**Machine Learning Implementation :**

**Group Members:**

Purbasha Chatterjee – 50% (contributed to Part 1 and Part 2 with partial contribution 10% in part 3)

Meghamala Sinha – 50% (contributed to Part 3 with partial 10% contribution in part 1)

**Introduction:**

We have implemented linear regression with quadratic regularization, which learns from a set of N training examples {xi,yi}Ni=1 an weight vector W that optimize the regularized Sum of Squared Error (SSE) objective as below

We have implemented gradient descent algorithm in order to optimize the above SSE.

We have initially read both the data files (train-p1.csv and tes-p1.csv). Then we normalize the data and extract the X-value and Y-value. We have initializes w0 as a zero vector of 1x45. So far, we have experimented with several learning rate and

value in order to receive the best possible outcome.

1. **Learning Rate:**

|  |  |  |
| --- | --- | --- |
| Learning rate (alpha) | SSE\_Test | SSE\_Train |
| 0.900 | 0.0223 | 0.0156 |
| 0.800 | 0.05574 | 0.049037 |
| 0.700 | 0.10193 | 0.095222 |
| 0.600 | 0.15407 | 0.14736 |
| 0.500 | 0.20637 | 0.19966 |
| 0.400 | 0.254 | 0.24729 |
| 0.300 | 0.2931 | 0.2864 |
| 0.200 | 0.32078 | 0.31408 |
| 0.100 | 0.3351 | 0.3284 |
|  |  |  |

From the above computation we found that as we decrease the value of learning rate, the value of SSE increases. We explored certain possible values between 0 to 1. Hence we decreased the learning rate to obtain a certain range of SSE values. For the complete case, is considered to be 1.

The best learning rate that we could find in our case is 0.900 as we find the least SSE value in this case.

1. **Experiment with different values :**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Learning rate | lambda | SSE Test | SSE Train | Cost Function |
| 0.9 | 0.001 | 0.0092476 | 0.0025437 | 0.0025307 |
| 0.9 | 0.01 | 0.009756 | 0.003052 | 0.0025307 |
| 0.9 | 0.1 | 0.020966 | 0.014262 | 0.0025307 |
| 0.9 | 1 | 0.21779 | 0.21109 | 0.0025307 |
| 0.9 | 0 | 0.0092346 | 0.0025307 | 0.0025307 |
| 0.9 | 10 | 3.2679 | 3.2612 | 0.0025307 |
| 0.9 | 100 | 46.934 | 46.928 | 0.0025307 |

1. As we increase the value, the SSE training value increases except for the point 0, where the SSE training value is the least.
2. It behaves same as above, with the increase in value, the SSE test value increases except for the point 0, where the value is observed to be the least.
3. As the value increases, the value of the|W| increases too. Thus, when we add the cost function with this corresponding value, the SSE tends to increase for both the case.
4. For = 1; Mean=4.698

|  |  |  |
| --- | --- | --- |
| w(1-10) | a(1-10) | 0.0439 |
| w(11-20) | a(11-20) | 0.2083 |
| w(21-30) | a(21-30) | 0.2883 |
| w(31-40) | a(31-40) | 0.0045 |
| w(41-50) | a(41-50) | 0.1466 |
| w(51-60) | a(51-60) | 18.5936 |
| w(61-70) | a(61-70) | 4.4265 |
| w(71-80) | a(71-80) | 18.0679 |
| w(81-90) | a(81-90) | 3.1626 |
| w(91-100) | a(91-100) | 2.0425 |
|  |  |  |

For = 0; Mean=4.69462

|  |  |  |
| --- | --- | --- |
| w(1-10) | a(1-10) | 0.0401 |
| w(11-20) | a(11-20) | 0.2053 |
| w(21-30) | a(21-30) | 0.2854 |
| w(31-40) | a(31-40) | 0.0017 |
| w(41-50) | a(41-50) | 0.1449 |
| w(51-60) | a(51-60) | 18.5907 |
| w(61-70) | a(61-70) | 4.4237 |
| w(71-80) | a(71-80) | 18.0649 |
| w(81-90) | a(81-90) | 3.1498 |
| w(91-100) | a(91-100) | 2.0397 |

For =0.00 1; Mean=4.69561

|  |  |  |
| --- | --- | --- |
| w(1-10) | a(1-10) | 0.04 |
| w(11-20) | a(11-20) | 0.2053 |
| w(21-30) | a(21-30) | 0.2854 |
| w(31-40) | a(31-40) | 0.0017 |
| w(41-50) | a(41-50) | 0.1449 |
| w(51-60) | a(51-60) | 18.05907 |
| w(61-70) | a(61-70) | 4.4237 |
| w(71-80) | a(71-80) | 18.0649 |
| w(81-90) | a(81-90) | 3.1598 |
| w(91-100) | a(91-100) | 2.0397 |

For = 0.01; Mean=4.69559

|  |  |  |
| --- | --- | --- |
| w(1-10) | a(1-10) | 0.04 |
| w(11-20) | a(11-20) | 0.2053 |
| w(21-30) | a(21-30) | 0.2854 |
| w(31-40) | a(31-40) | 0.0017 |
| w(41-50) | a(41-50) | 0.1449 |
| w(51-60) | a(51-60) | 18.05908 |
| w(61-70) | a(61-70) | 4.4237 |
| w(71-80) | a(71-80) | 18.0650 |
| w(81-90) | a(81-90) | 3.1598 |
| w(91-100) | a(91-100) | 2.0397 |

For = 0.1; Mean=4.69

|  |  |  |
| --- | --- | --- |
| w(1-10) | a(1-10) | 0.04 |
| w(11-20) | a(11-20) | 0.2056 |
| w(21-30) | a(21-30) | 0.2857 |
| w(31-40) | a(31-40) | 0.002 |
| w(41-50) | a(41-50) | 0.1450 |
| w(51-60) | a(51-60) | 18.05910 |
| w(61-70) | a(61-70) | 4.4239 |
| w(71-80) | a(71-80) | 18.0652 |
| w(81-90) | a(81-90) | 3.16 |
| w(91-100) | a(91-100) | 2.04 |

For = 10; Mean=4.724

|  |  |  |
| --- | --- | --- |
| w(1-10) | a(1-10) | 0.0784 |
| w(11-20) | a(11-20) | 0.2358 |
| w(21-30) | a(21-30) | 0.3142 |
| w(31-40) | a(31-40) | 0.0299 |
| w(41-50) | a(41-50) | 0.1622 |
| w(51-60) | a(51-60) | 18.6192 |
| w(61-70) | a(61-70) | 4.4523 |
| w(71-80) | a(71-80) | 18.0942 |
| w(81-90) | a(81-90) | 3.1879 |
| w(91-100) | a(91-100) | 2.0682 |

For = 100; Mean= 4.981

|  |  |  |
| --- | --- | --- |
| w(1-10) | a(1-10) | 0.4234 |
| w(11-20) | a(11-20) | 0.5100 |
| w(21-30) | a(21-30) | 0.5737 |
| w(31-40) | a(31-40) | 0.2835 |
| w(41-50) | a(41-50) | 0.3180 |
| w(51-60) | a(51-60) | 18.8732 |
| w(61-70) | a(61-70) | 4.7104 |
| w(71-80) | a(71-80) | 18.3581 |
| w(81-90) | a(81-90) | 3.4414 |
| w(91-100) | a(91-100) | 2.3244 |

1. = 0 is the best as it has the lowest mean among all the values.
2. Cross validated SSE is smaller than SSE train.
3. increases, the value increases .