Task 1

|  |  |
| --- | --- |
| N | # of Validation Errors |
| 5 | 107 |
| 50 | 40 |
| 100 | 27 |
| 200 | 17 |
| 400 | 9 |
| 800 | 12 |

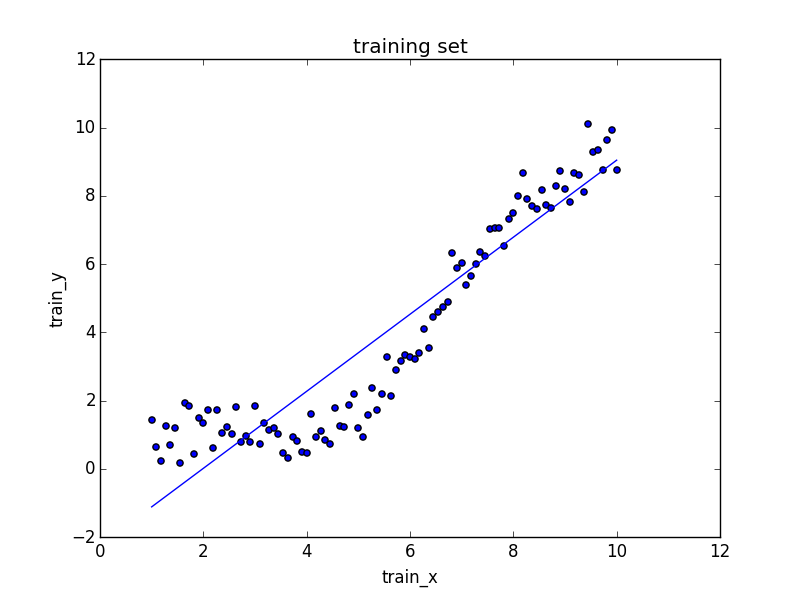
Generally a higher N value resulted in a lower number of validation errors, thus a higher N is best for performance.

Task 2

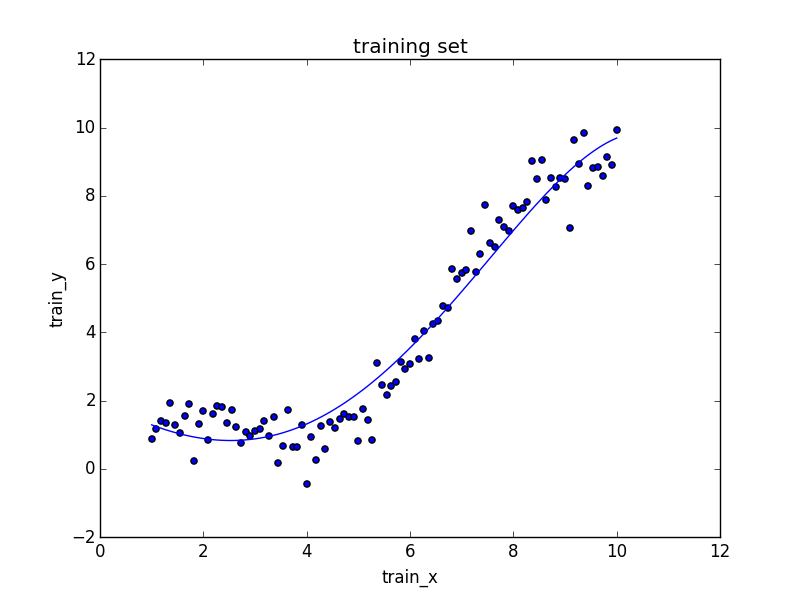
|  |  |
| --- | --- |
| K | # of Validation Errors |
| 1 | 12 |
| 3 | 8 |
| 5 | 10 |
| 7 | 9 |
| 21 | 11 |
| 101 | 24 |
| 401 | 52 |

Generally a smaller K value performed better.

Task 3



Task 4



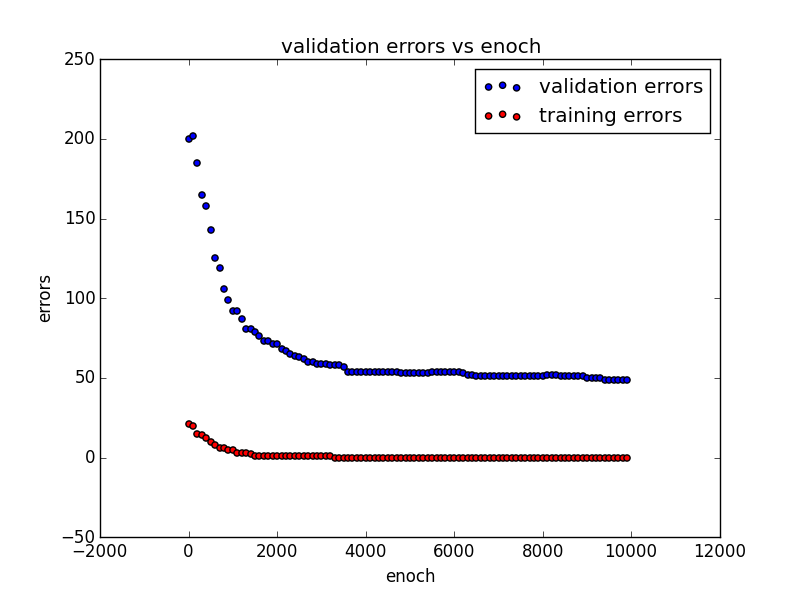
The effect of nonlinearity allows the curve to match the data a lot better than the line and hence have a lower cost.

Task 5

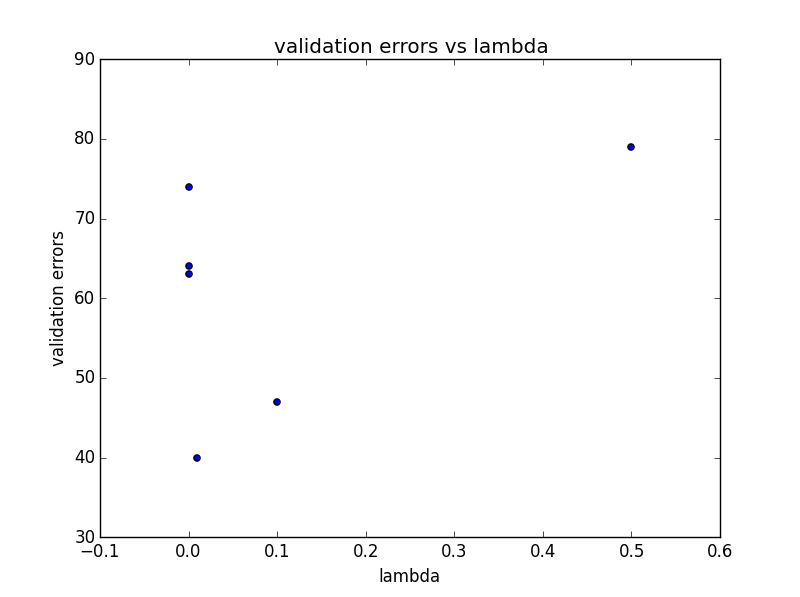
|  |  |
| --- | --- |
| N | # of validation errors |
| 100 | 72 |
| 200 | 54 |
| 400 | 37 |
| 800 | 22 |

The number of validation errors were seen to decrease as N increased, suggesting that a large N is best.

Task 6



Task 7



The weight decay parameter seemed to decrease the number of validation errors in the middle range, and an optimal value of 0.01 was found for lambda.