混沌系工学特論 #9

情報科学研究科　井上純一

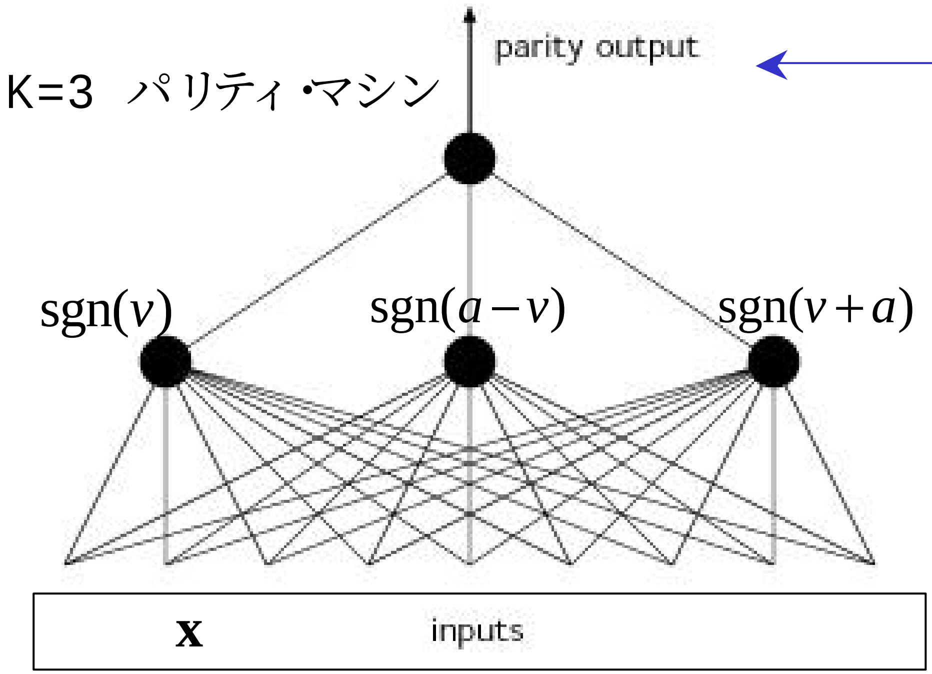
URL : http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/

平成17年1月31日　第10回（最終回）講義

実現不可能な規則のオンライン学習

前回の復習

教師機械



$$T_{a\to\infty}(v) = \mathrm{sgn}[v]$$ (学習可能となる極限)

$$S(u) = \mathrm{sgn}[u], u = \sqrt{N}(\mathbf{J}\cdot\mathbf{x})/|\mathbf{J}|$$

$$T_a(v) = \mathrm{sgn}[v(a-v)(a+v)]$$
$$v = \sqrt{N}(\mathbf{J}^0\cdot\mathbf{x})/|\mathbf{J}^0|$$

最終出力は3つのパーセプトロンの出力のパリティとなる

性能 ：実現することのできる入出力関係の数は

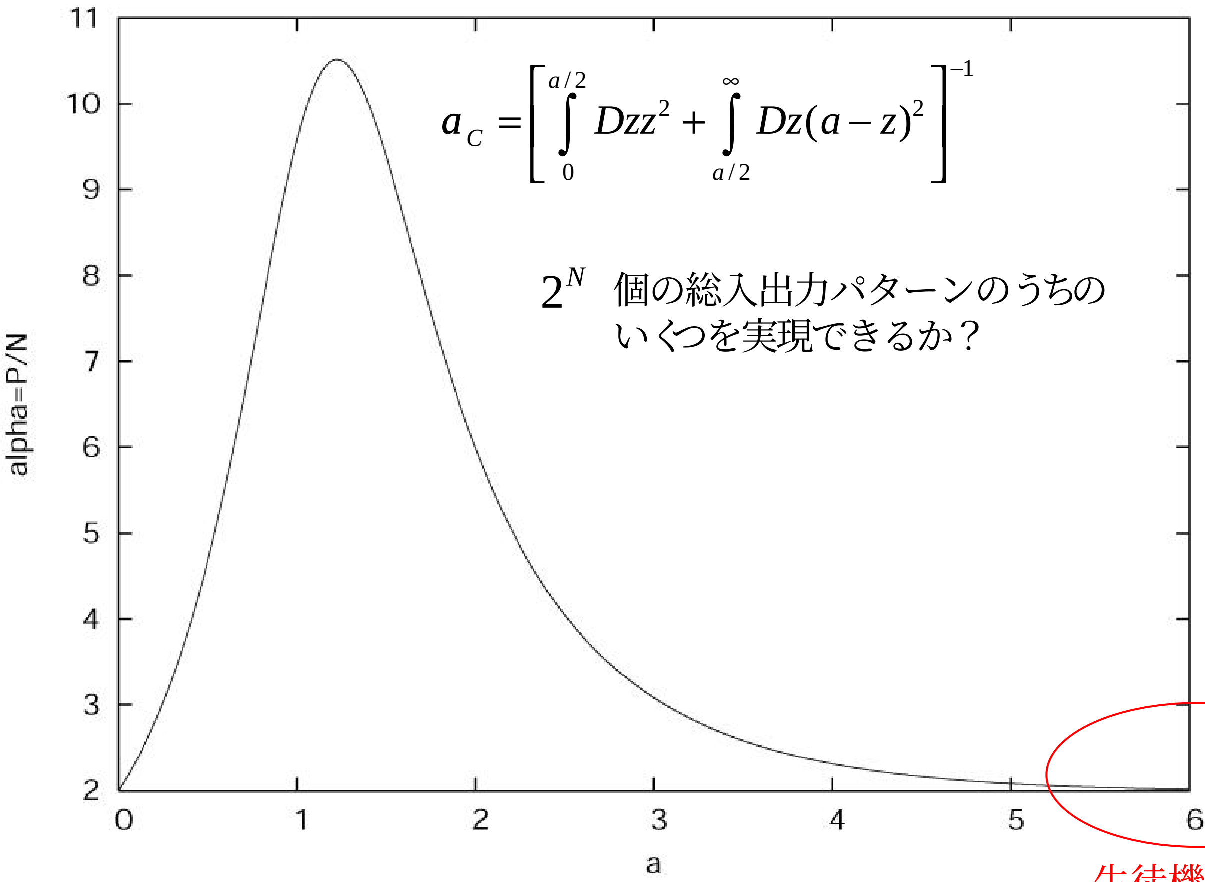
$\sim 10N$ (レプリカ法の解析による)

