 1.c

$Error_{tr}$  vs.  $k$  for  $\lambda = 0$



$Error_{te}$  vs.  $k$  for  $\lambda = 0$

$Error_{tr}$  vs.  $k$  for  $\lambda = 0.2$



$Error_{te}$  vs.  $k$  for  $\lambda = 0.2$

Based on these graphs, we find the following to be true -

- B: Regularization decreases the test error for  $k \geq 5$
- D: Regularization increases the training error for all (or almost all)  $k$
- H: Regularization decreases overfitting

### Answer to Question 1.d

$$\epsilon_{iu} = 2 \times (R_{iu} - (\mu + b_u + b_i + q_i \cdot p_u^T))$$

$$q_i = q_i + \eta_{LF}(\epsilon_{iu}p_u - \lambda q_i)$$

$$p_u = p_u + \eta_{LF}(\epsilon_{iu}q_i - \lambda p_u)$$

$$b_{i_i} = b_{i_i} + \eta_{b_i}(\epsilon_{iu} - \lambda b_{i_i})$$

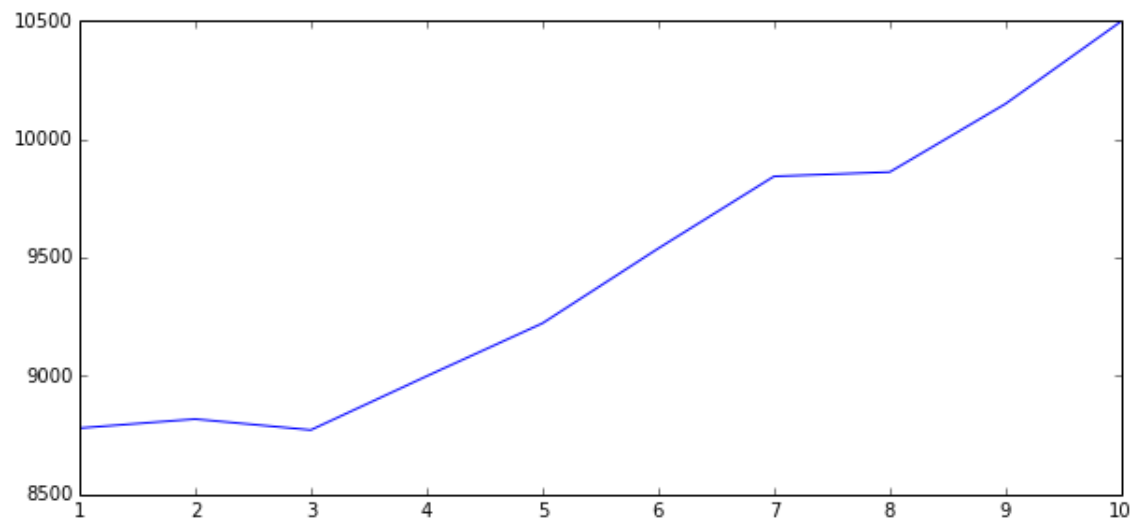
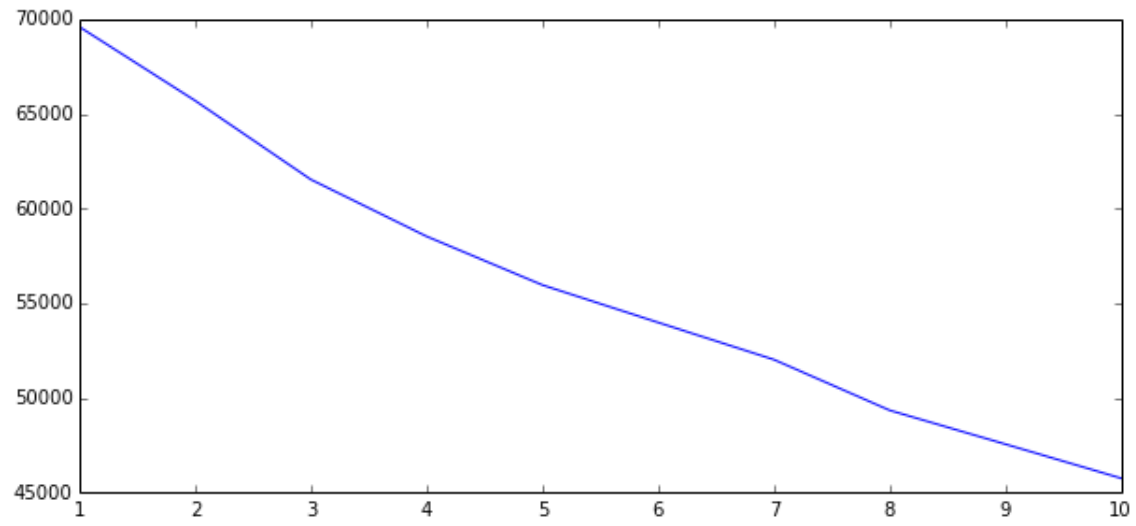
$$b_{u_u} = b_{u_u} + \eta_{b_u}(\epsilon_{iu} - \lambda b_{u_u})$$

