Incremental Gradient Descent

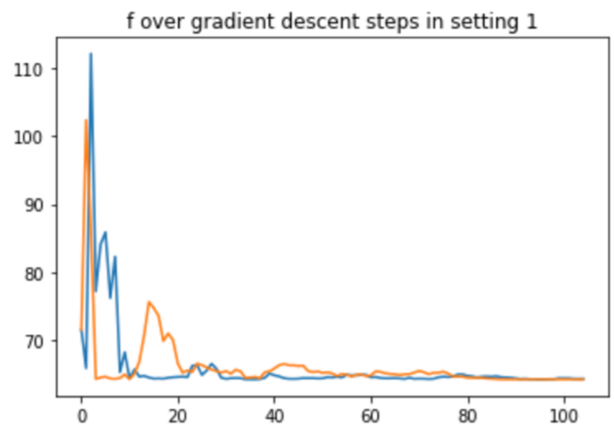
Meng Xu, Jiaxiang Peng

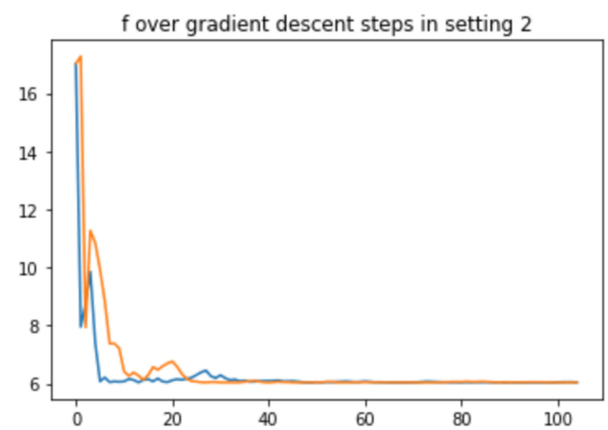
Task 1:

We run the GD functions once, the final x values are as follows, respectively:

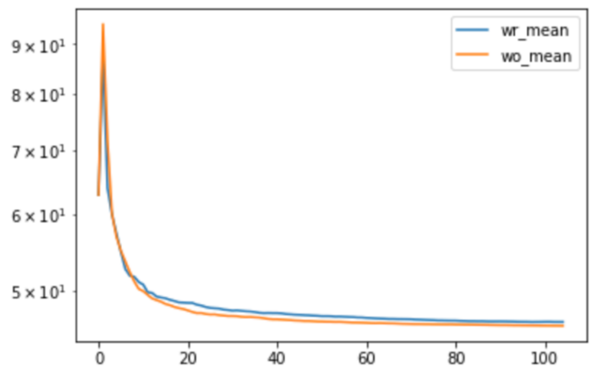
```
final x value w/ replacement in setting 1 is  0.5754400766013175
final x value w/o replacement in setting 1 is 0.5666535136867069
final x value w/ replacement in setting 2 is  0.570752244996383
final x value w/o replacement in setting 2 is 0.5529313839183578
```

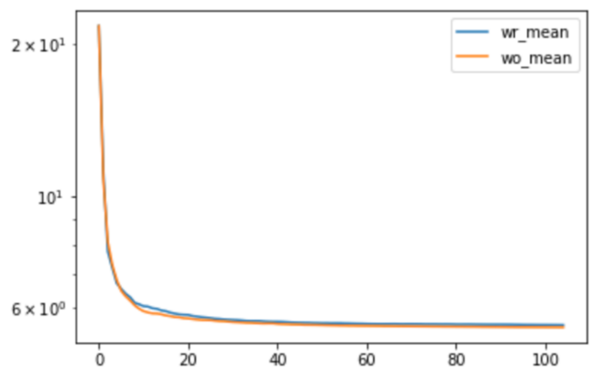
And the histories of objective function are as following, with blue lines indicating the with-replacement ordering and orange lines indicating the without-replacement ordering:





Run the GD for 200 times and take the average of the histories of the objective function and then take log:





As we can see, the without-replacement ordering is better since the final x value is closer to the true solution and the final f value is smaller.

