HW1

Zeyuan Xu

October 9, 2017

1 Task 1

define $\mathbf{Z} = \{\mathbf{z_1}, \mathbf{z_2}, ..., \mathbf{z_n}\}$. Then the function to minimize is:

$$\begin{aligned} J(\mathbf{w}) &= \frac{1}{2}\sum_{n=1}^{N}(y(x_n, \mathbf{w}) - t_n)^2 \\ &= \frac{1}{2}\sum_{n=1}^{N}(\mathbf{z}_n^\top \mathbf{w} - t_n)^2 \\ &= \frac{1}{2}||\mathbf{Z}\mathbf{w} - \mathbf{t}||^2 \end{aligned}$$

Using chain rule to differentiate the cost function against $\mathbf{w}$ with the variable declaration:

$$\begin{aligned} \mathbf{u} &= \mathbf{Z}\mathbf{w} - \mathbf{t} \\ \mathbf{J} &= \frac{1}{2}||\mathbf{u}||^2 \end{aligned}$$

Then we have that:

$$\begin{aligned} &D_{\mathbf{u}}(\mathbf{J}) = D_{\mathbf{u}}(\frac{1}{2}||\mathbf{u}||^2) = \mathbf{u}^\top \\ &D_{\mathbf{w}}(\mathbf{u}) = D_{\mathbf{w}}(\mathbf{Z}\mathbf{w} - \mathbf{t}) = \mathbf{Z} \\ &D_{\mathbf{w}}(\mathbf{J}(\mathbf{w})) = D_{\mathbf{u}}(\mathbf{J})D_{\mathbf{w}}(\mathbf{u}) = \mathbf{u}^\top\mathbf{Z} = (\mathbf{Z}\mathbf{w} - \mathbf{t})^\top\mathbf{Z} \end{aligned}$$

Set gradient to zero to find the point of minimal cost function:

$$\begin{aligned} &D_{\mathbf{w}}\mathbf{J}(\mathbf{w}) = 0 \\ &\implies 0 = (\mathbf{Z}\mathbf{w} - \mathbf{t})^\top\mathbf{Z} \\ &\implies 0 = (\mathbf{Z}\mathbf{w})^\top\mathbf{Z} - \mathbf{t}^\top\mathbf{Z} \\ &\implies 0 = \mathbf{w}^\top\mathbf{Z}^\top\mathbf{Z} - \mathbf{t}^\top\mathbf{Z} \\ &\implies \mathbf{t}^\top\mathbf{Z} = \mathbf{w}^\top\mathbf{Z}^\top\mathbf{Z} \\ &\implies \mathbf{Z}^\top\mathbf{t} = \mathbf{Z}^\top\mathbf{Z}\mathbf{w} \end{aligned}$$

so we know that we can write $\nabla J(\mathbf{w}) = \mathbf{0}$ in $\mathbf{A}\mathbf{w} = \mathbf{b}$, where $\mathbf{A} = \mathbf{Z}^\top\mathbf{Z}$ and $\mathbf{b} = \mathbf{Z}^\top\mathbf{t}$.

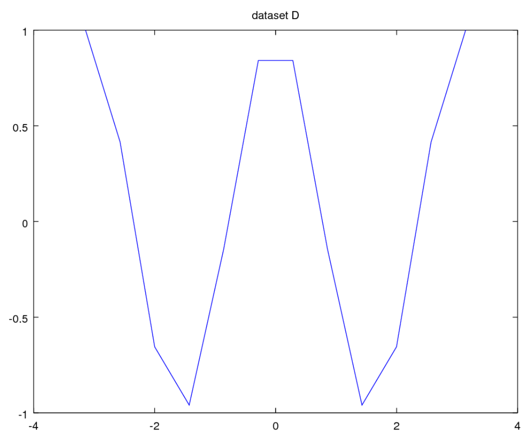

Figure 1: dataset-D

2 Task 2

the solution for the equation in Task 1 gives the solution:

$$\mathbf{w} = (\mathbf{Z}^{\top}\mathbf{Z})^{-1}\mathbf{Z}^{\top}\mathbf{t}$$

the solution always exist (i.e. there are always more than one solutions), since $(\mathbf{Z}^{\top}\mathbf{Z})^{-1}\mathbf{Z}^{\top}$ is the Moore-Penrose pseudoinverse and exists for all matrices (page 142, PRML). The solution is unique only when $\mathbf{Z}$ is of full rank, or equivalently $\mathbf{Z}^{\top}\mathbf{Z}$ is invertible. Since Z is a n x (m+1) matrix, we have$\mathbf{Z}^{\top}\mathbf{Z}$ with dimension (M+1)x(M+1). we have rank($\mathbf{Z}^{\top}\mathbf{Z}$) = rank($\mathbf{Z}$). So that if $M < N$, we always have unique solution. If $M \geq N$ we have more than one solutions.