$$\eta_{LF} = 0.005$$

$$\eta_{b_i} = 0.01$$

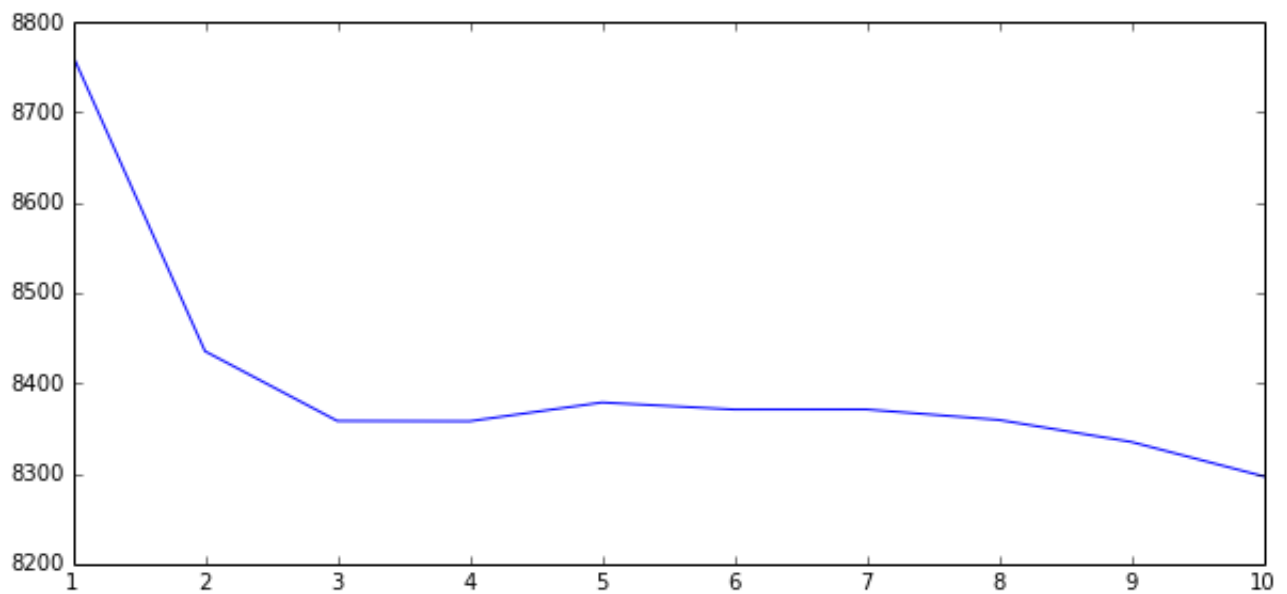
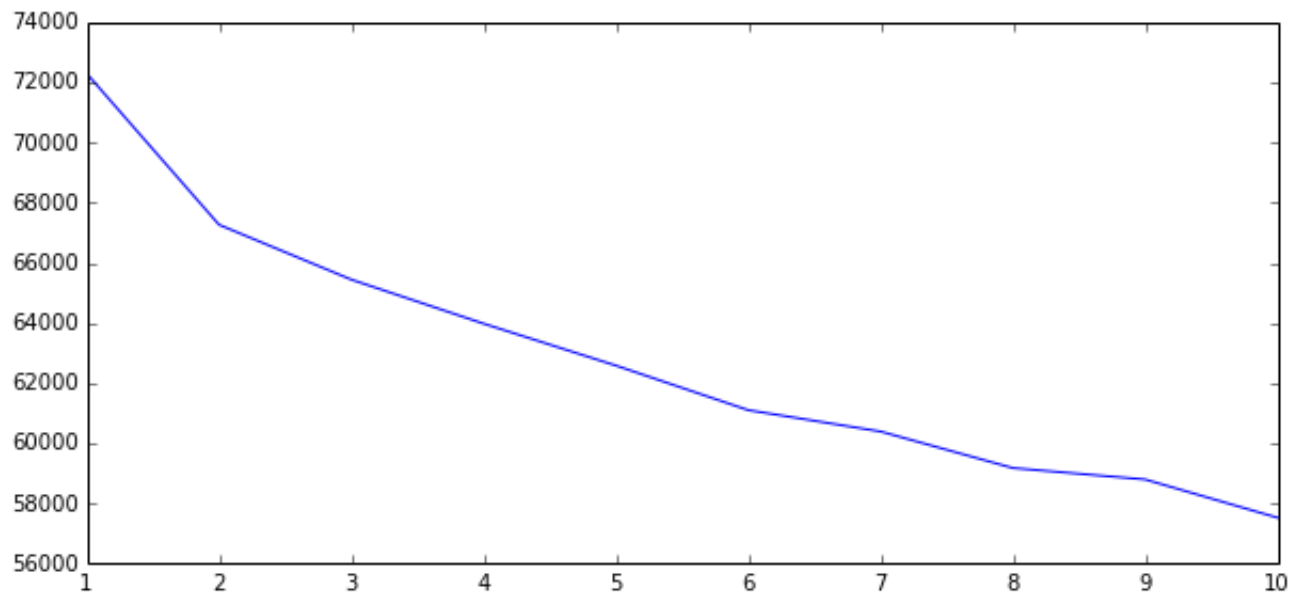
$$\eta_{b_u} = 0.01$$

$Error_{tr}$  vs.  $k$  for  $\lambda = 0$



$Error_{te}$  vs.  $k$  for  $\lambda = 0$

$Error_{tr}$  vs.  $k$  for  $\lambda = 0.2$



$Error_{te}$  vs.  $k$  for  $\lambda = 0.2$

Based on these graphs, we find the following to be true -

- B: Regularization decreases the test error for  $k \geq 5$
- D: Regularization increases the training error for all (or almost all)  $k$
- H: Regularization decreases overfitting



## Answer to Question 2a

## Answer to Question 2b

## Answer to Question 2c

## Answer to Question 2d

Runtime for 40 iterations of power iteration - 791  $\mu s$ .

Runtime for Monte Carlo with  $R = 1$  - 11  $\mu s$ .

Runtime for Monte Carlo with  $R = 3$  - 24  $\mu s$ .