The IGD_wo_task1 must converge to the true solution (mean value of y).

Brief Proof:

$$x_k = x_{k-1} - \gamma_{k-1}(x_{k-1} - y_{i_{k-1}})$$

$$x_k = x_{k-1} - \frac{1}{k}(x_{k-1} - y_{i_{k-1}})$$

$$x_k = \frac{kx_{k-1} - (x_{k-1} - y_{i_{k-1}})}{k}$$

$$x_k = \frac{(k-1)x_{k-1} + y_{i_{k-1}}}{k}$$

$$kx_k = (k-1)x_{k-1} + y_{i_{k-1}}$$

$$n x_n = (n-1) x_{n-1} + y_{i_{n-1}}$$

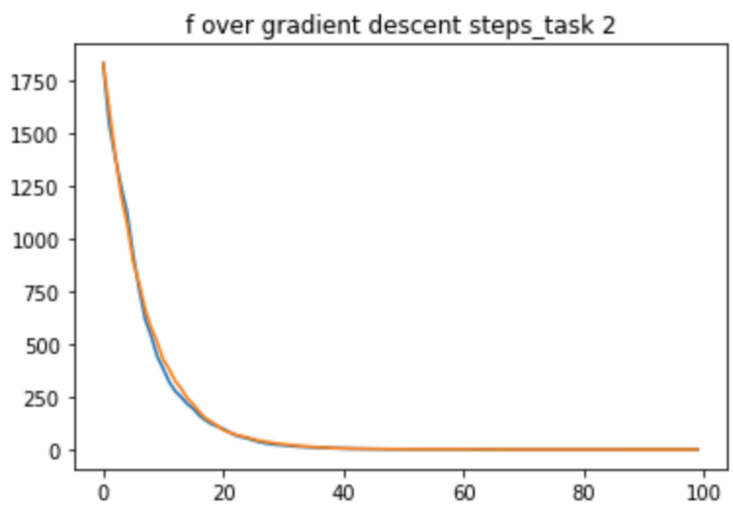
Since we ordered without placement, we have $i_k \neq j_k$ for $i \neq j$, so in each iteration, we get different y_i. Thus after we loop through k=1, k=2, …, to k=n, we would have summed up each y_i value, and the optimal solution for x is

$$x_n = \frac{1}{n} \sum_{i=1}^{n} y_i = \mu,$$

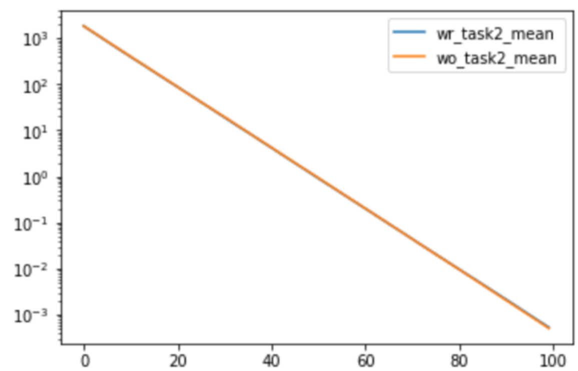
which is the mean value of y.

Task 2:

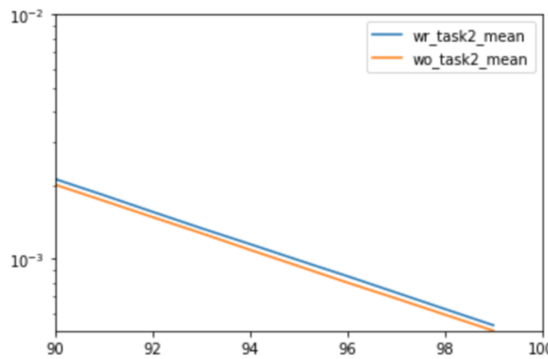
Running the GD once gives: (with blue line representing the with-replacement ordering and orange line the without-replacement ordering)



