5.6 (a) $Q_0 = 1$. $Q1 = M_1 + 1$. $Q_2 = M_1(M_2+1)+1$ $Q_3 = M_1(M_2(M_3+1)+1)+1$ ...

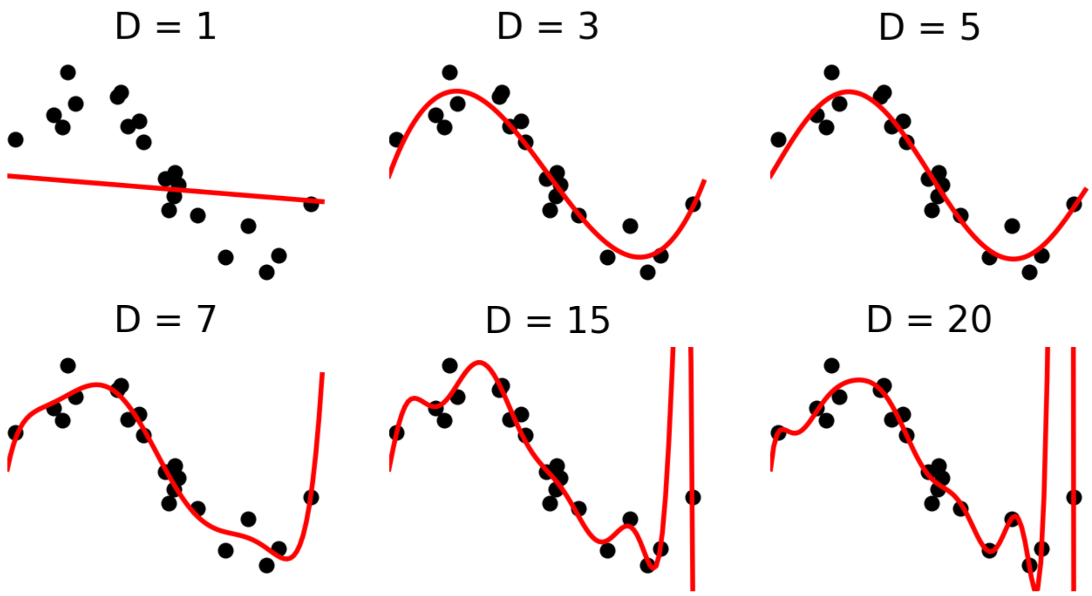
$Q_L = M_{L+1} M_L \cdots M_0 + M_L M_{L-1} \cdots M_0 + \cdots + M_1 M_0 + M_0 = \sum_{i=0}^{L+1} (\prod_{j=0}^{i} M_j)$

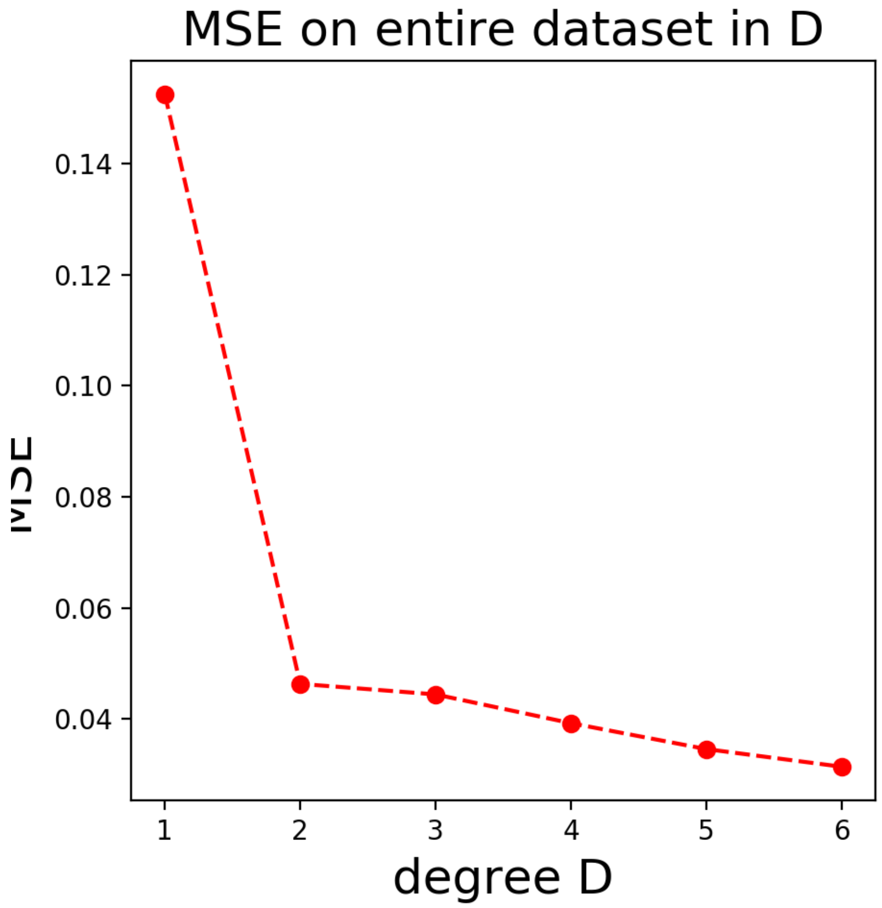
(b). N is getting larger, the number of parameter Q is getting more

