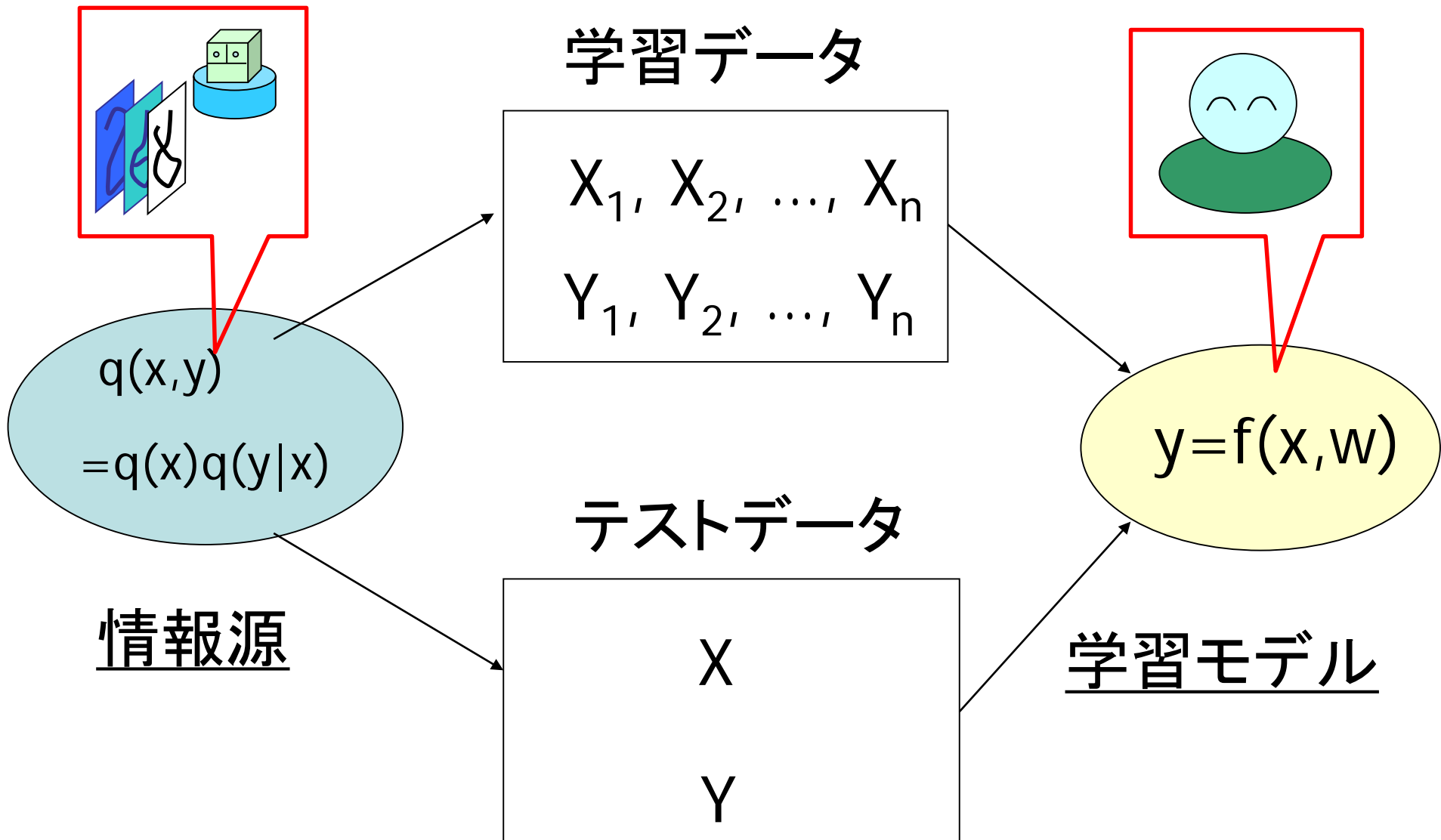


誤差逆伝播法

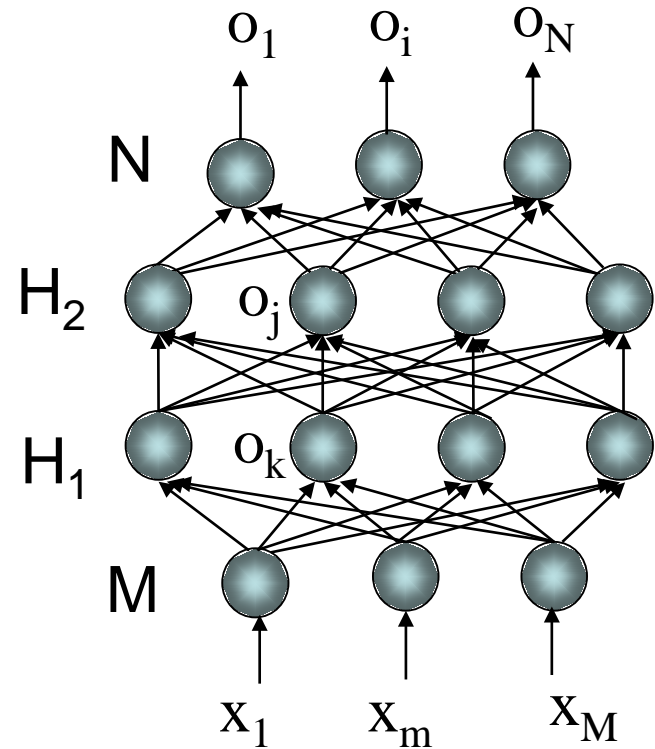
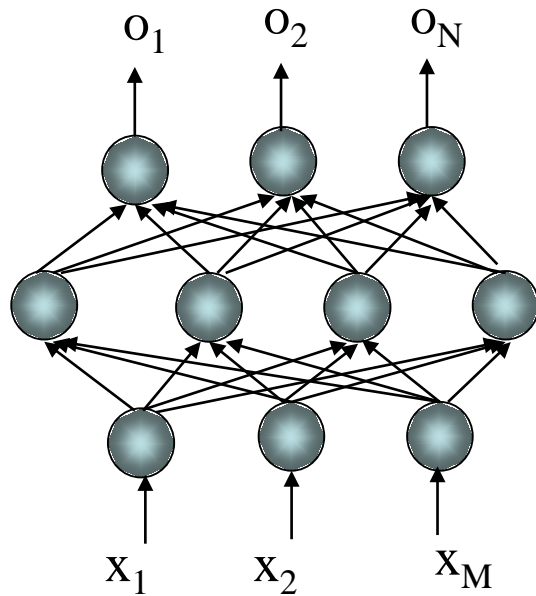
渡辺澄夫
東京工業大学

教師あり学習



深層学習

深い階層のモデルを定義することは簡単にできる。



$$o_i = \sigma \left(\sum_{j=1}^{H_2} w_{ij} \sigma \left(\sum_{k=1}^{H_1} w_{jk} \sigma \left(\sum_{m=1}^M w_{km} x_m + \theta_k \right) + \theta_j \right) + \theta_i \right)$$

階層ごとに別の記号を使うのがメンドウなので同じ記号を使うが別のもの。

二乗誤差

学習モデルの出力を (o_i) , 教師データを $y = (y_i)$ とする。

$$\text{二乗誤差は } E(w) = \frac{1}{2} \sum_{i=1}^N (o_i - y_i)^2 \quad (N: \text{出力の次元})$$

学習モデルの出力 (o_i) はパラメータ (w, θ) の関数なので、 $E(w)$ の w に関する微分 $\nabla E(w)$ を計算することができる。

確率降下法

各時刻 t において学習データ (X_i, Y_i) をランダムに選んで

$$\Delta w = w(t+1) - w(t) = -\eta(t) \nabla \{ (Y_i - f(X_i, w(t)))^2 \}$$

