Backward Interpolation

Once we have predicted the output we compute the loss, L. We use the mean squared error as the loss function, that is the mean of the squared difference between the actual output, **yi** ,and the predicted output , which is given as follows:

Now we use back propagation to minimise the loss **L**. To minimise the loss we need to find the optimal values for our filter **w.** Our filter matrix consist of 4 values w1 , w2 , w3 , w4. To find the optimal filter matrix, we need to calculate the gradients of our loss function.

*Equations of output matrix*

**O1 = x1w1 + x2w2 + x4w3 + x5w4**

**O2 = x2w1 + x3w2 + x5w3 + x6w4**

**O3 = x4w1 + x5w2 + x7w3 + x8w4**

**O4 = x5w1 + x6w2 + x8w3 + x9w4**

To calculate the gradients we partially differentiate **L** wrt **w1 , w2 , w3 , w4**

Computing the derivatives of loss wrt filter matrix is just another convolution operation. The preceding equations look like the result of a convolution operation b/w the input matrix and the gradient of the loss wrt the output as a filter matrix.