



So, the update rule is:



Let

Let be the corresponding linear hypothesis which have the lowest square error on point x and y, which is the line constructed by x and y (the square error is zero).

So,



SGD:

(use point to approximate the true gradient)

So the update rule is:

The update rule in Q3 looks like this:

We can observe that

So, the two update rules are probably approximately same, meaning that when we add virtual example in the training data set and do normal SGD without regularization, it can reach the same result using GD with regularization.

For target function , the squared error for is:

To solve , we need to solve the equation . After calculation, we get:

So, for each x, the level of deterministic noise is