Runtime for Monte Carlo with  $R = 5$  - 34.7  $\mu s$ .

| R | K  | Error           |
|---|----|-----------------|
| 1 | 10 | 0.0040539431549 |
| 3 | 10 | 0.0024193828540 |
| 5 | 10 | 0.0019164964406 |
| 1 | 30 | 0.0024580176469 |
| 3 | 30 | 0.0016139479244 |
| 5 | 30 | 0.0011990209574 |
| 1 | 50 | 0.0018569538811 |
| 3 | 50 | 0.0012340305464 |
| 5 | 50 | 0.0009915055978 |

### Answer to Question 3.a

## Answer to Question 3.b

### Answer to Question 3.c

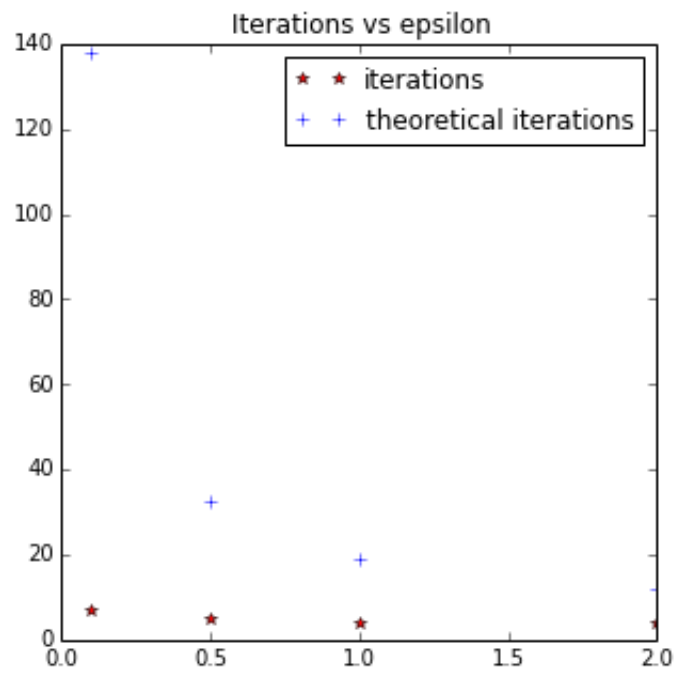
## Answer to Question 4a



## Answer to Question 4b

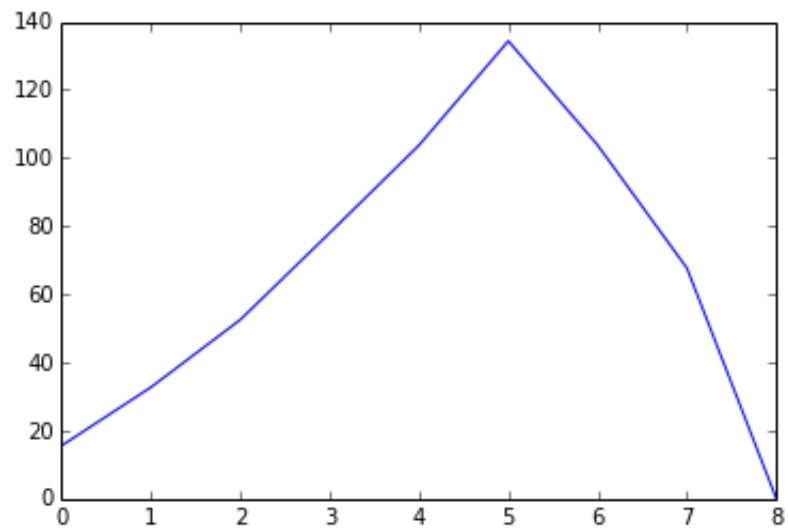
## Answer to Question 4c

1

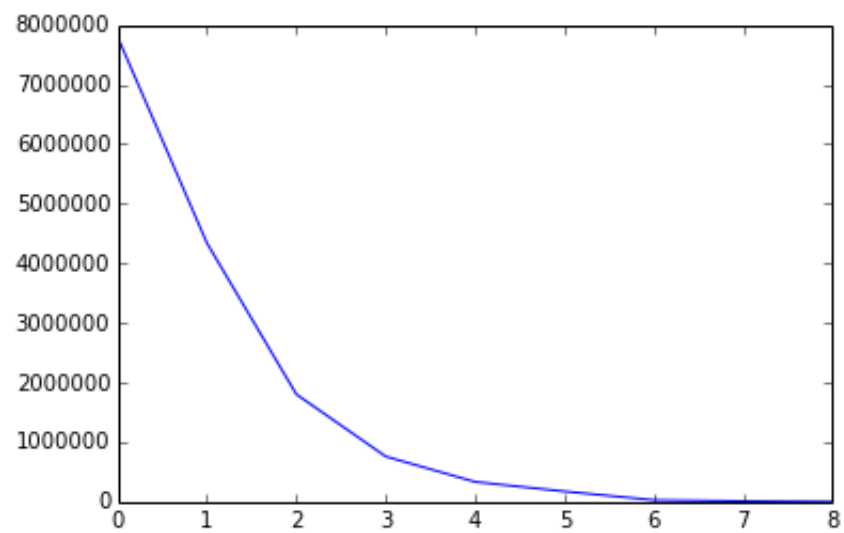


| $\epsilon$ | Iterations | Theoretical iterations |
|------------|------------|------------------------|
| 0.1        | 7          | 137.67                 |
| 0.5        | 5          | 32.36                  |
| 1          | 4          | 18.93                  |
| 2          | 4          | 11.94                  |

