



HOKKAIDO UNIVERSITY

Title	2004年度 混沌系工学特論講義ノート
Author(s)	井上, 純一; Inoue, Jun-ichi
Description	当講義資料は著者のホームページ http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/ からダウンロードできます。 http://chaosweb.complex.eng.hokudai.ac.jp/~j_inoue/
Issue Date	2004
Doc URL	https://hdl.handle.net/2115/370
Rights(URL)	https://creativecommons.org/licenses/by-nc-sa/2.1/jp/
Type	learning object
File Information	konton2004_8_present.pdf, 第8回講義スライド





混沌系工学特論 #8

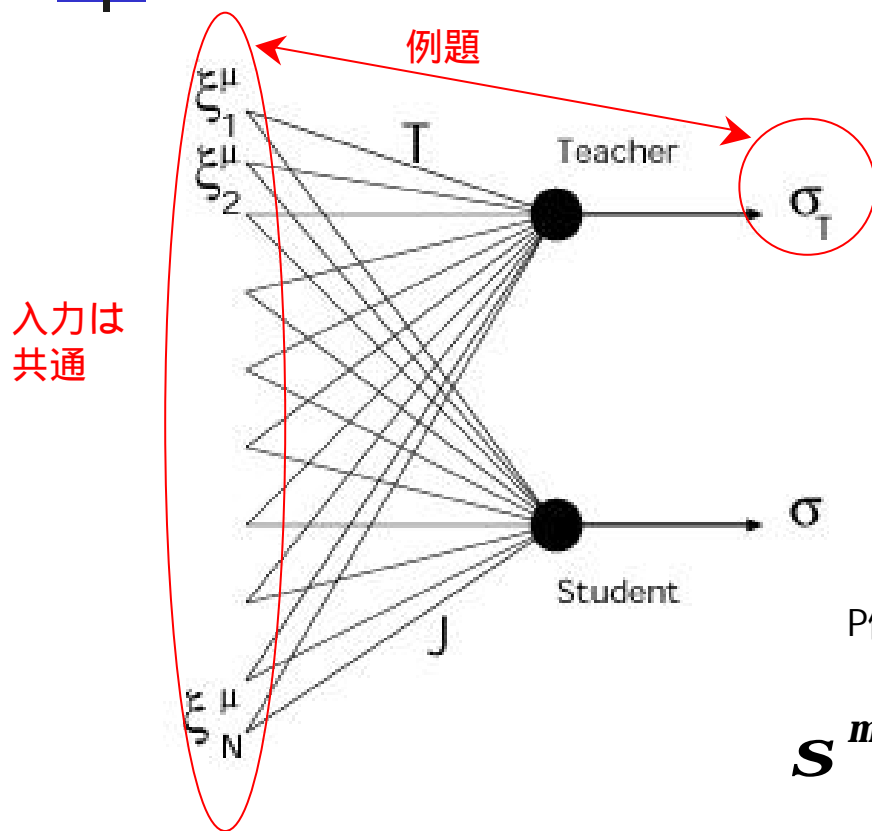
情報科学研究科 井上純一

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

平成17年1月24日 第9回講義

例題からの学習

前回の復習



$$\mathbf{s}_T^m = \text{sgn}(\mathbf{T} \cdot \mathbf{x}^m) = \text{sgn}\left(\sum_{i=1}^N T_i \mathbf{x}_i^m\right)$$

$$\mathbf{s}^m = \text{sgn}(\mathbf{J} \cdot \mathbf{x}^m) = \text{sgn}\left(\sum_{i=1}^N J_i \mathbf{x}_i^m\right)$$

