PRML Chapter 1

1　例：多項式曲線フィッティング

実数値の入力変数 x を観測し、それを用いて実数値の目標変数 t を予測する回帰問題を考える。ただし、ここでは関数 $\sin(2\pi x)$ にガウス分布に従うランダムノイズを加えて生成した人工データを用いる。

訓練集合として、N 個の観測地 x を並べた $\boldsymbol{x} \equiv (x_1, \cdots, x_N)^T$ と、それぞれに対応する観測値 t を並べた $\boldsymbol{t} \equiv (t_1, \cdots, t_N)^T$ が与えられたとする。図 1.1 は、$N = 10$ の場合の人工データの例である。

我々の目標は、この訓練集合を利用して、新たな入力変数 $\hat{x}$ に対して目標変数 $\hat{t}$ の値を予測することである。

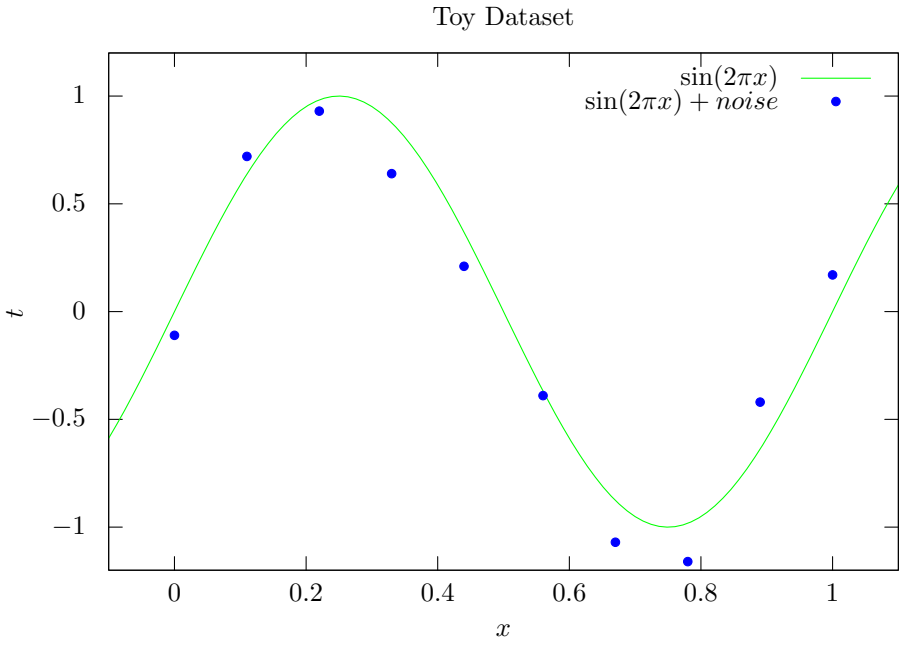

図 1.1　$N = 10$ 個の訓練データの例

ここでは、以下のような多項式を用いてデータへのフィッティングを行うことにする。

$$y(x, \boldsymbol{w}) = w_0 + w_1 x + w_2 x^2 + \cdots + w_M x^M = \sum_{j=0}^{M} w_j x^j \tag{1.1}$$

ただし、M は多項式の**次数 (order)** で、x^j は x の j 乗を表す。多項式の係数 $w_0, \cdots, w_M$ をまとめて $\boldsymbol{w}$ と書くことにする。多項式 $y(x, \boldsymbol{w})$ は、x の非線形関数であるが、係数 $\boldsymbol{w}$ の線形関数であることに注意する。このような、未知パラメータに対して線形な関数は**線形モデル**と呼ばれる。

訓練データに多項式をあてはめることで係数の値を求める。これは、$\boldsymbol{w}$ を任意に固定したときの関数 $y(x, \boldsymbol{w})$ と訓練集合のデータ点との間のずれを測る**誤差関数 (error function)** の最小化で達成できる。ここでは、誤差関数として単純で広く用いられている**二乗和誤差 (sum-of-squares error)** を用いる。式で書けば、

$$E(\boldsymbol{w}) = \frac{1}{2}\sum_{n=1}^{N}\{y(x_n, \boldsymbol{w}) - t_n\}^2 \tag{1.2}$$