the number of parameter doesn't change with the number of data points P.

5.7 (b) As D increases, it represents the phenomenon generating the data more and more accurately (from first figure, we could see the regression line is getting closer and closer to the data). At the same time, D↑, MSE is getting smaller and smaller, So. the more accurracy to represent data (figure 1), the smaller MSE is.

由 扫描全能王 扫描创建


D = 1
D = 3
D = 5
D = 7
D = 15
D = 20



MSE on entire dataset in D
MSE
0.14
0.12
0.10
0.08
0.06
0.04
1
2
3
4
5
6
degree D

5.9.

(a) (5.21):

$$g = \sum_{p=1}^{P}\left(b+\sum_{m=1}^{M} a(c_m+\bar{x}_p^T\bar{v}_m)w_m - y_p\right)^2$$

$$\frac{\partial g}{\partial b} = 2\sum_{p=1}^{P}\left(b+\sum_{m=1}^{M} a(c_m+\bar{x}_p^T\bar{v}_m)w_m - y_p\right)$$

$$\frac{\partial g}{\partial w_n} = 2\sum_{p=1}^{P}\left(b+\sum_{m=1}^{M} a(c_m+\bar{x}_p^T\bar{v}_m)w_m - y_p\right)\cdot\left(\frac{\partial a(c_n+\bar{x}_p^T\bar{v}_n)w_n}{\partial w_n}\right)$$

$$= 2\sum_{p=1}^{P}\left(b+\sum_{m=1}^{M} a(c_m+\bar{x}_p^T\bar{v}_m)w_m - y_p\right)\cdot a(c_n+\bar{x}_p^T\bar{v}_n)$$

$$\frac{\partial g}{\partial c_n} = 2\sum_{p=1}^{P}\left(b+\sum_{m=1}^{M} a(c_m+\bar{x}_p^T\bar{v}_m)w_m - y_p\right)\frac{\partial a(c_n+\bar{x}_p^T\bar{v}_n)w_n}{\partial c_n}$$

$$= 2\sum_{p=1}^{P}\left(b+\sum_{m=1}^{M} a(c_m+\bar{x}_p^T\bar{v}_m)w_m - y_p\right)a'(c_n+\bar{x}_p^T\bar{v}_n)w_n.$$

$$\nabla_{v_n} g = 2\sum_{p=1}^{P}\left(b+\sum_{m=1}^{M} a(c_m+\bar{x}_p^T\bar{v}_m)w_m - y_p\right)a'(c_n+\bar{x}_p^T\bar{v}_n)w_m\,\frac{\partial\,\bar{v}_n^T\bar{x}_p}{\partial v_n}$$

$$= 2\sum_{p=1}^{P}\left(b+\sum_{m=1}^{M} a(c_m+\bar{x}_p^T\bar{v}_m)w_m - y_p\right)a'(c_n+\bar{x}_p^T\bar{v}_n)w_n\cdot\bar{x}_p.$$