(単純パーセプトロンは2N)

⇒ 単純パーセプトロン(生徒機械)にとって実現不可能な規則である

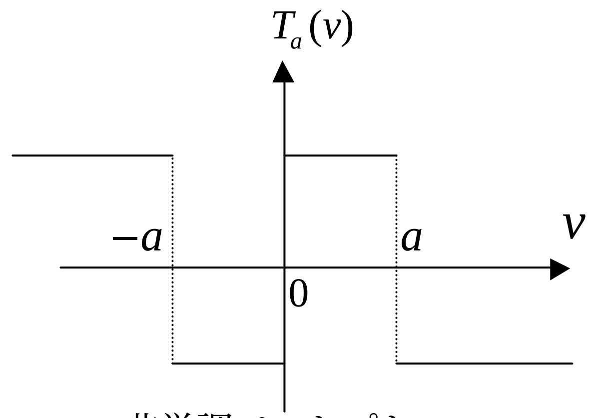
教師機械の記憶容量：補足

レプリカ法による解析によれば



$$a_C = \left[\int_0^{a/2} Dz z^2 + \int_{a/2}^{\infty} Dz (a-z)^2 \right]^{-1}$$

2^N 個の総入出力パターンのうちの
いくつを実現できるか？



非単調パーセプトロン

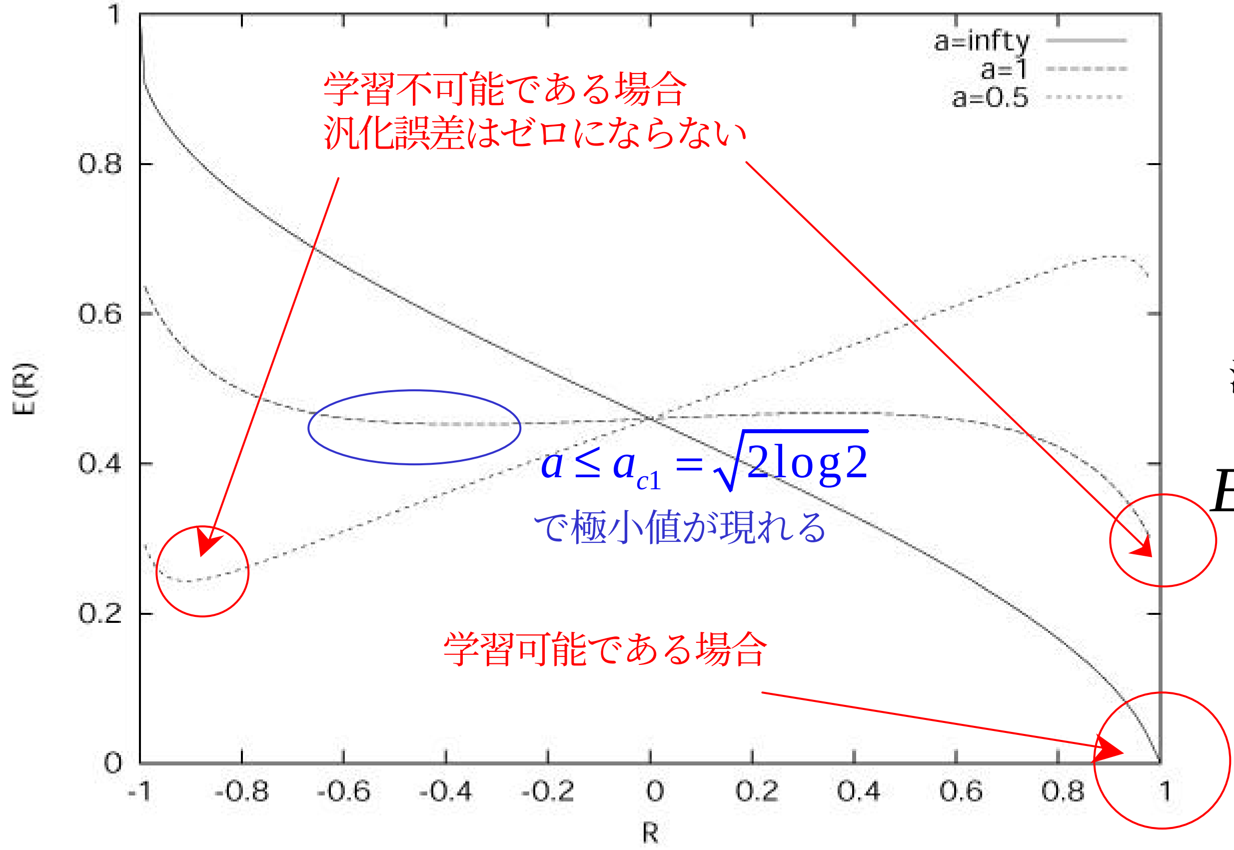
⇒ 教師機械の入出力関係は
生徒機械にとって実現不可能

生徒機械の記憶容量 $2N$

マクロな量の導入と汎化誤差

前回の復習

汎化誤差の学習則に依らない一般的性質



両機械の結合の重なり

$$R = \left(\mathbf{J}^{0} \cdot \mathbf{J}\right) / \left|\mathbf{J}^{0}\right|\left|\mathbf{J}\right|$$