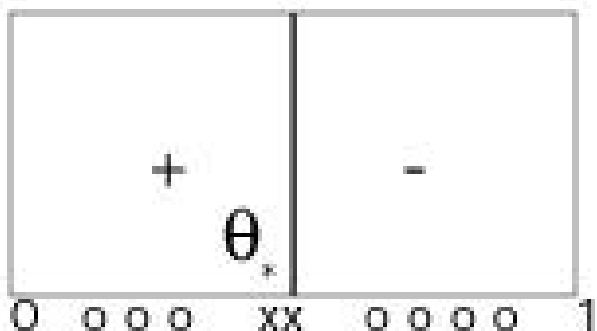
P個の例題全てに正しい答えをだす結合を選べれば

$$\mathbf{s}^m = \mathbf{s}_T^m \quad (m=1, \dots, P)$$

例題以外の入出力を正しく与える保障は汎化誤差によりわかる

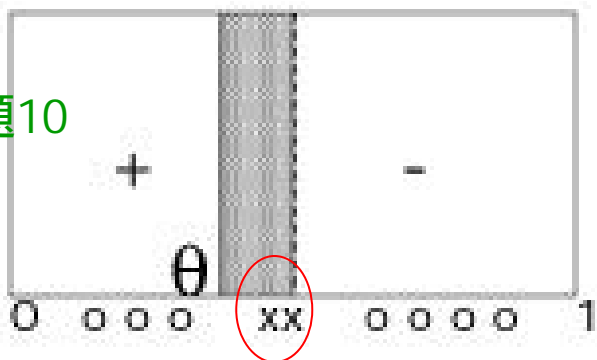
訓練誤差と汎化誤差

$$T \quad \mathbf{s}_T(t) = \text{sgn}[\mathbf{q}_* - x(t)]$$



$$S \quad \mathbf{s}_S(t) = \text{sgn}[\mathbf{q}(t) - x(t)]$$

問題10

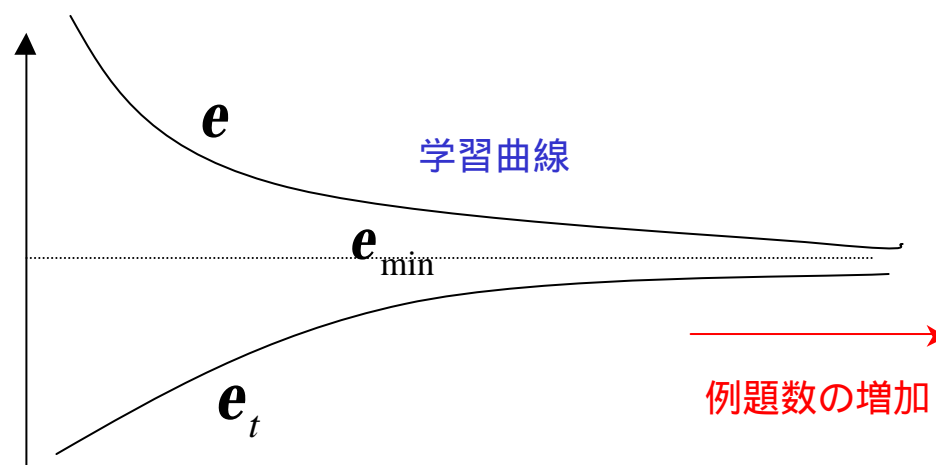


この領域に落ちた入力に対して
間違った結果を出力する

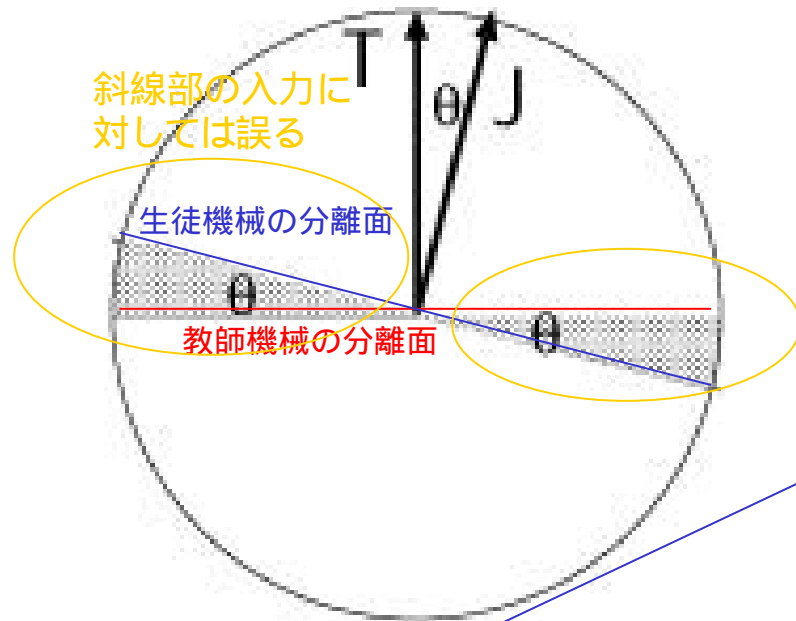
印の例題に対しては正解するが、×印は誤る

$$\text{訓練誤差} : e_t = \frac{1}{P} \sum_{m=1}^P \Theta(-\mathbf{s}_T \mathbf{s})$$

$$\text{汎化誤差} : e = \sum_{\{\mathbf{S}\}} P_S(\mathbf{S}) \Theta(-\mathbf{s}_T \mathbf{s})$$

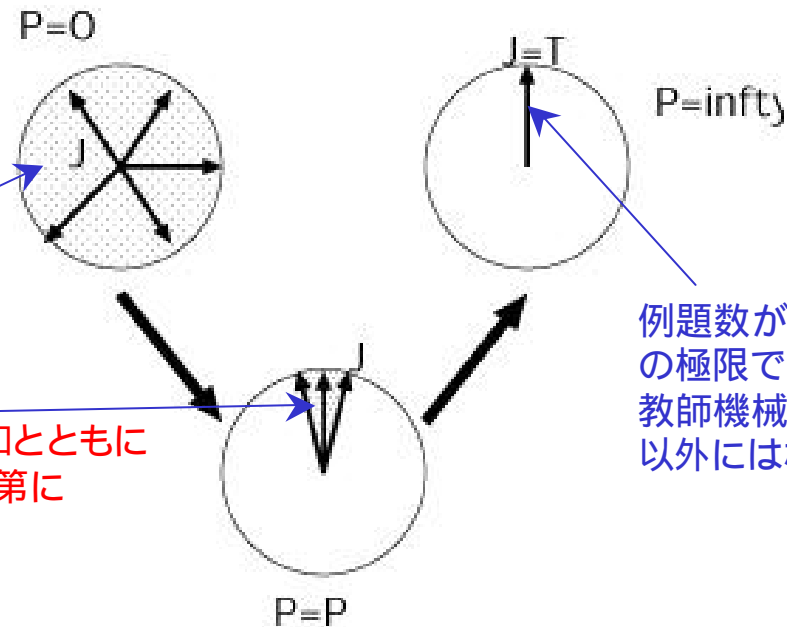


解空間の体積の例題数依存性



$$R = \frac{1}{N} (\mathbf{J} \cdot \mathbf{T}) = \cos q$$

