



In all Below learning rate is fixed

* GD – All records at a time
* SGD – One record at a time
* Mini Batch SGD – Records will pass in batches (Eg. 1000 at a time)
* SGD with Momentum – This is to reduce zigzag noise.

In all below learning rate is dynamic – large in starting epochs then reduce gradually

* SGD will have more noise than Mini Batch SGD because in SGD we pass one record at a time and in Mini Batch SGD we pass records in batches. Please check 2nd snapshot
* Computationally, which one is more expensive (GD > SGD > Mini Batch SGD)
* Bit in terms of conversion, which one takes more time (SGD > Mini Batch SGD > GD), because SGD takes longer path. Check in 2nd snapshot