## **Why should we normalize inputs**

Why does normalizing inputs help?

1. Let’s compare the before-after of normalization of inputs

|  |  |
| --- | --- |
| **Before Normalization** | **After Normalization** |
|  |  |
| Here, we observe that x1 has a much larger range of values than x2 | Here, x1 and x2 have both been brought into a comparable range of values. |
| When optimising the Loss function, one set of weights corresponding to the smaller feature x2 will end up being very large.  This leads to oscillations about the local minima during gradient descent as much larger updates are made due to large weights. | Here, since weights end up being smaller, we see a good reduction in the oscillation about the local minima during optimization. |
| Another point to note is that in Gradient Descent, our **updates would be very biased** to the feature corresponding to the larger weights.  Thus we may see the model becoming very sensitive to that particular feature | Here, the weights are brought in a comparable range.  This reduces the chance of updates biased to any particular feature |