

1

1.1

訓練集合として、 N 個の観測地 x を並べた $\mathbf{m}x := (x_1, \dots, x_N)^T$ とそれぞれ対応する観測値 t を並べた $\mathbf{m}t := (t_1, \dots, t_N)^T$ が与えられたとする。そして目標データ集合 $\mathbf{m}t$ は、まず $\sin(2\pi x)$ の関数値を作成したのち、ガウス分布に従う小さなランダムノイズを加えて対応する t_n を作った。すなわち

$$t_n \sim \sin(2\pi x_n) + \varepsilon \quad \varepsilon \sim N(0, \sigma^2)$$

となるデータを作成している。このようにして生成されたデータは多くの現実データ集合の持つ性質をよく表している。すなわち、データはこれから学習しようとする規則性を保持しているが、それぞれの観測値はランダムノイズによって不正確なものになっている。このノイズは、放射性崩壊のように、本質的に確率的なランダムプロセスによる場合もあるが、多くはそれ自身は観測されない信号源の変動によるものである。

我々の目標は、