Q1. Stochastic gradient descent is a method may in some situations be an effective method for optimizing machine learning models. However, it does have several disadvantages that may compromise the efficiency or reliability of the convergence. The variability of the input data may cause the gradient determined by the gradient descent algorithm to change significantly at every update step. This causes noisy convergence which may extend the time it takes to reach the optimal solution. By considering the previous gradients when determining the convergence can be made more smooth. Due to the similarity in with objects with mass in motion this method is called momentum.