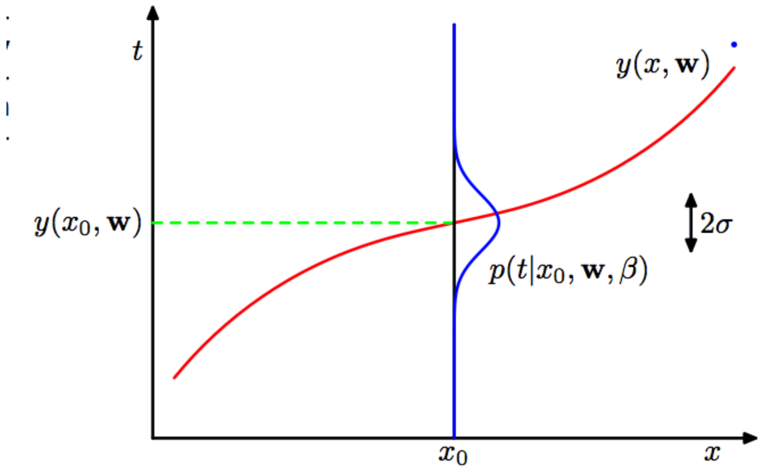


1　復習

$$p(\boldsymbol{w}, \mathcal{D}) = p(\mathcal{D})p(\boldsymbol{w}|\mathcal{D}) = p(\boldsymbol{w})p(\mathcal{D}|\boldsymbol{w}) \tag{1}$$

$$p(\boldsymbol{w}|\mathcal{D}) = \frac{p(\mathcal{D}|\boldsymbol{w})p(\boldsymbol{w})}{p(\mathcal{D})} \tag{2}$$

2　1.2.6 ベイズ曲線フィッティング

$$p(\boldsymbol{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta) = \frac{p(\mathbf{t}|\mathbf{x}, \boldsymbol{w}, \beta)p(\boldsymbol{w}|\alpha)}{p(\mathbf{x}, \mathbf{t}, \alpha, \beta)} \tag{3}$$

$$p(\boldsymbol{w}|\mathbf{x}, \mathbf{t}, \alpha, \beta) \propto p(\mathbf{t}|\mathbf{x}, \boldsymbol{w}, \beta)p(\boldsymbol{w}|\alpha) \qquad (1.66) \tag{4}$$

上記は、最大事後確率 (MAP; maximum a posteriori) となるが、最大となる $\mathbf{w}$ の点を推定しているだけ。完

全なベイズというには $\mathbf{w}$ のすべての値で積分する必要がある。(すなわち、いろいろな w がそれぞれの事後確率で出力され、それぞれの w にて x から t を予想する)。問題としては、x から t を予測する問題で、訓練例として $\mathbf{x}, \mathbf{t}$ が与えられるので次のようになる。

$$p(t|x, \mathbf{x}, \mathbf{t}) = \int p(t|x, \boldsymbol{w}) p(\boldsymbol{w}|\mathbf{x}, \mathbf{t}) d\boldsymbol{w} \qquad (1.68) \tag{5}$$

ここで式 (1.60) から、

$$p(t|x, \mathbf{t}, \beta) = \mathcal{N}(t|y(x, \boldsymbol{w}), \beta^{-1}) \qquad (1.60)$$

$$p(t|x, \mathbf{t}) = \mathcal{N}(t|y(x, \boldsymbol{w})) \qquad (1.60')$$

式 (1.66) から、事後分布

$$p(\boldsymbol{w}|\mathbf{x}, \mathbf{t}) = \frac{p(\mathbf{t}|\mathbf{x}, \boldsymbol{w}, \beta) p(\boldsymbol{w}|\alpha)}{\sum p(\mathbf{t}|\mathbf{x}, \boldsymbol{w}, \beta) p(\boldsymbol{w}|\alpha)}$$

となり、これらから、式 (1.68) は解析的に解くことができ、

$$p(t|x, \mathbf{x}, \mathbf{t}) = \mathcal{N}(t|m(x), s^2(x)) \qquad (1.69)$$

ここで、平均と分散は、

$$m(x) = \beta \boldsymbol{\phi}(x)^{\mathrm{T}} \mathbf{S} \sum_{n=1}^{N} \phi(x_n) t_n \qquad (1.70)$$

$$s^2(x) = \beta^{-1} + \boldsymbol{\phi}(x)^{\mathrm{T}} \mathbf{S} \boldsymbol{\phi}(x) \qquad (1.71)$$

となり行列 $\mathbf{S}$ は、

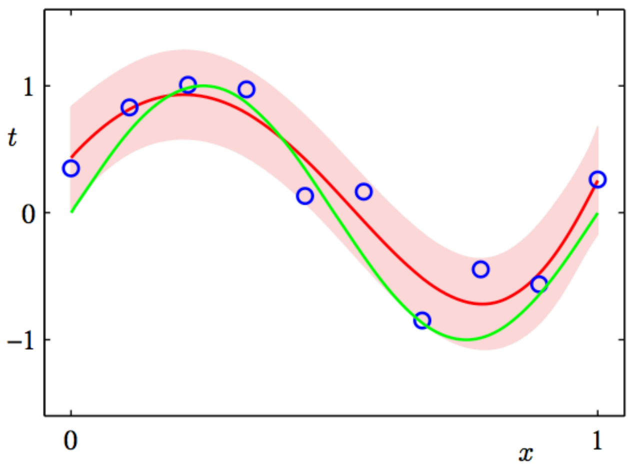
$$S^{-1} = \alpha \mathbf{I} + \beta \sum_{n=1}^{N} \boldsymbol{\phi}(x_n) \boldsymbol{\phi}(x_n)^{\mathrm{T}}$$

で与えられる。ここで $\mathbf{I}$ は単位行列で、$\phi_i(x) = x^i (i = 0, ..., M)$

3 1.3 モデル選択

最小二乗法の例にて、最適な次数があった。次数のようなモデルの複雑さを、最小二乗法では正則化係数 λ により制御するし、他の学習手法もそのようなパラメータがある。これらは未知の事例に対する予測が最も良いものを選ぶ。

過学習があるので、訓練データとは別に、**確認用集合 (検証用集合; validation set)** を用意しておき、これらのパラメータを調整し、最後に、**テスト用集合 (test set)** にて性能の最終評価を行う。



1
t
0
−1
0
x
1