# Neural Network

**Chung Tsz Ting 1155110208 CSCI3230 Assignment 1**



1. We need the learning rate to adjust the step size for doing a gradient descent search through weight space. This parameter is usually significantly smaller than 1, this is to prevent searching overshoot. For example, the local or global minima is only 0.5 far away from the current location in the weight space in one axis indicating a specific dimension, without , g’(x) may have a value larger than 0.5 and causing the update in x overshoot the targeted local or global minima.

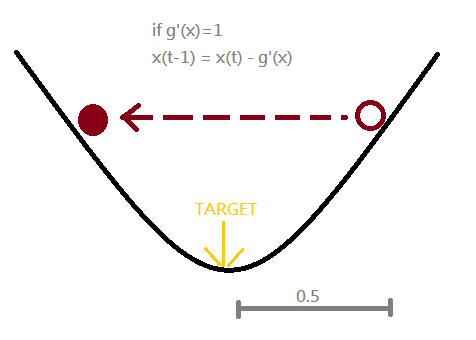


Fig.1: Visualization in 2-dimensional weight space (t represents the time step)

1. =

1. Starting from the error term , using multivariate chain rule as shown in above sub-question (, and ), the partial derivative of (calculated in () by using () and ()) can be calculated. This gives the information of the direction that the gradient should descent to the minima. The hyperparameters of the network which is the parameters can be updated recursively in the below format,

The whole process will stop when the difference between and is small than .