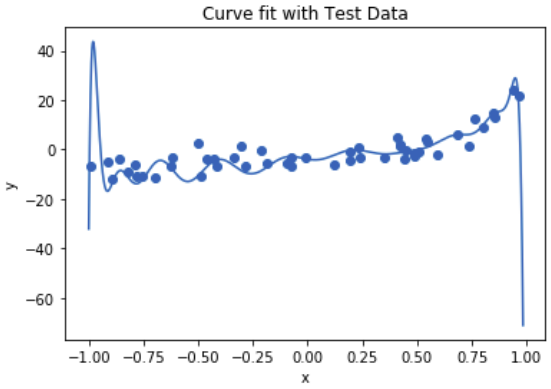
1 Model Selection

1. The training MSE is : 6.47474592081
   The validation MSE is : 1418.50796451



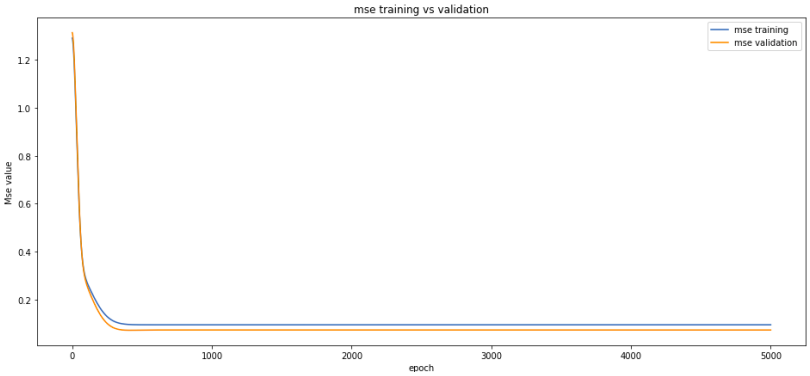
As we can see, the fit seems to be a good one except around the edges of the interval where it seems a bit off (fluctuations)

2. The fit is actually better that the previous one, there are no oscillations at the edges of the interval after adding L2 regularization.

3. Just by looking at the previous fit, we can deduce that the degree of the original polynomial is definitely less than 20, in fact judging by the real data and the previous fits, I'd say the original degree is sround 5 or 4.

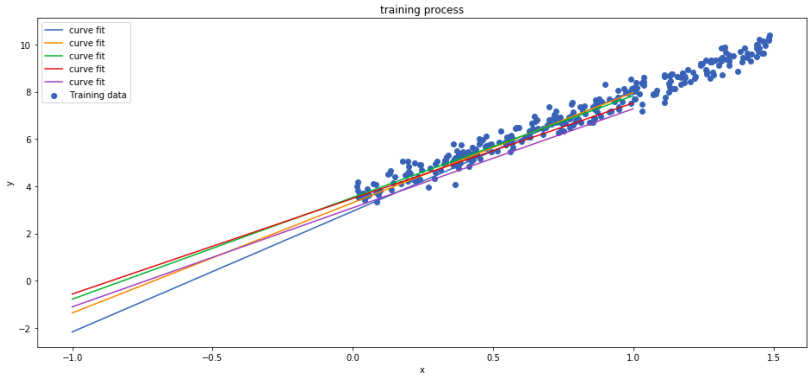
2 Gradient Descent for Regression

1. Here is the learning plot.



2. The chosen model is at step size equal to 0.001, We get a test MSE equal to 0.208061016794.

3. Here is the evolution of the learning process in 5 different plots.



3 Real life dataset

1. this is obviously not the best way, a better approach as seen in class, is to use the K-nearest neighbor or maybe even regression.

2. The MSE found are:

```
The MSE over the first test data is:  0.0190055064924
The MSE over the second test data is:  0.0176292678266
The MSE over the third test data is:  0.0211232731655
The MSE over the fourth test data is:  0.0204193773437
The MSE over the fifth test data is:  0.0161002723889
The average MSE is:  0.0188555394434
```

3. It is possible to do feature selection with this experiment. If for example the regularization gives very small number for one of the parameters (i.e $|\omega_i| < 0.001$). Then we could exclude this parameters as the regularization minimizes the effect of these features in the process."