生徒機械の結合の長さ

$$l = \left|\mathbf{J}\right| / \sqrt{N}$$

汎化誤差

$$E_a(R) = \left\langle \Theta\left(-T_a(v) S(u)\right) \right\rangle$$

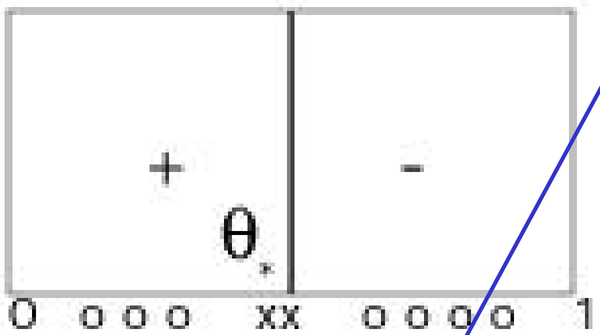
汎化誤差の例題数の増加にともなう振る舞いは具体的に学習則を与えることにより明らかとなる

パーセプトロン学習とそのダイナミックス

前回の復習

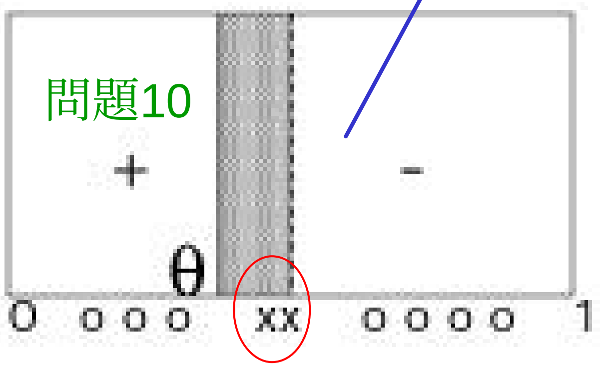
T

$$s_T(t) = \mathrm{sgn}[q_* - x(t)]$$



S

$$s_S(t) = \mathrm{sgn}[q(t) - x(t)]$$



この領域に落ちた入力に対して
間違った結果を出力する

これの高次元版を考える　　パーセプトロン学習

$$\mathbf{J}^{\mathbf{m+1}} = \mathbf{J}^{\mathbf{m}} - \Theta(-T_a(v)S(u))S(u)\mathbf{x}$$

教師/生徒の出力が異なる
場合のみ結合が修正される

マクロな量は次の微分方程式に従う　$P = aN$

$$2l\frac{dl}{da} = -2l\langle\Theta(-T_a(v)S(u))u\rangle + \langle\Theta(-T_a(v)S(u))\rangle$$

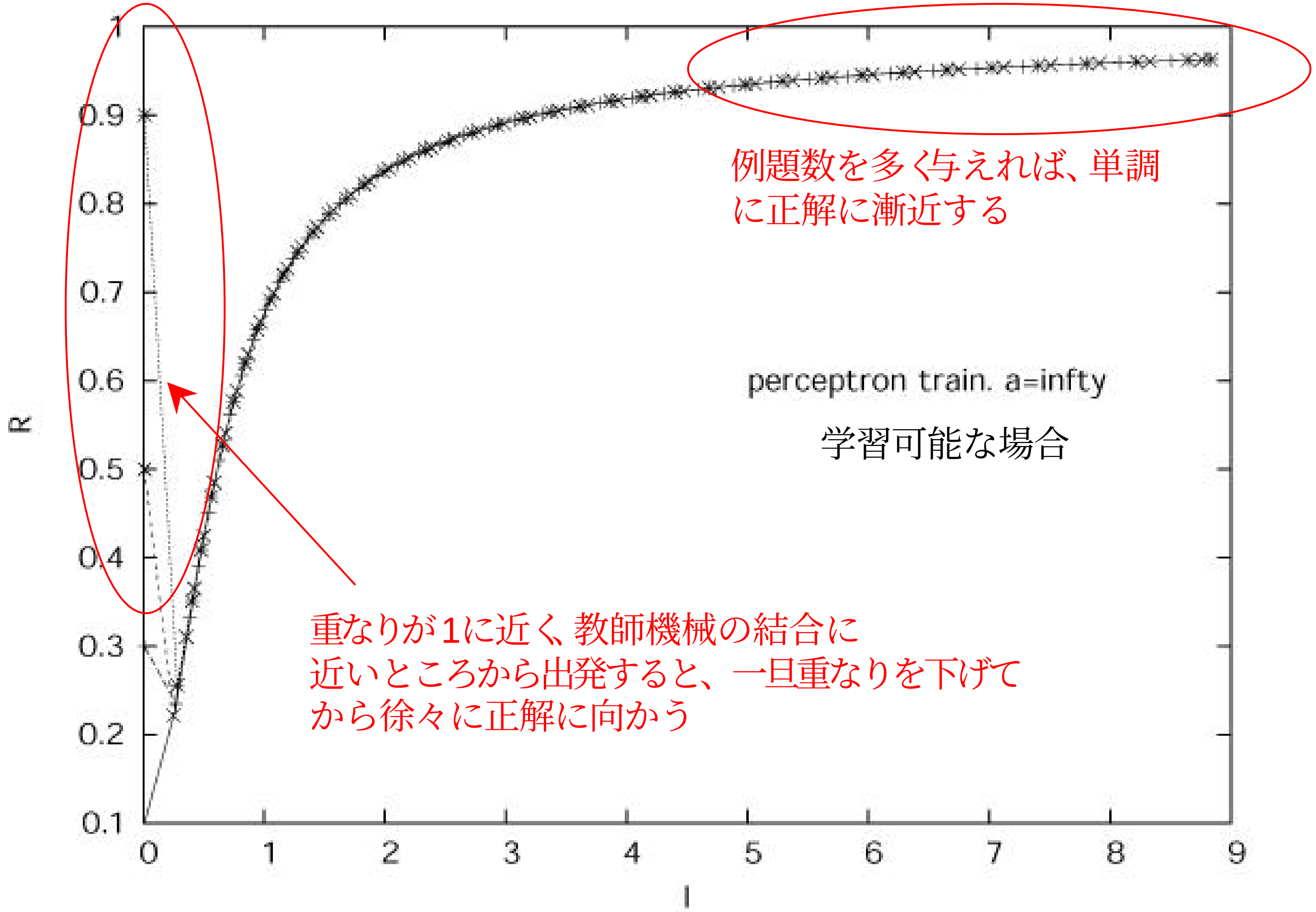
$$l^2\frac{dR}{da} = -\frac{R}{2}\langle\Theta(-T_a(v)S(u))v\rangle + l(R\langle-\Theta(T_a(v)S(u))u\rangle$$