(b) according to part (a),

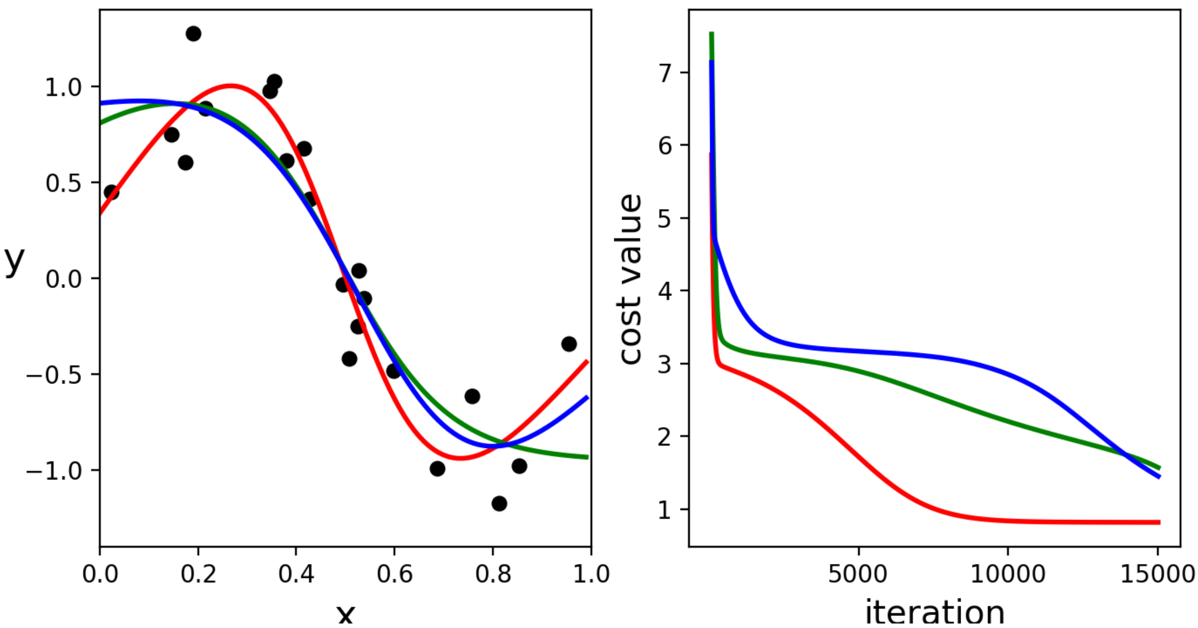
$$\frac{\partial g}{\partial b} = 2\cdot\bar{1}_{p\times1}^T\begin{bmatrix} b+\sum_{m=1}^{M} a(c_m+\bar{x}_1^T\bar{v}_m)w_m-y_1 \\ b+\sum_{m=1}^{M} a(c_m+\bar{x}_2^T\bar{v}_m)w_m-y_2 \\ \vdots \\ b+\sum_{m=1}^{M} a(c_m+\bar{x}_p^T\bar{v}_m)w_m-y_p \end{bmatrix} = 2\cdot\bar{1}_{p\times1}^T\cdot\bar{q}$$

the first half portion is the same as above.

$$\frac{\partial g}{\partial w_n} = 2\cdot\bar{1}_{p\times1}^T\cdot\bar{q}\odot\begin{bmatrix} a(c_n+\bar{x}_1^T\bar{v}_n) \\ \vdots \\ a(c_n+\bar{x}_p^T\bar{v}_n)\end{bmatrix}$$

$$= 2\cdot\bar{1}_{p\times1}^T\cdot\bar{q}\odot t_n.$$

由 扫描全能王 扫描创建



1.0
0.5
0.0
−0.5
−1.0
y
0.0
0.2
0.4
0.6
0.8
1.0
x
7
6
5
4
3
2
1
cost value
5000
10000
15000
iteration

$$\frac{\partial g}{\partial C_n} = 2 \cdot \bar{1}_{p\times1}^T \cdot \bar{q} \odot \begin{bmatrix} a'(C_n + \bar{x}_1^T \bar{V}_n) \\ \vdots \\ a'(C_n + \bar{x}_p^T \bar{V}_n) \end{bmatrix} \cdot W_n$$

$$= 2 \cdot \bar{1}_{p\times1}^T (\bar{q} \odot S_n) W_n .$$

the method is the same as ~~below~~. above

$$\frac{\partial g}{\partial V_n} = 2 \cdot \bar{1}_{p\times1}^T (\bar{q} \odot X \odot S_n) W_n .$$

c) According to figure, we could see the more ~~exact~~ closed to approximate the dataset, the less cost value is. (like red line).

We adopt more iterations, the cost value gets much less.

由 扫描全能王 扫描创建