$$\text{汎化誤差は } \mathbf{e} = \frac{q}{p} = \frac{1}{p} \arccos(R)$$



得られた解空間の中で
どの結合を選ぶかは学習則に
依存する
ランダムに選べばギブス学習

例題数の増加とともに
解空間は次第に
縮んでいく

例題数が無限大
の極限での解は
教師機械の結合
以外にはない

解空間のエントロピーの評価

$$P = \mathbf{a}N$$

P個の例題数に対する解空間の体積

$$\Omega_P(\mathbf{e}) = \Omega_{P-1}(\mathbf{e})(1-\mathbf{e}) = \dots = \Omega_0(\mathbf{e})(1-\mathbf{e})^P$$

$$\Omega_0(\mathbf{e})(1-\mathbf{e})^P = (1-\mathbf{e})^P \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} d\mathbf{J} d(\mathbf{J}^2 - N) d\left(\frac{1}{N}(\mathbf{J} \cdot \mathbf{T}) - \cos(\mathbf{p}\mathbf{e})\right)$$

$$= \exp(Ns(\mathbf{e}, \mathbf{a}))$$

(導出の詳細は講義ノート参照)

$$s(\mathbf{e}, \mathbf{a}) = \frac{1}{2}(1 + \log 2\mathbf{p}) + \frac{1}{2} \log \sin^2(\mathbf{p}\mathbf{e}) + \mathbf{a} \log(1-\mathbf{e})$$