とする (∇ はパラメータに関する微分を表す)。

◎ 1個ずつデータを取ってくるだけで学習できる。

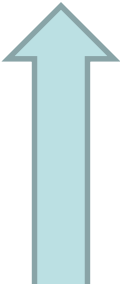
◎ $\eta(t)$ は下記を満たすように定める。

$$\eta(t) \rightarrow 0$$


$$\eta(0) + \eta(1) + \dots + \eta(t) \rightarrow \infty$$

計算の手順


(3) 中間2→出力


$$o_i = \sigma \left(\sum_{j=1}^{H_2} w_{ij} o_j + \theta_i \right)$$


(2) 中間1→中間2


$$o_j = \sigma \left(\sum_{k=1}^{H_1} w_{jk} o_k + \theta_j \right)$$

(1) 入力→中間1



$$o_k = \sigma \left(\sum_{m=1}^M w_{km} x_m + \theta_k \right)$$

(4) 中間2→出力


$$\delta_i = (o_i - y_i) o_i (1 - o_i)$$

$$\Delta w_{ij} = -\eta(t) \delta_i o_j$$
$$\Delta \theta_i = -\eta(t) \delta_i$$

(5) 中間1→中間2


$$\delta_j = \sum_i \delta_i w_{ij} o_j (1 - o_j)$$

$$\Delta w_{jk} = -\eta(t) \delta_j o_k$$
$$\Delta \theta_j = -\eta(t) \delta_j$$