となる。ただし、後で便利なように係数 $1/2$ をかけている。

このように、$E(\boldsymbol{w})$ をできるだけ小さくするような $\boldsymbol{w}$ を選ぶことで曲線当てはめ問題を解くことができる。では、誤差関数を最小にする解 $\boldsymbol{w}^* = \{w_i\}$ を求める。誤差関数を最小にする解をもとめるには、$E(\boldsymbol{w})$ を w_i について微分し、その微分がゼロになるような w_i を求めればよい。つまり、

$$\frac{\delta E}{\delta w_i} = 0 \tag{1.3}$$

を求める。はじめに、$E(\boldsymbol{w})$ を w_i について微分する。

$$\begin{aligned}\frac{\delta E}{\delta w_i} &= \frac{1}{2}\sum_{n=1}^{N}\left\{2\left(\sum_{j=0}^{M} w_j x_n^j - t_n\right) - x_n^i\right\} \\ &= \sum_{n=1}^{N}\left(x_n^i \sum_{j=0}^{M} w_j x_n^j\right) - \sum_{n=1}^{N} t_n x_n^i\end{aligned}$$

ここで、$\frac{\delta E}{\delta w_i} = 0$ を求めるので、

$$\sum_{n=1}^{N}\left(x_n^i \sum_{j=0}^{M} w_j x_n^j\right) = \sum_{n=1}^{N} t_n x_n^i \tag{1.4}$$

左辺を変形すると、

$$\begin{aligned}(\text{左辺}) &= \sum_{n=1}^{N}\left\{x_n^i\left(w_0 x_n^0 + w_1 x_n^1 + \cdots + w_M x_n^M\right)\right\} \\ &= \sum_{n=1}^{N}\left\{w_0 x_n^{i+0} + w_1 x_n^{i+1} + \cdots + w_M x_n^{i+M}\right\} \\ &= \sum_{n=1}^{N}\sum_{j=0}^{M}\left(w_j x_n^{i+j}\right) \\ &= \sum_{j=0}^{M}\left\{\sum_{n=1}^{N}(x_n)^{i+j}\right\} w_j\end{aligned}$$

また、

$$A_{ij} = \sum_{n=1}^{N}(x_n)^{i+j}, T_i = \sum_{n=1}^{N} t_n (x_n)^i \tag{1.5}$$

とおくと、

$$\sum_{j=0}^{M} A_{ij} w_j = T_i \tag{1.6}$$

となる。これは線形方程式であり、これを解くことで、誤差関数を最小にする解 $\boldsymbol{w}^*$ を求めることができる。

多項式の次数 M を選ぶのは、**モデル比較 (model comparison)** あるいは**モデル選択 (model selection)** と呼ばれる問題である。例として図 1.2 に、$M = 0, 1, 3, 9$ の場合のフィッティング結果を示す。