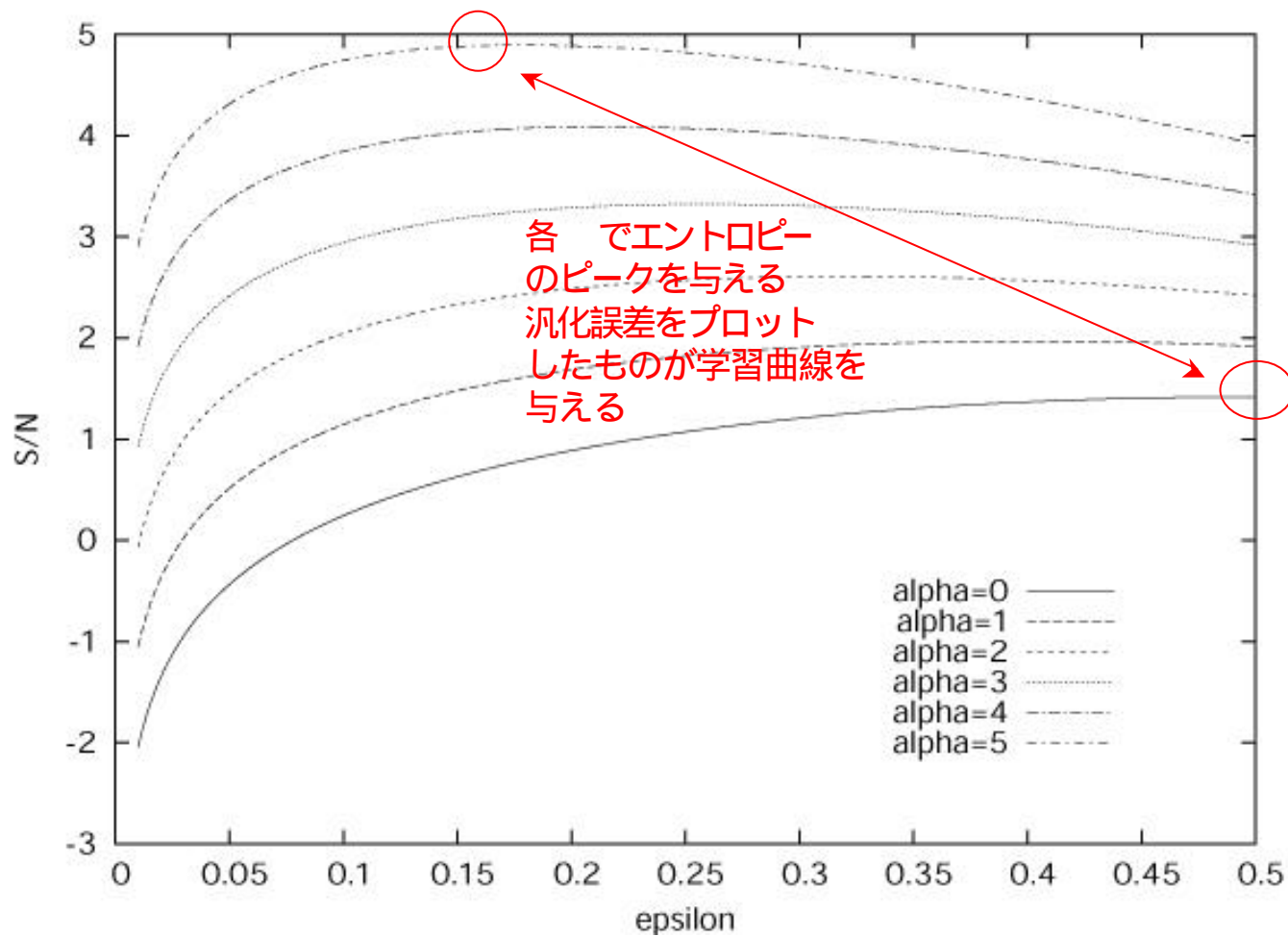
解空間のエントロピー

$$\mathbf{e}(\mathbf{a}) = \arg \max_{\mathbf{e}} s(\mathbf{e}, \mathbf{a})$$

