3) e) optimal weights = wopt = [w0, w1, w2] = [ 0.07398882 -0.15041019 0.28511816]

f) Randomly picked weights for PTA: [w0’, w1’, w2’] = [ 0.07611122 0.26930747 -0.57670416]

j) vii) Final weights for learning rate 1 = [ 11.07611122 -22.19536636 42.95259398]. These weights are a lot different compared to the optimal weights above.

n) Based on my results, I found no relationship between learning rate and no of epochs needed for PTA to converge. It makes sense as the number of epochs mainly depend on the observations i.e, data

|  |  |  |
| --- | --- | --- |
| Number of samples | Learning Rate | Number of epochs for convergence |
| 100 | 0.1 | 12 |
| 100 | 1 | 11 |
| 100 | 10 | 12 |
| 1000 | 0.1 | 6 |
| 1000 | 1 | 11 |
| 1000 | 10 | 11 |

o) Yes. We would get same results i.e. there wouldn’t be any correlation between the learning rate and number of epochs needed for convergence. (I tried running with different weights)

p) I got higher weights in case of n = 1000. It makes sense as there are 1000 samples. It seems the ratio of the weights has correlation with learning rate for both n = 100 and n = 1000. But the number of epochs for convergence has no correlation with the number of samples.

|  |  |  |  |
| --- | --- | --- | --- |
| Number of samples | Learning Rate | Final Weights | Number of epochs for convergence |
| 100 | 0.1 | [1.17611122 -2.33942677 4.49016557] | 12 |
| 100 | 1 | [11.07611122 -22.19536636 42.95259398] | 11 |
| 100 | 10 | [110.07611122 -221.60765386 432.90900933] | 12 |
| 1000 | 0.1 | [10.47611122 -21.29689005 40.43167596] | 6 |
| 1000 | 1 | [105.07611122 -215.03790194 409.4783406] | 11 |
| 1000 | 10 | [1060.07611122 -2160.25697083 4093.18590864] | 11 |