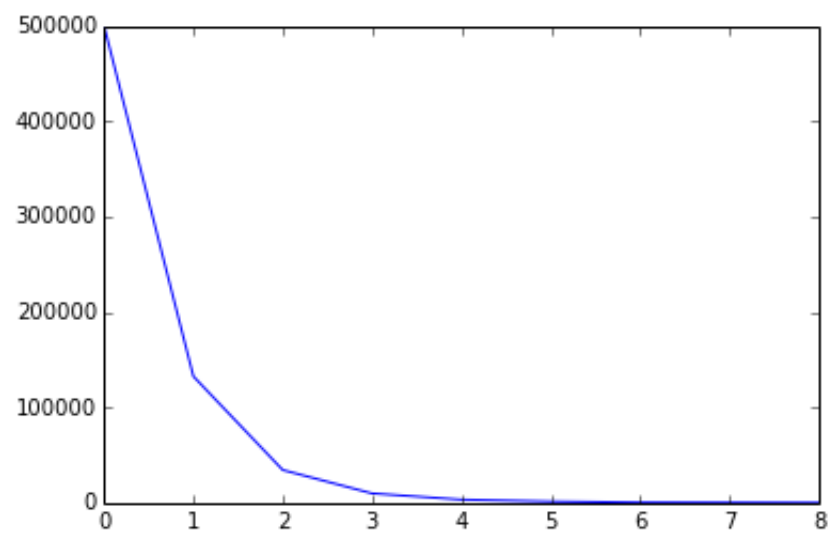
Density ( $\rho(S_i)$ ) vs iteration



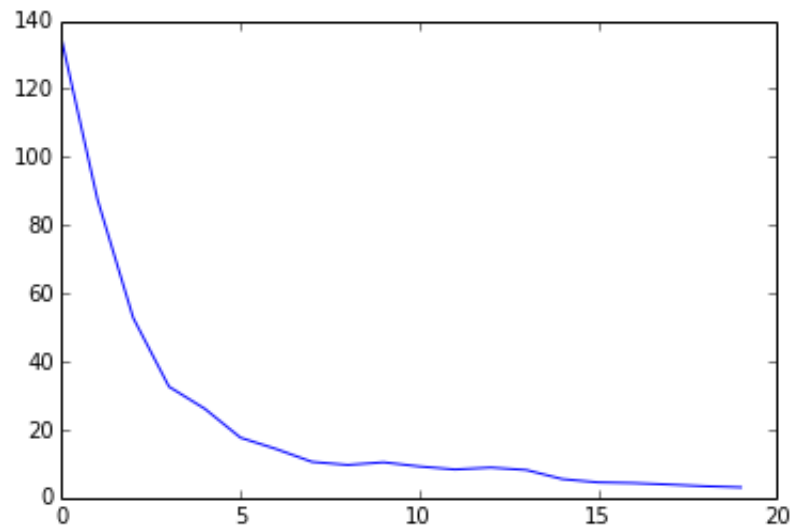
$|E(S_i)|$  vs iteration



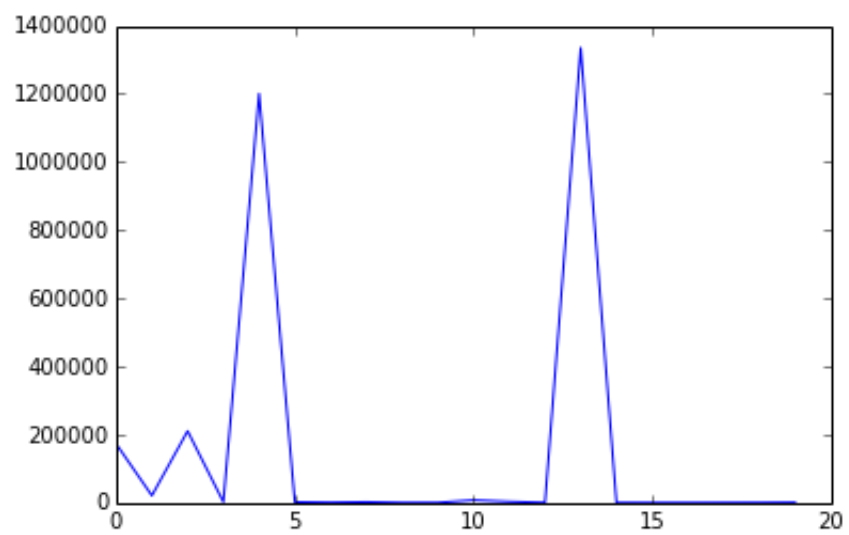
