超平面からの距離による評価

- 2クラス識別器

$$g(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x}$$

- 重みベクトル $\boldsymbol{w}$ と識別線 $g(\boldsymbol{x}) = \boldsymbol{w}^T \boldsymbol{x} = 0$ との距離

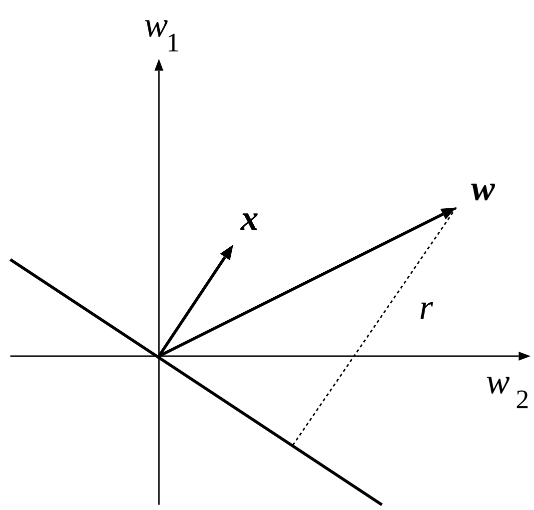
$$r = \frac{\|\boldsymbol{w}^T \boldsymbol{x}\|}{\|\boldsymbol{x}\|}$$

- 評価関数 $J = \|\boldsymbol{w}^T \boldsymbol{x}\|$ として逐次修正を行う．ただし，正しく識別されたときには $J = 0$ とする．

$$J(\boldsymbol{w}) = \begin{cases} -\boldsymbol{w}^T \boldsymbol{x} & (\boldsymbol{x} \in \omega_1 \text{かつ} \boldsymbol{w}^T \boldsymbol{x} < 0) \\ \boldsymbol{w}^T \boldsymbol{x} & (\boldsymbol{x} \in \omega_2 \text{かつ} \boldsymbol{w}^T \boldsymbol{x} > 0) \\ 0 & (\text{その他のとき}) \end{cases}$$

- 最急降下法はパーセプトロンと等価

$$\begin{aligned} \boldsymbol{w}' &= \boldsymbol{w} - \rho \frac{\partial J}{\partial \boldsymbol{w}} \\ &= \begin{cases} \boldsymbol{w} + \rho \boldsymbol{x} & (\boldsymbol{x} \in \omega_1 \text{かつ} \boldsymbol{w}^T \boldsymbol{x} < 0) \\ \boldsymbol{w} - \rho \boldsymbol{x} & (\boldsymbol{x} \in \omega_2 \text{かつ} \boldsymbol{w}^T \boldsymbol{x} > 0) \\ \boldsymbol{w} & (\text{その他のとき}) \end{cases} \qquad (1) \end{aligned}$$



- 多層パーセプトロン (非線形写像)
- ユニット j への入力 (ユニット i の出力の重みつき線形和)

$$h_{jp} = \sum_i w_{ij} g_{ip}$$

- ユニット j の出力 (非線形写像)

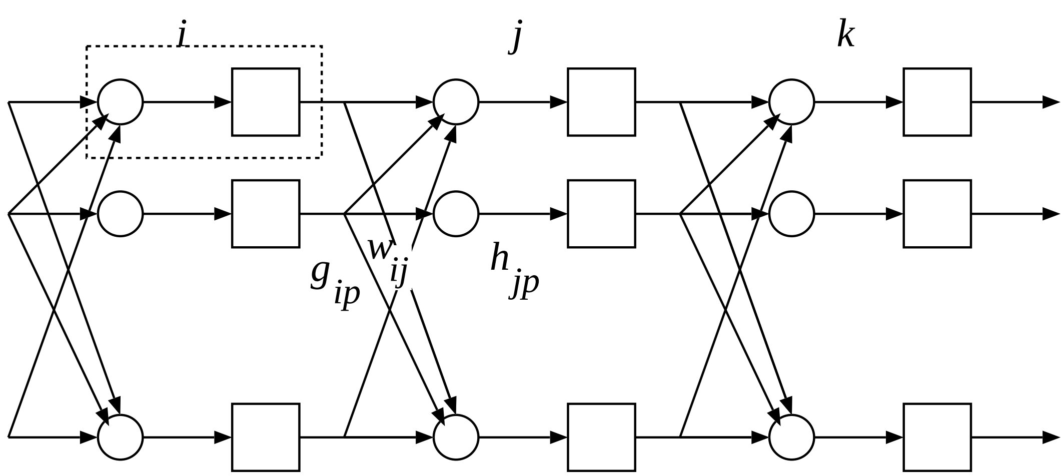
$$g_{jp} = f_j(h_{jp})$$

- 出力層: ℓ，出力: $g_{\ell p}$，教師信号 $b_{\ell p}$
- 誤差:

$$J_p = \frac{1}{2} \sum_\ell (g_{\ell p} - b_{\ell p})^2$$

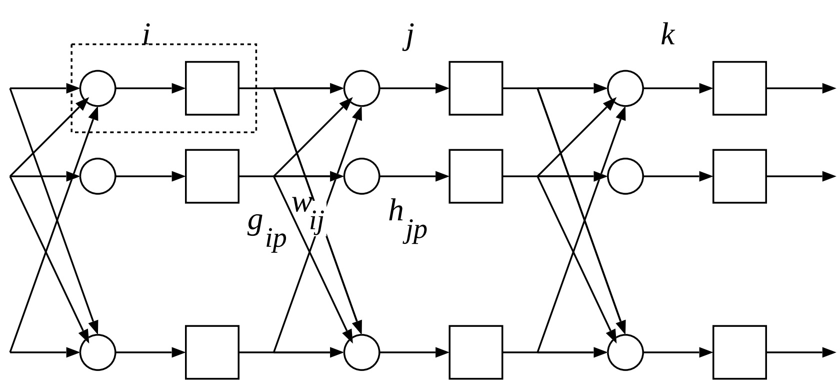
- 全パターンに対する誤差:

