|  |  |  |
| --- | --- | --- |
| **Tolerance** | **Alpha value** | **Iteration to converge** |
| 1e-3 | 0.00002 | 964 |
| 1e-3 | 0.0004 | 668 |
| 1e-3 | 0.004 | 221 |
| 1e-3 | 0.04 | 1 |
| 1e-5 | 0.003 | 766 |
| 1e-5 | 0.0007 | 2582 |
| 1e-5 | 0.7 | 1 |
| 1e-5 | 0.08 | 1 |
| 1e-1 | 0.00000008 | 1 |
| 1e-8 | 0.00000008 | 10000(still too much outliers |

**Question 2 a)**

Following different values were tried for evaluating the gradient descent.

**Question 2 b)**

Please see next page for the code of Newton’s Method, as mentioned only the function code is provided.

|  |  |  |  |
| --- | --- | --- | --- |
| **Tolerance** | **Alpha value** | **Convergence(gradient)** | **Convergence (Newton)** |
| 1e-3 | Not included in newton method | 221(with max learning rate from above table) | 588 |
| 1e-5 | Not included in newton method | 766 | 2437 |
| 1e-1 | Not included in newton method | 1 | 52 |
| 1e-8 | Not included in newton method | 1 | 6137 |
|  |  |  |  |

**Question 2 c)**

See next page for the stochastic descent gradient code,

Convergence values are given below ,most noticeable thing about sgd was the time execution that it was much faster than traditional gradient descent see next table(2d).

|  |  |  |  |
| --- | --- | --- | --- |
| **Tolerance** | **Alpha value** | **Convergence(gradient)** | **Convergence (SGD)** |
| 1e-3 | 0.0004 | 668 | 105 |
| 1e-3 | 0.004 | 221 | 24 |
| 1e-3 | 0.04 | 1 | 16 |
| 1e-5 | 0.003 | 766 | 719 |
| 1e-5 | 0.0007 | 2582 | 719 |
| 1e-5 | 0.7 | 1 | 2 |
| 1e-5 | 0.08 | 1 | 16 |
| 1e-1 | 0.00000008 | 1 | 2 |
| 1e-8 | 0.00000008 | 10000(still too much outliers | 10000 |

**Question 2 d)**

**Time comparison for all the method for different number of samples**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Method** | **alpha** | **time** | **iterations** | **tolerance** | **N\_samples** |
| **Gradient** | 0. 0000008 | 26.28 | 7135 | **1e-8** | **10000** |
| **Gradient** | 0.0000009 | 115.60 | 3379 | **1e-8** | **100000** |
| **Gradient** | 0.0000009 | 41.62 | 1321 | **1e-3** | **100000** |
| **Newton** | Not applicable | 52.29 | 1585 | **1e-8** | **100000** |
| **Newton** | Not applicable | 8.51 | 1466 | **1e-8** | **10000** |
| **Newton** | Not applicable | 3.82 | 573 | **1e-3** | **10000** |
| **SGD** | 0. 0000008 | 9.01 | 10000 | **1e-8** | **10000** |
| **SGD** | **0.** 0000009 | 0.80 | 3 | **1e-3** | 100000 |
| **SGD** | **0.** 0000009 | 8.9 | 10000 | **1e-8** | 100000 |