$$-\langle\Theta(-T_a(v)S(u))v\rangle)$$

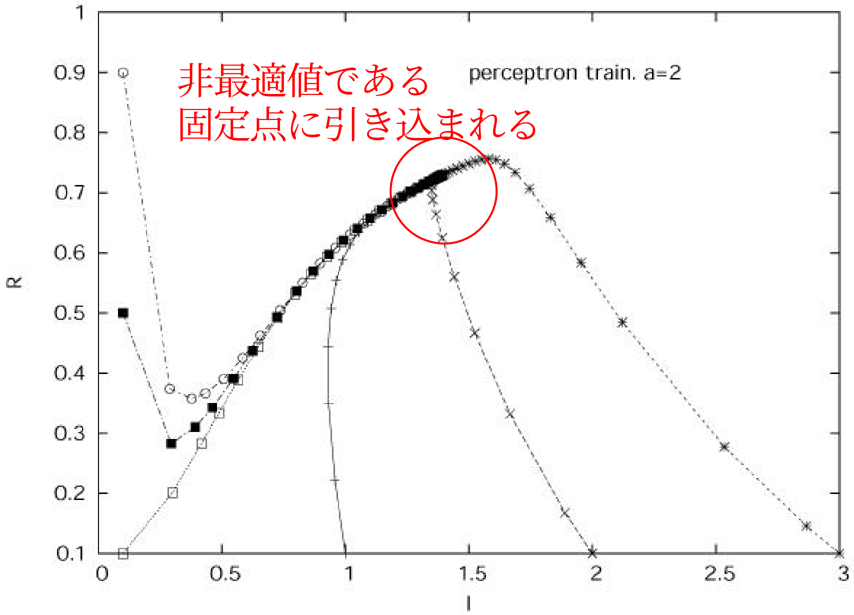
状態更新式

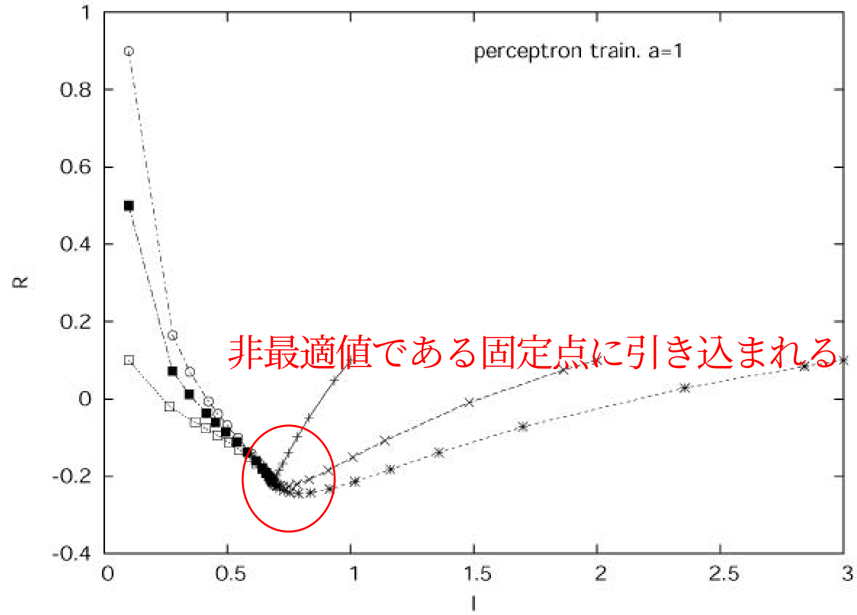
導出の詳細は講義ノート参照

マクロな量のフロー図#1



マクロな量のフロー図#2

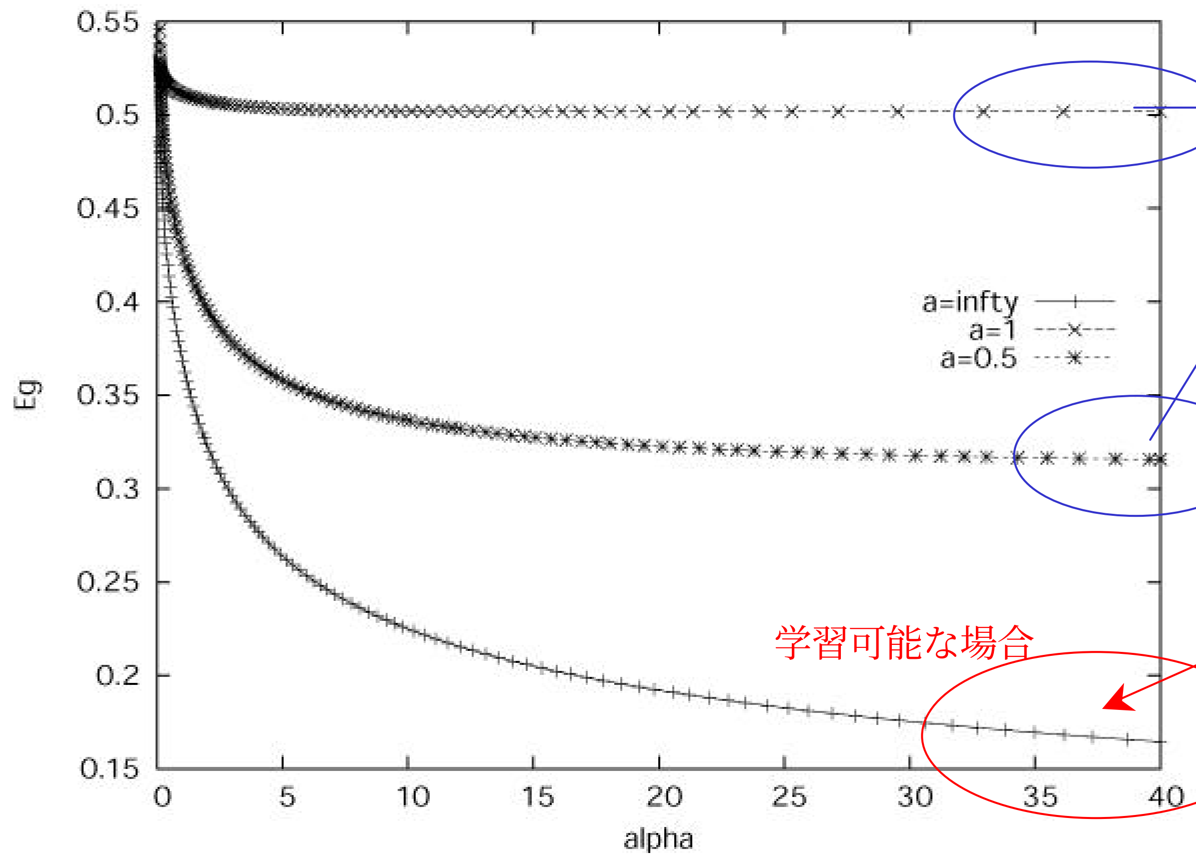




汎化誤差は

$$E_a(R)=\left\langle \Theta(-T_a(v)S(u)) \right\rangle$$ で算出される

パーセプトロン学習の学習曲線



学習不可能な場合、
非最適値に指数関数
的に漸近してしまう

学習可能な場合

バッチ学習に比べて
精度が悪い
⇒何とか加速できないか？

$a^{-1/3}$ 則に従う

学習レートの導入

パーセプトロン学習に学習レートを導入する

$$\mathbf{J}^{\mathbf{m+1}} = \mathbf{J}^{\mathbf{m}} - \underbrace{g(a)}_{\text{学習レート}}\Theta(-T_a(v)S(u))S(u)\mathbf{x}$$

マクロな発展方程式は

$$\frac{dR}{da} = \frac{1}{l^2}\left[-\frac{R}{2}g(a)^2 E_a(R) + g(a)\left[F_a(R)R - G_a(R)\right]l\right] \equiv L(g(a))$$

