Deep Learning Assignment 5

# In your own words, please describe the difference between Gradient Descent with Momentum and regular Gradient Descent, and how Gradient Descent with Momentum speeds up the NN?

In normal gradient descent what happens is that the gradient of the objective function is used in order to find the minima. This is usually a long process and it takes a lot of steps for the minima to be reached. It has a problem though because it can bounce around in the search space because of the structure and shape of the search space. Sometimes it can also get stuck in some of the flat spots if the gradient is low. On the other hand, if a momentum term is added, the direction of search of the minima term will gain some inertia in a direction. This will enable it to easily cross low gradient spots in the search space and enable the neural network to converge faster.