T-61.3050 Machine Learning: Basic Principles Solutions of Homework problems: Exercise Session 2

Md Mohsin Ali Khan(Student No: 336790)

September 22, 2013

1 Derive the general least squares solution:

Assume that we have a set of training data of N observations $D = \{(x^t, r^t)\}_{t=1}^N$ where $x^t, y^t \in \mathbb{R}$. We want to find out a function $f : \mathbb{R} \to \mathbb{R}$ so that,

$$E = \frac{1}{N} \sum_{t=1}^{N} (r^t - f(x^t))^2$$

is minimized. Let us consider basis functions $\phi_i : \mathbb{R} \to \mathbb{R}$. And let us also consider a $1 \times k$ verctor $W = (w_0\, w_1\, w_2 \,..\, w_k)$. Let us assume:

$$f(x^t) = \sum_{i=1}^{k} w_i \cdot \phi_i(x^t)$$

That means we can write:

$$E = \frac{1}{N} \sum_{t=1}^{N} \left(r^t - \sum_{i=1}^{k} w_i \cdot \phi_i(x^t) \right)^2$$

In the equation above, only the w_i's are unknown. In order to find out the optimum $f(x)$, we need to find out the w_i's so that E is minimized. That means we need to find out the w_i's so that the gradient of E becomes zero. That means the partial derivatives becomes zero. So, we can write:

$$\frac{\delta E}{\delta w_j} = \frac{1}{N} \sum_{t=1}^{N} 2 \left((r^t - \sum_{i=1}^{k} w_i \cdot \phi_i(x^t) \right) \phi_j(x^t)$$

Which is after simplification:

$$\sum_{t=1}^{N} r^t \phi_j(x^t) \quad = \quad \sum_{t=1}^{N} \sum_{i=1}^{k} w_i \cdot \phi_i(x^t) \phi_j(x^t) \tag{1}$$

The equations (1) over all the possible values of j from 0 to k can be represented in matrix notation as below:

$$W\Phi\Phi^T = R\Phi^T \tag{2}$$

Where Φ is a $k \times N$ matrix containing all the $\phi_i(x^t) = \Phi_{i,t}$ and R is a $1 \times N$ vector containing all the r^t. Now if we multiply both side of (2) by $(\phi\phi^T)^{-1}$ then we can write:

$$W = R\Phi^T(\Phi\Phi^T)^{-1} \tag{3}$$

2 Polynomial Regression of different Orders:

Set the basis function of the above derviation to $\phi_i(x) = (x)^i$. And for different values of $k = 0$ to 6, solve the vecotr W as it has been dervied in problem 1 of this tutorial.

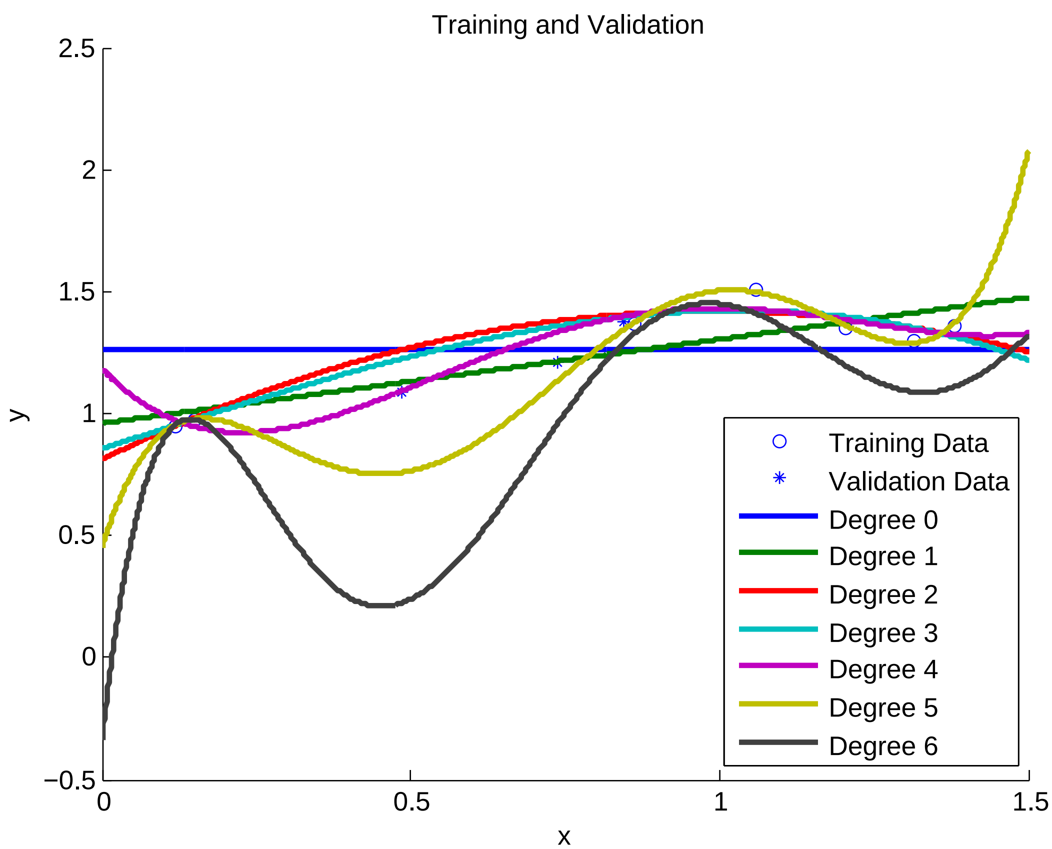
Here Below is the procedure:

1. The first 35 rows of the training data set have been considered for training and the rest 15 has been considered for validation.
2. At first W has been solved for different orders, (thats is, for different values of $k = 0..6$) using the 35 training data. That means the polynomials are extracted.
3. Then using those different polynomials, validation errors have been calculated for the 15 data left out for validation.
4. It has been found that validation error is minimum when $k = 4$ and the polynomial is $1.1865 - 2.6468x + 8.0724x^2 - 7.2140x^3 + 2.034x^4$ (See the below table)
5. Then using all these polynomials, error has been calculated over all the 50 training data. It has been found that the polynomial which generated the lowest error on validation data has generated the lowest error in this case also. So, this polynomial($1.1865 - 2.6468x + 8.0724x^2 - 7.2140x^3 + 2.034x^4$) has been chosen as the solution.
6. Then using all the polynomials, errors have been calculated on the test data and found that the chosen polynomial is the one which generates the least error on the test data as well

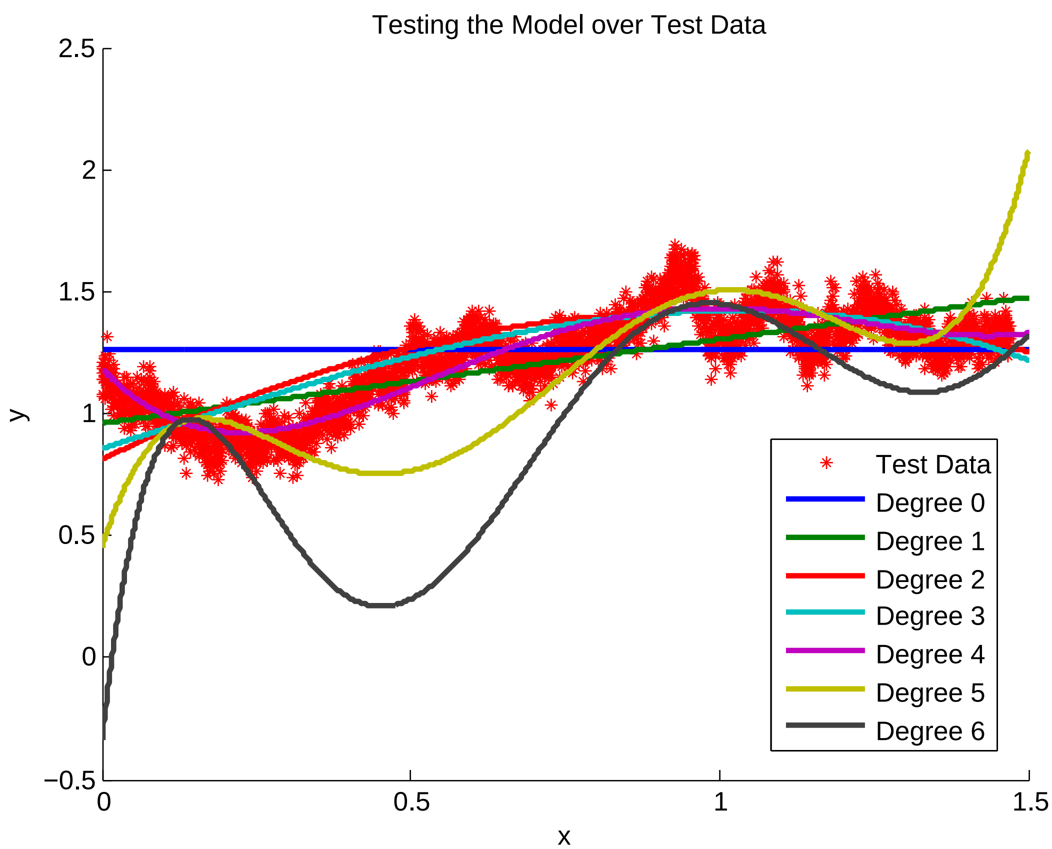
In below table E_T, E_V, E_{T+V} and E_{Test} represents the **Average Squared Training Error, Validation Error, Training+Validation Error and Test Error** respectively

Table 1: Polynomial Regression Result

k	E_T	E_V	E_{T+V}	E_{Test}	$f(x\|w_0, \ldots, w_k) = \sum_{i=0}^{k} w_i x_i$
0	0.0383	0.0152	0.0314	0.0391	1.2591
1	0.0096	0.0057	0.0084	0.0160	$0.9595 + 0.3444x$
2	0.0022	0.0201	0.0076	0.0177	$0.8094 + 1.2336x - 0.6244x^2$
3	0.0021	0.0138	0.0056	0.0142	$0.8552 + 0.8106x + 0.0232x^2 - 0.2675x^3$
4	0.0019	**0.0052**	**0.0029**	**0.0076**	$1.1865 - 2.6468x + 8.0724x^2 - 7.2140x^3 + 2.034x^4$
5	$5.4786e-05$	0.0385	0.0116	0.0519	$0.4516 + 7.9751x - 39.9080x^2 + 75.8405x^3 - 58.8348x^4 + 15.9794x^5$
6	0.0205	0.2724	0.0961	0.2210	$-0.3356 + 22.7281x - 131.0983x^2 + 295.4919x^3 - 306.8257x^4$ $+148.8833x^5 - 27.3936x^6$



Training and Validation
2.5
2
1.5
1
0.5
0
−0.5
y
0
0.5
1
1.5
x
Training Data
Validation Data
Degree 0
Degree 1
Degree 2
Degree 3
Degree 4
Degree 5
Degree 6


Testing the Model over Test Data
2.5
2
1.5
1
0.5
0
−0.5
y
0
0.5
1
1.5
x
Test Data
Degree 0
Degree 1
Degree 2
Degree 3
Degree 4
Degree 5
Degree 6