$|S_i|$  vs iteration



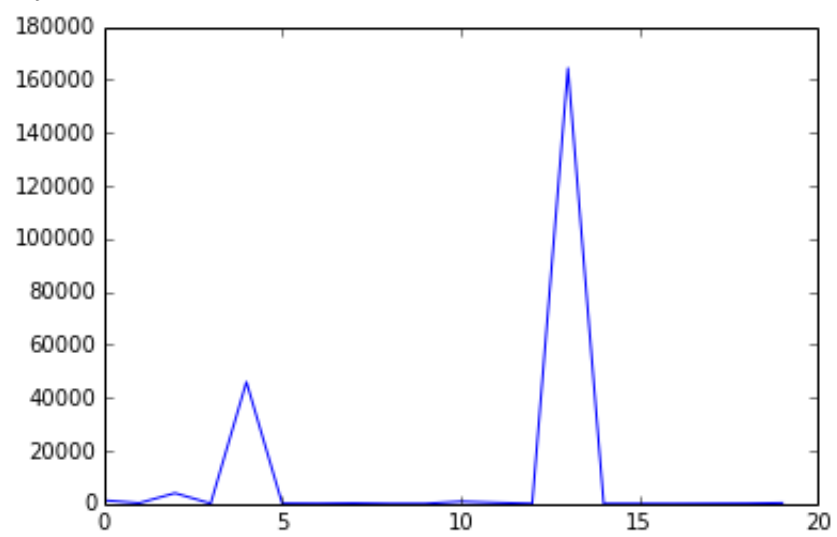
Density ( $\rho(\bar{S}_j)$ ) vs iteration



$|E(\bar{S}_j)|$  vs iteration



$|\bar{S}_j|$  vs iteration



Code for Q1

Code for Q2



Code for Q4