第3章　線形回帰モデル

回帰の目的は、観測点と対応する目標値からなる訓練データが与えられたとき、新たなxに対するtの値を予測することである。
本章では、回帰モデルの中でも、入力に対して基底関数と呼ばれる非線形な関数を適用した値に対して、線形結合を行ったモデルを扱う。

最も簡単な線形回帰モデルは入力変数の線形結合である。

$$y(x, w) = w_0 + w_1 x_1 + \cdots + w_D x_D$$

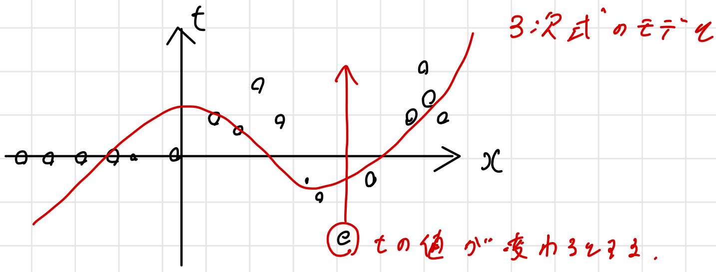
これでは表現力が乏しいため、入力変数に関して非線形な関数の線形結合を考える。

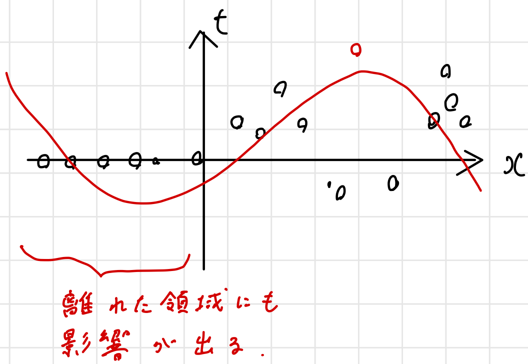
$$\begin{aligned} y(x, w) &= w_0 + \sum_{j=1}^{M-1} w_j \phi_j(x) \\ &= \sum_{j=0}^{M-1} w_j \phi_j(x) \qquad (ここで\ \phi_0(x) = 1\ とした) \\ &= w' \phi(x) \end{aligned}$$

例えば、多項式モデルであれば、基底関数は次のようになる。

$$\phi(x) = \begin{pmatrix} 1 \\ x \\ x^2 \\ \vdots \\ x^K \end{pmatrix}$$

（スプライン関数について）自信はないが、たぶんこうかな、ということだけ記す。
「多項式は入力変数の大域的な関数であるため、入力空間のある領域の変化が他の領域に及んでしまう」とは、あるサンプルにおける目標変数が変動したら、そのサンプルのx以外の領域の回帰結果も変動することを言っているのだと思う。

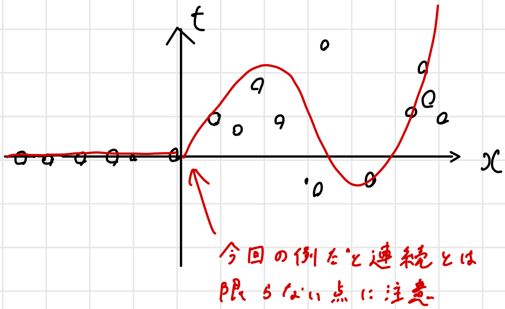




基底関数を次のようなスプライン関数とすると、異なる領域に影響を及ぼさなくなる。

$$\phi(x) = \begin{cases} \begin{pmatrix} 1 \\ x \\ x^2 \\ x^3 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} & (x \leq 0) \\ \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ x \\ x^2 \\ x^3 \end{pmatrix} & (x \geq 0) \end{cases}$$

回帰の結果のイメージ



他にも次のような基底関数がある。

ガウス基底関数

$$\phi_j(x) = \exp\left\{-\frac{(x-\mu_j)^2}{2s^2}\right\}$$

シグモイド基底関数

$$\phi_j(x) = \sigma\left(\frac{x-\mu_j}{s}\right) \qquad \left(ここで\ \sigma(a) = \frac{1}{1+e^{-a}}\right)$$