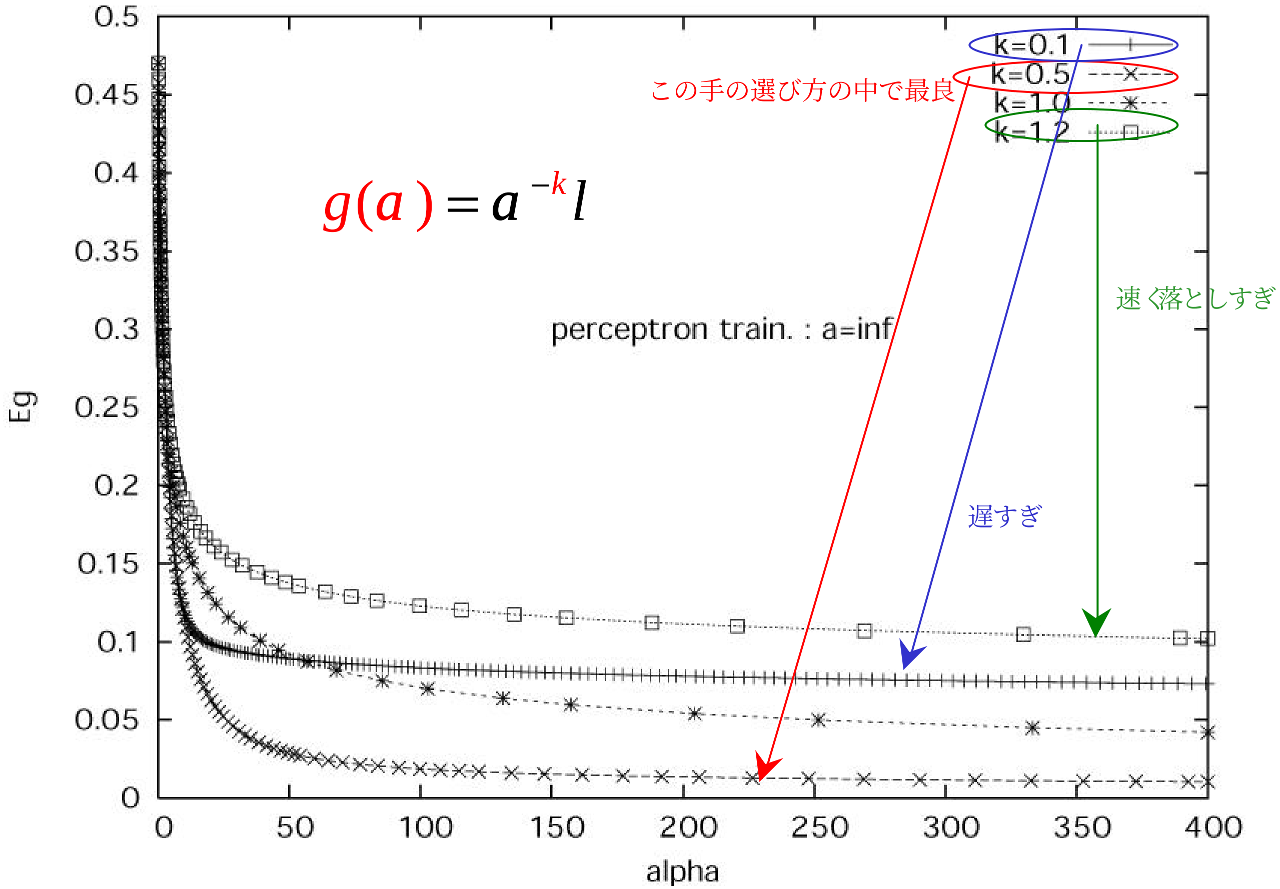
$$\frac{dl}{da} = \frac{1}{l}\left[\frac{g(a)^2 E_a(R)}{2} - g(a)F_a(R)l\right]$$

学習レートの例題数に関する「スケジューリング」を考える

$$g(a) = a^{-k}l$$

$\Rightarrow$ 様々なkの値に対してパフォーマンスを調べる

ad hoc な学習レートのパフォーマンス



システマティックに
最適な学習レートを
選ぶ方法はないものか？

学習レートの最適化

学習の各ステップで重なりの増加率が最大になるように学習レートを決定する

$$\frac{dR}{da} = \frac{1}{l^2}\left[-\frac{R}{2} g(a)^2 E_a(R) + g(a)\left[F_a(R)R - G_a(R)\right] l \right] \equiv \underline{L(g(a))}$$

