Difference b/w batch Gradient Descent and Stochastic Gradient Descent:

Batch Gradient Descent (BGD)– slower and heavy algorithm as it loads up all the data into the memory and waits for all the rows to complete its calculation. It is deterministic algorithm unlike stochastic which is random in nature.

In BGD if we start the NN with the same set of weights then the number of iterations when starting the BGD every time will be the same and the same results every time we run this algorithm but with stochastic GD even if we start the training with the same weights the next row might get a different weights to work with.

Combo of both these is called mini batch gradient descent method

Pseudo inverse

In Mathematics, inverse is only found for square matrices but not all sq matrixes have an inverse. Therefore, we find a pseudo inverse and work with it.

In machine learning, not all real data have an inverse, necessarily the dataset vectors do not form a square matrix. There are many types of pseudo inverse.

Using Single Value Decomposition (SVD) we can find a pseudoinverse.