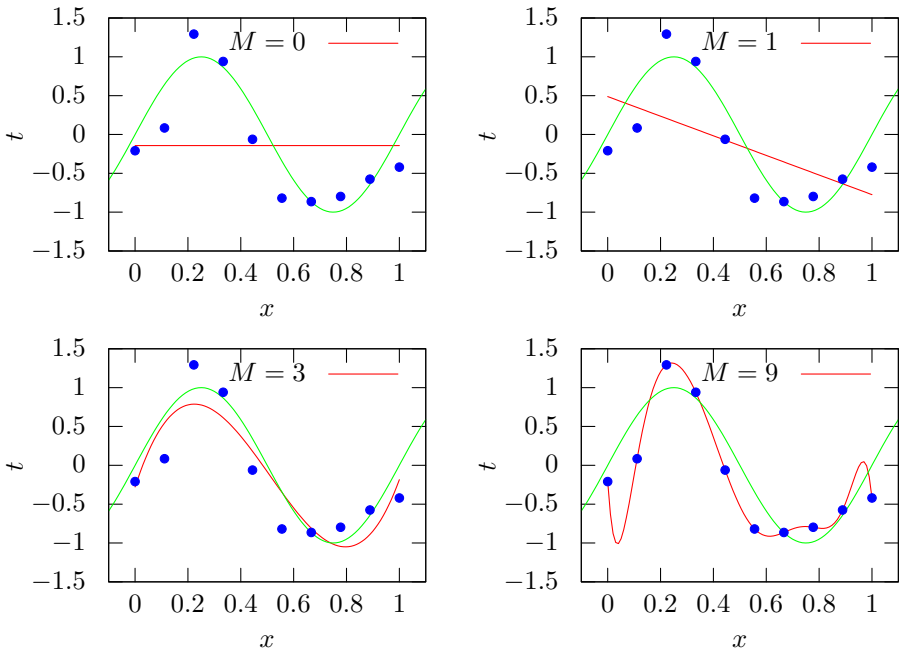

図 1.2　いろいろな次数 M のプロット

3 次の場合が $\sin(2\pi x)$ に最もよく当てはまっている。次数をもっと高い $M = 9$ にすると、訓練データにはよく当てはまっているが、$\sin(2\pi x)$ の表現としては明らかに不適切である。このような振る舞いは**過学習 (過適合;over-fitting)** として知られている。

汎化性能が M にどう依存するかを定量的に評価するために、100 個のデータ点からなる独立したテスト集合を、訓練集合と同じやり方で生成する。すると、選んだ M の各値に対して、訓練データに対して 1.2 で与えられる $E(\boldsymbol{w}^*)$ が計算できるが、テスト集合に対しても同じように $E(\boldsymbol{w}^*)$ を計算することができる。このとき、

$$E_{\mathrm{RMS}} = \sqrt{2E(\boldsymbol{w}^*)/N} \tag{1.7}$$

で定義される**平均二乗平方根誤差 (root-mean-square error, RMS error)** を用いると便利なことがある。N で割ることによってサイズの異なるデータ集合を比較することができ、平方根をとることによって、E_{RMS} は目的変数 t と同じ尺度であることが保証される。いろいろな M に対する訓練集合とテスト集合の RMS 誤差のグラフを図 1.3 に示す。図 1.3 を見てわかるように、M が小さいと、$\sin(2\pi x)$ の振動を捉えることができないため、訓練集合とテスト集合の RMS 誤差が大きくなる。また、M が大きいと、訓練集合に対してはよく当てはまるが、テスト集合に対しては当てはまらなくなる。

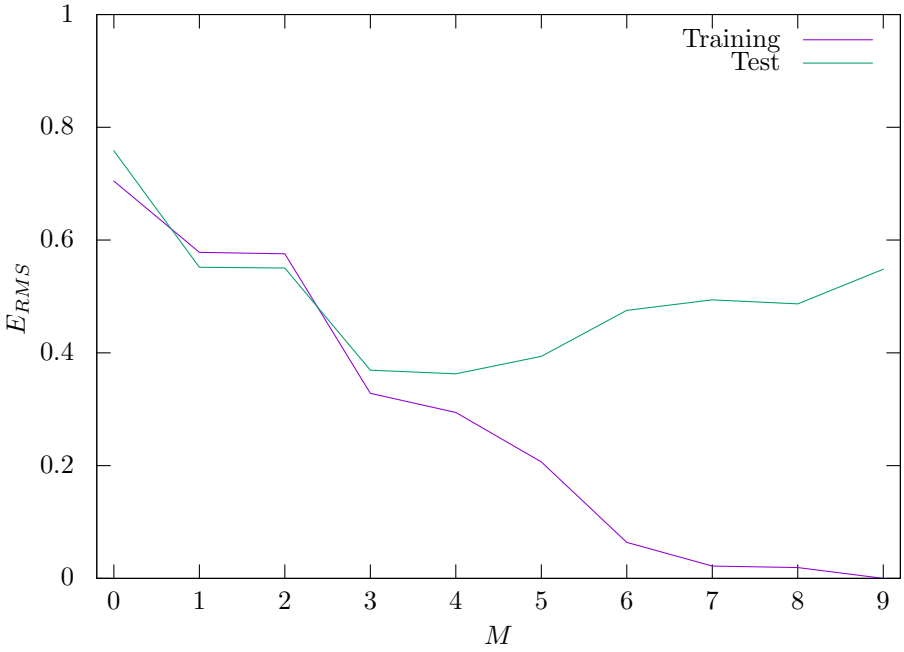

図 1.3　訓練集合とテスト集合の RMS 誤差