学習レートに関する
汎関数を最大化

$$\boxed{g_{opt}(a) = \frac{\left[F_a(R)R - G_a(R)\right] l}{RE_a(R)}}$$

F_a, E_a, G_a, R, l

を介して例題数に依存する

最適な学習レートのスケジューリング

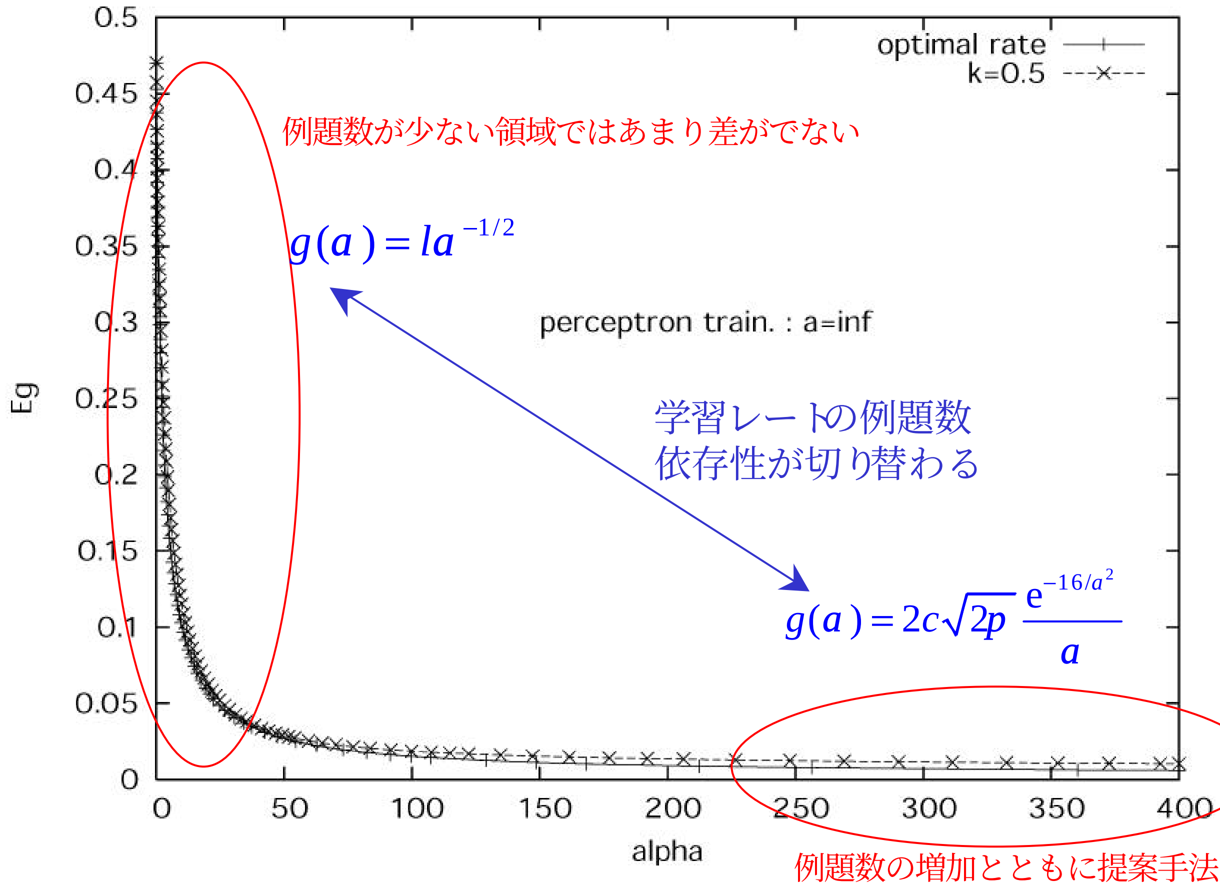
$$\boxed{\begin{aligned} &(1+R)^{-(1+A)/A}(1-R)^{(1-A)/A} R = cl \\ &A = 1 - 2\Delta = 1 - 2e^{-a^2/2} \end{aligned}}$$

