1. Assume that you want to balance a (real) ball on a (real) saddle.
   * Why is this hard?

A saddle point is a point on the surface of the graph of a function where the slopes in orthogonal directions are all zero (a critical point), but which is not a local extremum of the function. When we try to make it stand still at this point, a little disturbance will get it out of the saddle point.

* + Can you exploit this effect also for optimization algorithms?

When we want to test whether an optimal solution is a local optimal extremum or just a saddle point, we can exert a small random disturbance on the optimal solution and do the optimization again. If the new-found optimal solution is almost the same as the original one, that can be seen as the local optimal.

1. What changes when we perform SGD with momentum? What happens when we use minibatch SGD with momentum?

Adding momentum make convergence faster.

Momentum prevents stalling of the optimization process that is much more likely to occur for stochastic gradient descent.

The effects of momentum on SGD are almost same as on the minibatch SGD.