**Homework 1 Experiment Report**

**Problem 1.**

Experiment with the step size (learning rate).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Non-Normalized Data** | | **Normalized Data** | |
| **Learning Rate** | **Training Error** | **Testing Error** | **Training Error** | **Testing Error** |
| 0.05 | 0.507 | 0.4733 | 0.0919 | 0.103 |
| 0.1 | 0.2586 | 0.2679 | 0.0893 | 0.0991 |
| 0.5 | 0.5057 | 0.472 | 0.0851 | 0.0978 |
| 1 | 0.5083 | 0.4733 | 0.0828 | 0.0932 |
| 5 | **0.2348** | **0.2386** | 0.0796 | 0.0939 |
| 10 | 0.3671 | 0.3937 | 0.0796 | 0.0945 |
| 15 | 0.5086 | 0.4732 | **0.0743** | **0.0769** |
| 20 | 0.5103 | 0.4739 | 0.0802 | 0.0821 |

Based on the experiment results, before data normalization, the best learning rate is 5. After normalizing the data set, the best learning rate is 15.

**Problem 2.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Learning Rate** | **Training Error** | **Testing Error** | **Running Time**  (second) |
| 0.00005 | 0.3182 | 0.3390 | 23.50 |
| **0.0001** | **0.2811** | **0.3018** | 21.18 |
| 0.0005 | 0.2729 | 0.3083 | 21.87 |
| 0.001 | 0.2902 | 0.3162 | 21.70 |
| 0.005 | 0.3270 | 0.3585 | 21.71 |
| 0.01 | 0.3534 | 0.3781 | 21.89 |
| 0.05 | 0.3042 | 0.3227 | 21.74 |
| 0.1 | 0.3316 | 0.3559 | 23.31 |
| 0.5 | 0.2957 | 0.3122 | 21.89 |

In each experiment, I applied 50 epochs, and set each min batch with size 10.

(Note: the performance is not stable, since the result is impact the random shuffle of data and stochastic gradient descent process.)