⇒ ad hoc な学習レートの選び方の
最良値と比較する

マクロな量のフローが陽に書ける

最適化された学習曲線

学習可能な場合



例題数が少ない領域ではあまり差がでない

$g(a) = la^{-1/2}$

学習レートの例題数
依存性が切り替わる

$g(a) = 2c\sqrt{2p}\,\dfrac{e^{-16/a^2}}{a}$

例題数の増加とともに提案手法が優位になる

汎化誤差の漸近形

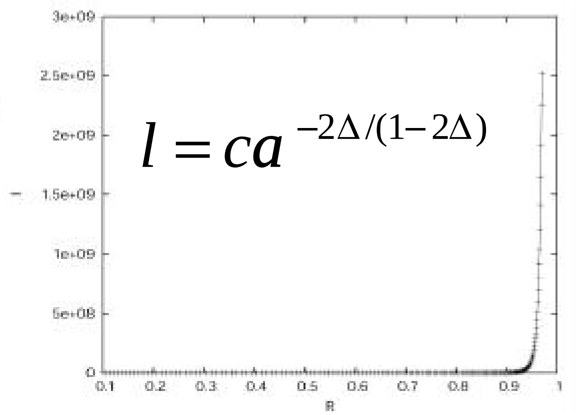
$$e_g = \frac{4}{p} a^{-1}$$

バッチ学習と同じ精度

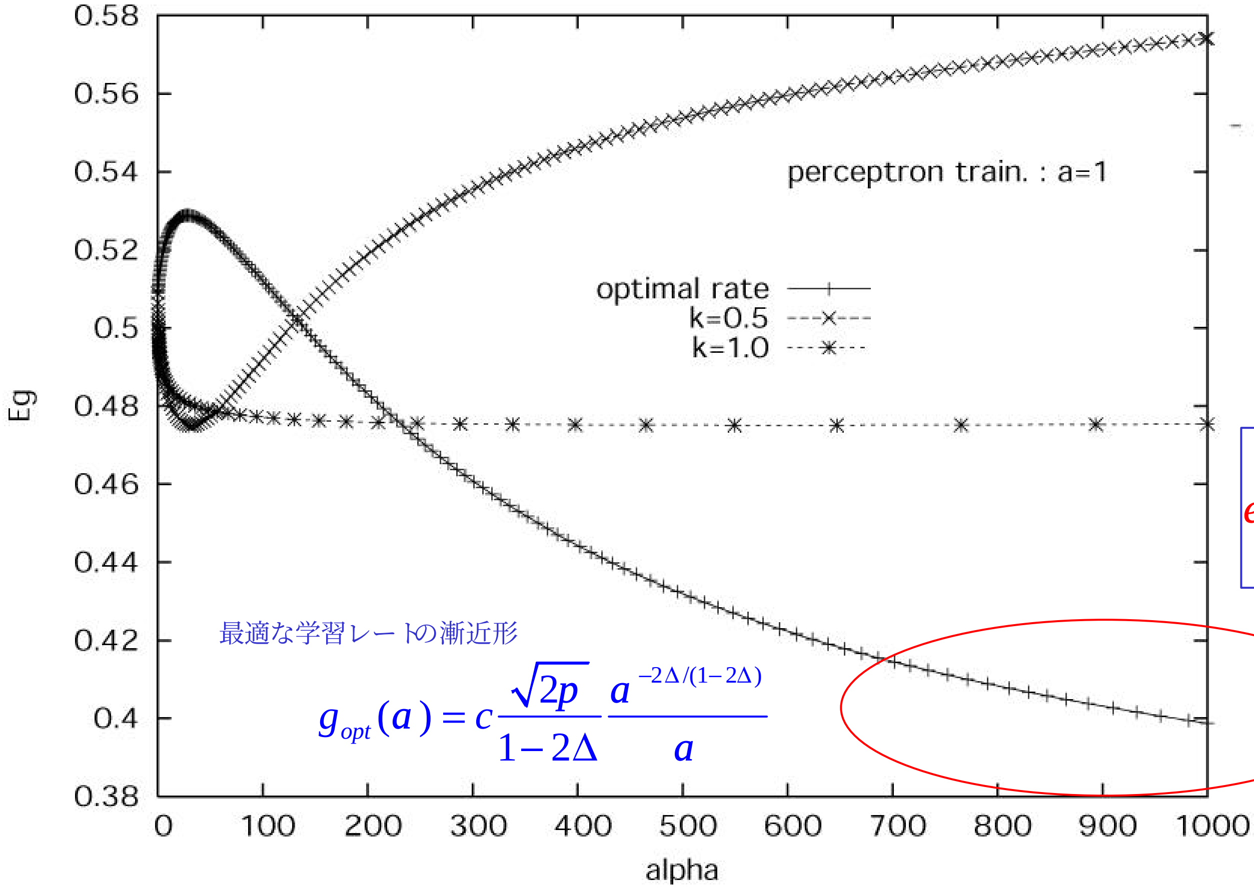
最適化された学習曲線

学習不可能な場合

結合の長さは学習とともに発散



$$l = ca^{-2\Delta/(1-2\Delta)}$$



最適な学習レートの漸近形

$$g_{opt}(a) = c\frac{\sqrt{2p}}{1-2\Delta}\frac{a^{-2\Delta/(1-2\Delta)}}{a}$$