フーリエ基底

ちょっとイメージつかなかった。。基底関数がある単一の周波数に対応するとは？？

3.1.1 最尤推定と最小二乗法

既に学習した内容については割愛する。誤差にガウスノイズを仮定して、最尤推定によってパラメータ推定を行うと、最小二乗法に一致する。

最小二乗法の解は、正規方程式を解くことによって得られる。

$$W_{ML} = (\Phi^T \Phi)^{-1} \Phi^T t$$

計画行列

$$\begin{pmatrix} \phi_0(x_1) & \cdots & \phi_{M-1}(x_1) \\ \vdots & & \\ \phi_0(x_N) & \cdots & \phi_{M-1}(x_N) \end{pmatrix}$$

ここで、

$$\Phi^{\dagger} \equiv (\Phi'\Phi)^{-1}\Phi'$$

を ムーアペンローズの擬似逆行列 というらしい。

ガウスノイズの精度パラメータβについても、次の最尤推定量が得られる。

$$\frac{1}{\beta_{ML}} = \frac{1}{N}\sum_{n=1}^{N}\left(t_n - w_{ML}'\phi(x_n)\right)^2$$

3.1.2 最小二乗法の幾何学

水色の教科書p42あたりの内容と同等のことが書いてある。
軽くおさらいする。

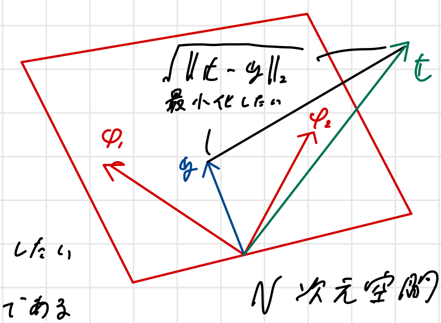
結論から言うと、最小二乗法によるパラメータ推定は、真の目標変数ベクトルを、ある空間に正射影したベクトルを求めることに相当する。

・ある空間とは、計画行列の列べくとるで張られる空間である。

・正射影したベクトルは、予測ベクトル $y\ (=\Phi w_{ML})$ に相当する。

$$\Phi = (\varphi_1\ \ \varphi_2\ \ \cdots\ \ \varphi_m)$$

↓ N次元ベクトル



右図でいう、$\sqrt{\|t-y\|_2}$ を最小化したい というのは、y が t の正射影である ということと同義である。実際に、

$$y = \Phi w_{ML} = \Phi(\Phi'\Phi)^{-1}\Phi'\,t$$

ここが $\{\varphi_i\}$ の張る空間への正射影となっている。(*)

(※) これを示す。（演習 3.2）

$\boldsymbol{t} - \boldsymbol{y}$ が、任意の φ_i と直交することを示せば良い。

$$\varphi_i'(\boldsymbol{t} - \boldsymbol{y}) = \varphi_i'(\mathbb{I} - \Phi(\Phi'\Phi)^{-1}\Phi')\boldsymbol{t}$$

これを m 個縦に並べて、

$$\underbrace{\begin{pmatrix} \varphi_1' \\ \varphi_2' \\ \vdots \\ \varphi_m' \end{pmatrix}}_{=\Phi'}(\boldsymbol{t} - \boldsymbol{y}) = \underbrace{\begin{pmatrix} \varphi_1' \\ \varphi_2' \\ \vdots \\ \varphi_m' \end{pmatrix}}_{=\Phi'}(\mathbb{I} - \Phi(\Phi'\Phi)^{-1}\Phi')\boldsymbol{t}$$

$$= (\Phi' - \underbrace{\Phi'\Phi(\Phi'\Phi)^{-1}}_{I}\Phi')\boldsymbol{t}$$

$$= (\Phi' - \Phi')\boldsymbol{t}$$

$$= \boldsymbol{0}$$

よって示せた。