3 Task 4, 5

Task 3 is missing in the lab assignment.
the plot result is shown in Figure 1. The matlab code for task 5 is shown below:

```
function w = hw1_task5(X,t,n)
 % Z matrix is of dimension: N x (M+1)
 Z = ones(12,n+1);
  for i = 2:(n+1)   %modify column vectors
    Z(:,i)= X.^(i-1);
    end
 % after computing the Z matrix apply the formula
 % a should be of dimension: (M+1)x(M+1)
  a = inv(transpose(Z)*Z);
  Z
 % w should be of dimension (M+1) x 1: a column vector
  w = a * transpose(Z) * transpose(t);
 end
```

4 Task 6

the plots are shown with their corresponding value of dimension N: From the plots, it is clear

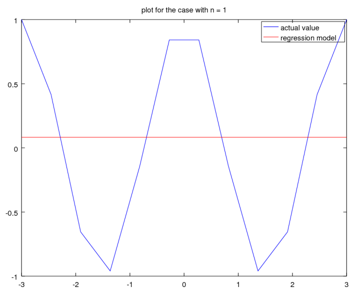

Figure 2: M=1

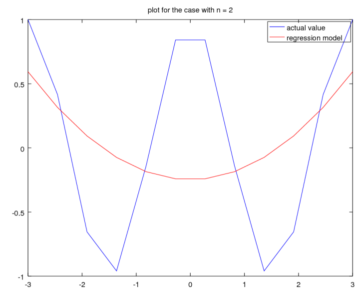

Figure 3: M=2

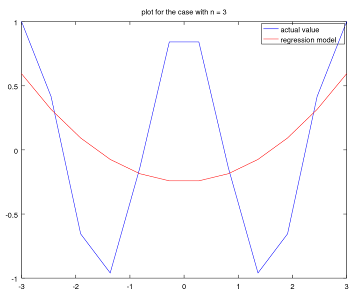

Figure 4: M=3

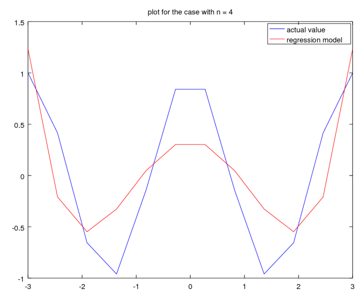

Figure 5: M=4

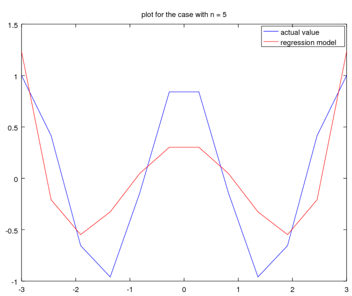

Figure 6: M=5

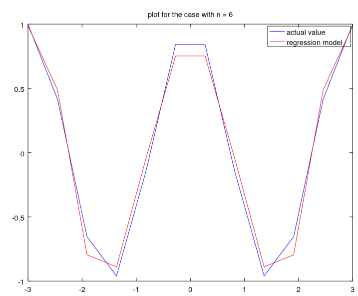

Figure 7: M=6

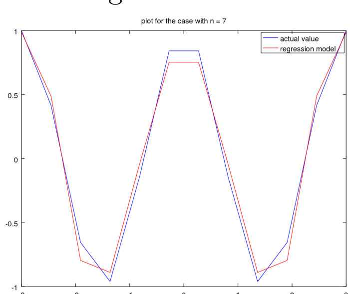

Figure 8: M=7

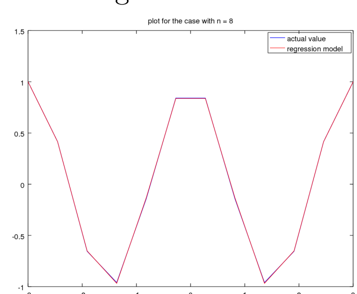

Figure 9: M=8

that approximation is more and more precise as higher dimension terms are introduced, from M