学習曲線の漸近形

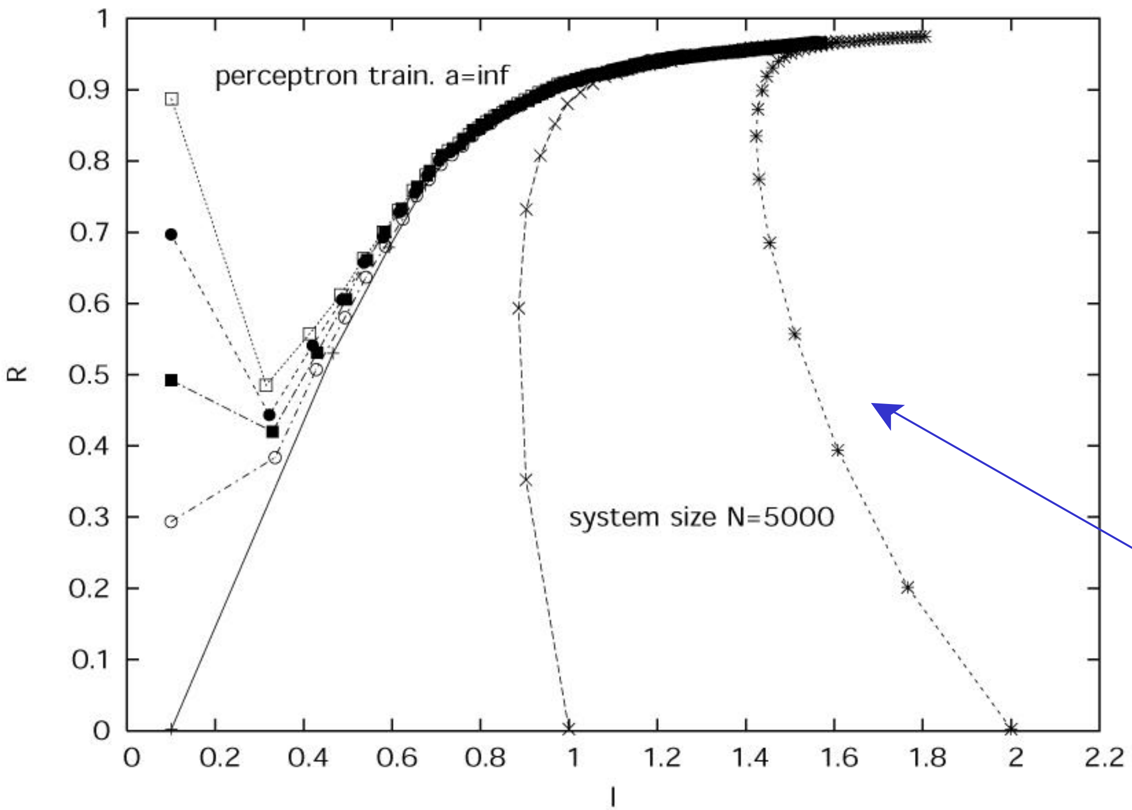
$$e_g = \frac{\sqrt{2}}{p}\frac{\sqrt{2H(a)}}{1-2\Delta}\frac{1}{\sqrt{a}} + 2H(a)$$

オンライン学習の計算機シミュレーション

オンライン・パーセプトロン学習に関するサンプルプログラム

http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/KONTON2004/cnline.c

各自がダウンロードし、必要と
あらば修正を加えて使ってください



課題13

次式の意味で最適化された
オンラインHebb学習をシミュレートせよ

$$\mathbf{J}^{\mathbf{m+1}} = \mathbf{J}^{\mathbf{m}} + g(a)T_a(v)\mathbf{x}$$

$$g(a) = la^{-k}$$

上掲サンプル・プログラムの
出力結果の一例