$$J = \sum_p J_p$$

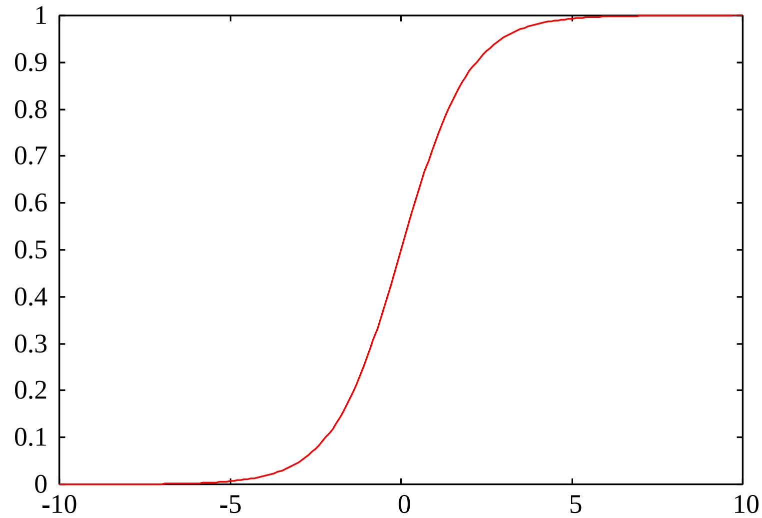


誤差最小化

- Jの最小化
$$\frac{\partial J_p}{\partial w_{ij}} = \frac{\partial J_p}{\partial h_{ip}}\frac{\partial h_{jp}}{\partial w_{ij}} = \varepsilon_{jp} g_{jp}$$
ただし$\varepsilon_{jp} = \partial J_p / \partial h_{jp}$.
- 最急降下法
$$w'_{ij} = w_{ij} - \rho \varepsilon_{jp} g_{jp}$$
- ε_{jp}の決め方.
$$\varepsilon_{jp} = \frac{\partial J_p}{\partial h_{jp}} = \frac{\partial J_p}{\partial g_{jp}}\frac{\partial g_{jp}}{\partial h_{jp}} = \frac{\partial J_p}{\partial g_{jp}} f'_j(h_{jp})$$
- $\partial J_p / \partial g_{jp}$の決め方.
  - ユニットjが出力層にあるとき($J_p = \frac{1}{2}\sum_j (g_{jp} - b_{jp})^2$)
  $$\frac{\partial J_p}{\partial g_{jp}} = g_{jp} - b_{jp}$$
  - ユニットjが中間層にあるとき($h_{kp} = \sum_j w_{jk} g_{jp}$)
  $$\frac{\partial J_p}{\partial g_{jp}} = \sum_k \frac{\partial J_p}{\partial h_{kp}}\frac{\partial h_{kp}}{\partial g_{jp}} = \sum_k \varepsilon_{kp} w_{jk}$$



誤差逆伝播法 (Back Propagation Method)



- しきい値関数の近似 (シグモイド関数: Sigmoid Function)

$$S(u) = \frac{1}{1 + \exp -u}, S'(u) = S(u)(1 - S(u))$$

- f_j をシグモイド関数にすると $(g_{jp} = f_j(h_{jp}))$

$$f'_j = g_{jp}(1 - g_{jp})$$

- 重み更新

$$\begin{aligned} w'_{ij} &= w_{ij} - \rho \varepsilon_{jp} g_{jp} \\ \varepsilon_{jp} &= \begin{cases} (g_{jp} - b_{jp}) g_{jp} (1 - g_{jp}) \\ \qquad \text{(ユニット } j \text{ が出力層にあるとき)} \\ (\sum_k \varepsilon_{kp} w_{jk}) g_{jp} (1 - g_{jp}) \\ \qquad \text{(ユニット } j \text{ が中間層にあるとき)} \end{cases} \end{aligned}$$

汎化 (generalization)

- 一般的には: 個別の事例から背後に潜む一般法則を抽出すること
- 学習理論では: 未知パターンに対してどの程度正しい出力を示せるか(予測能力)であり，一般法則の抽出とは関係ない.
- 教師つき学習では，学習は入出力間関数のパラメータ推定問題
- パーセプトロンでは線形重みの推定(重回帰分析と等価)
- ニューラルネットワークでは非線形近似
- 非線形関数を用いれば複雑な関数を高精度で近似可能
- 少数の学習パターンを多数のパラメータを用いて近似すれば学習誤差は減少するが，未知学習パターンには対応できなくなる.
- パラメータの数が多くなれば十分な学習パターンを準備することは困難(次元の呪い)
- パラメータの数に対して学習パターンが少ない場合，過学習(言葉が適切ではない)が発生する.
- 過学習は関数のオーバーフィッティングであり，少数のデータに対して多数のパラメータを用いた複雑な関数で誤差ゼロ近似をしてしまうこと.