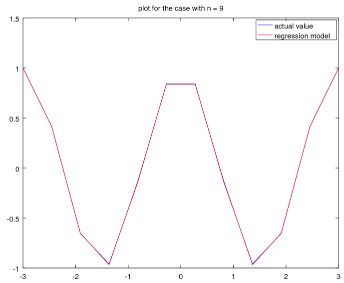

Figure 10: M=9

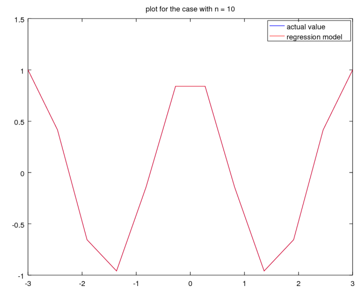

Figure 11: M=10

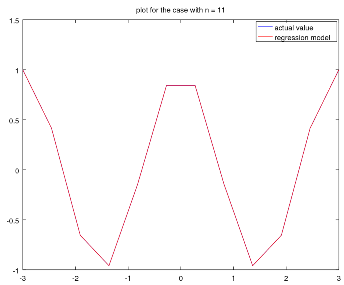

Figure 12: M=11

= 9 and on, the plot is almost exactly the same as f.The matlab code is shown below:

```
domain = linspace(-3,3,12);
x = linspace(-pi, pi, 12);
t = cos(2.*x);
%case
for n = 1:11
  w = hw1_task5(x,t,n);
  Z = ones(12,n+1);
  for i = 2: (n+1)    %modify column vectors
    Z(:,i)= x.^(i-1);
  end
  % dimension check:
  % w' of 1 by (M+1)
  % Z' of (M+1) by N
  y = (transpose(w)* transpose(Z) )
  %plot
  fig = figure;
  plot(domain, t, 'color', 'b');
  hold on;
  plot(domain, y, 'color', 'r');
  title(sprintf('plot for the case with n = %d',n));
  legend('actual value', 'regression model');
  saveas(fig, sprintf("hw1_task6_fig%d.png",n))
end
```

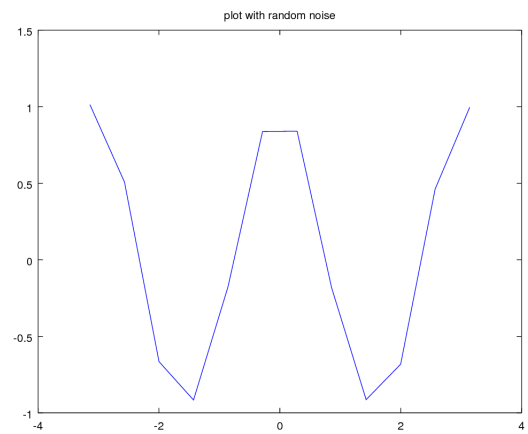

Figure 13: random noise plot cos(x)

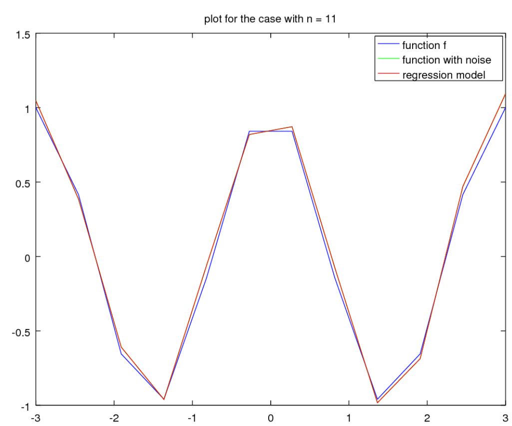

Figure 14: M=11

5 Task 7

the random noise here is uniform random noise, the plot is shown in figure 13.

6 Task 8

The comparison shows that with value of M approach N, the result is generally more and more similar to f. However, it takes larger value of M when random noise is introduced in the model.And in our case, when $M = 10$, we have a better approximation than $M = 11$ for function f. The result for $M = 11$ is shown in figure 14.