(6) 入力→中間1


$$\delta_k = \sum_j \delta_j w_{jk} o_k (1 - o_k)$$

$$\Delta w_{km} = -\eta(t) \delta_k x_m$$
$$\Delta \theta_k = -\eta(t) \delta_k$$

学習と汎化

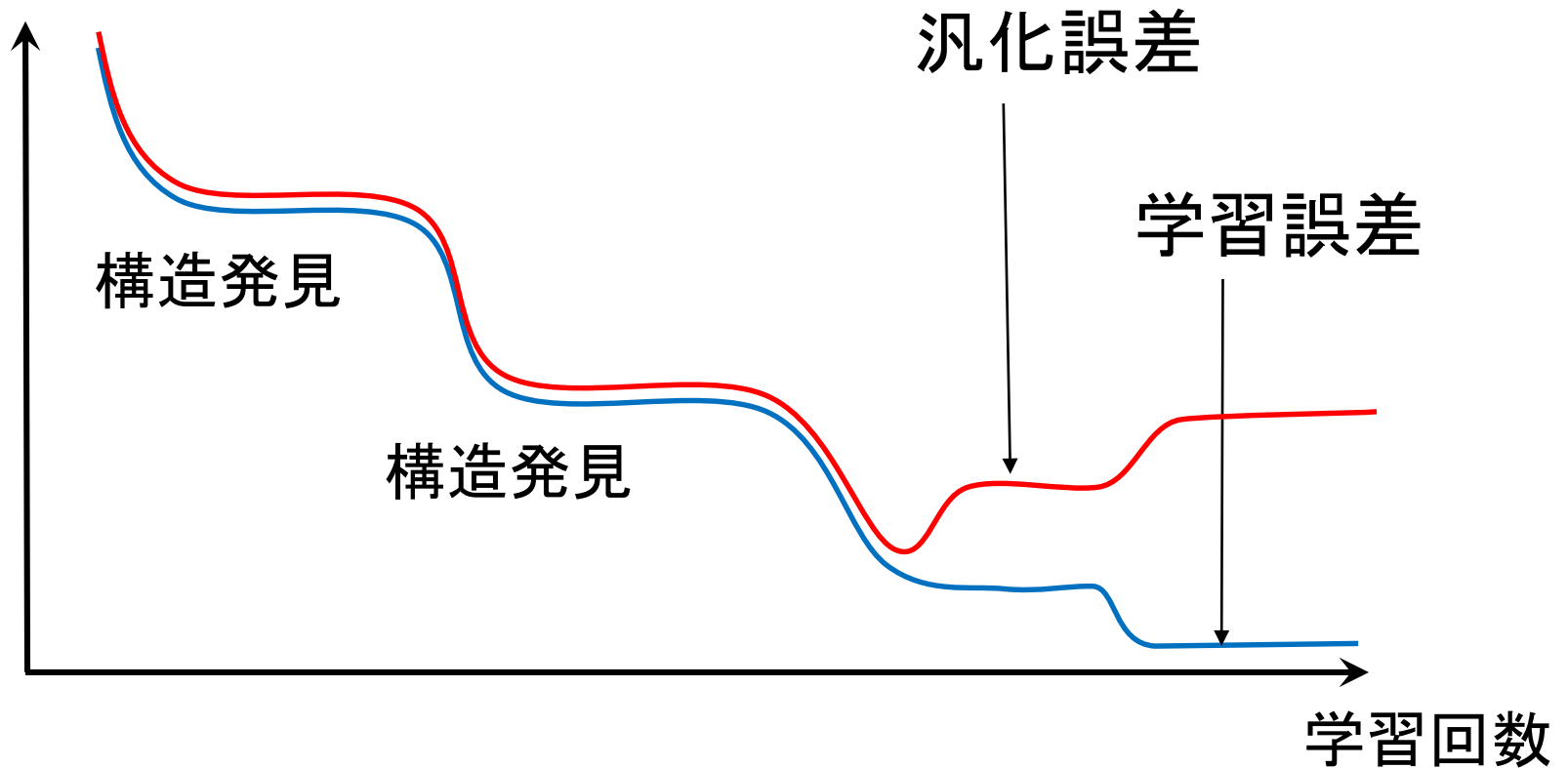
学習誤差関数

$$E(w) = (1/n) \sum_{i=1}^n (Y_i - f(X_i, w))^2$$

汎化誤差関数

$$F(w) = \iint (y - f(x, w))^2 q(x, y) dx dy$$

学習誤差と汎化誤差



良好な学習結果を得るには途中で学習を停止する必要がある。
しかし、どこで停止させるのが最適であるかはわかっていない。