Running the GD for 200 times and taking average of the histories of the objective function and taking log gives:



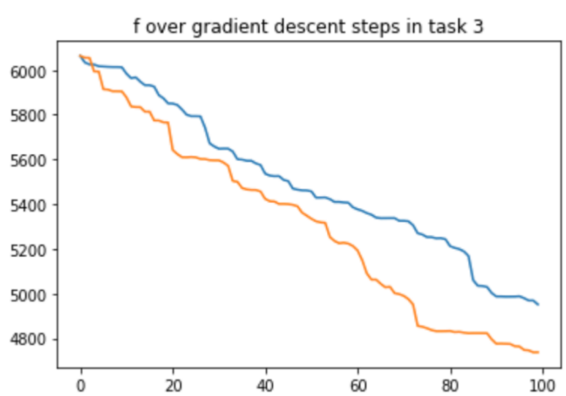
If we zoom in at the tail:



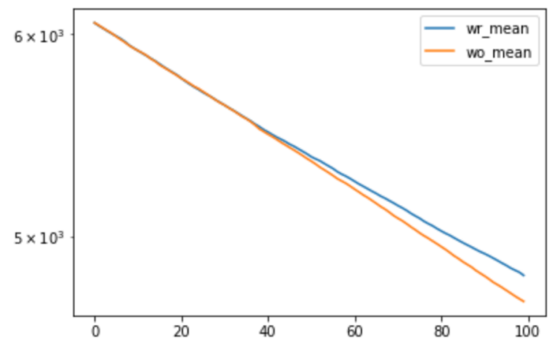
As we can see, the without-replacement ordering is better.

Task 3:

Running the GD once gives: (with blue line representing the with-replacement ordering and orange line the without-replacement ordering)

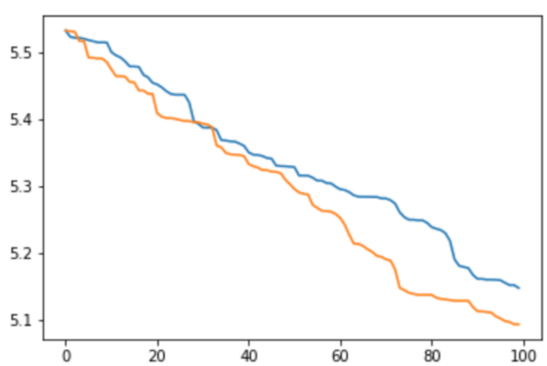


Running the GD for 200 times and taking average of the histories of the objective function and taking log gives:

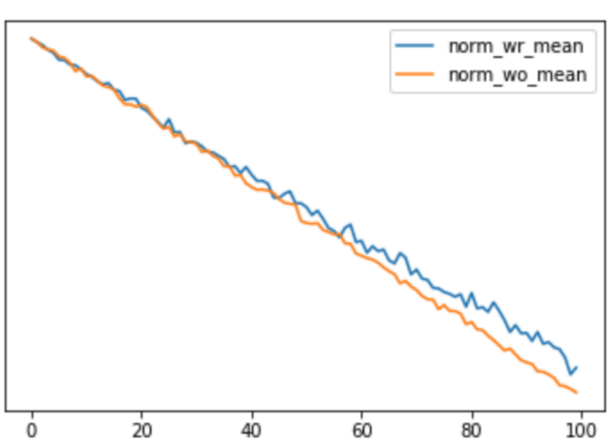


Let’s take a look at the histories of the norm $\| x_k - x^* \|$:

Running the GD once gives: (with blue line representing the with-replacement ordering and orange line the without-replacement ordering)



Running it for 50 times gives;



As we can see, the without-replacement ordering is better. The f values diminish faster and x converges faster using the without-replacement ordering.