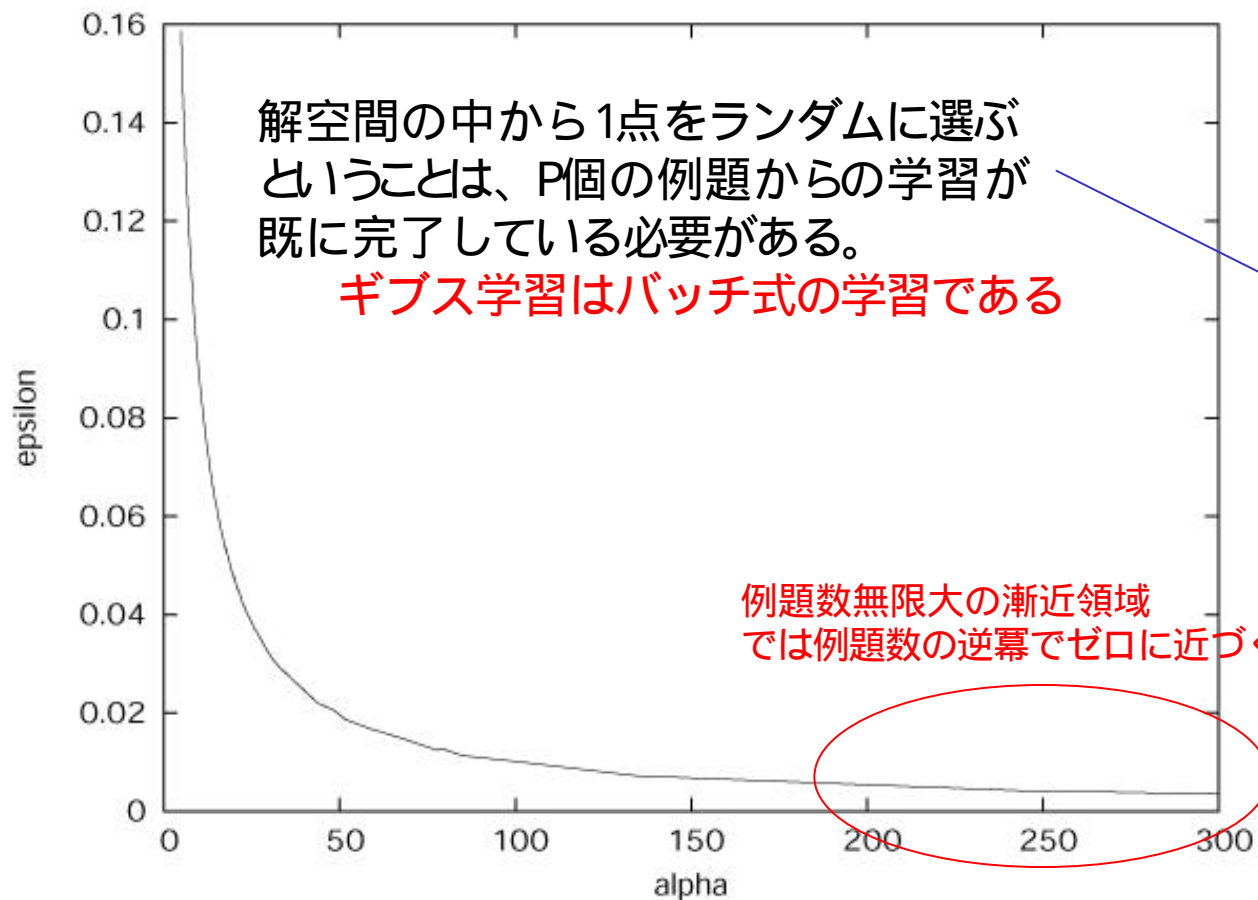
学習曲線

本来は訓練データに関する平均操作をしなければならない。ここでの解析はその部分を「粗視化」している。正確にはレプリカ法を用いるべきだが、この講義ではそれに触れない

解空間のエントロピー



ギブス学習の学習曲線

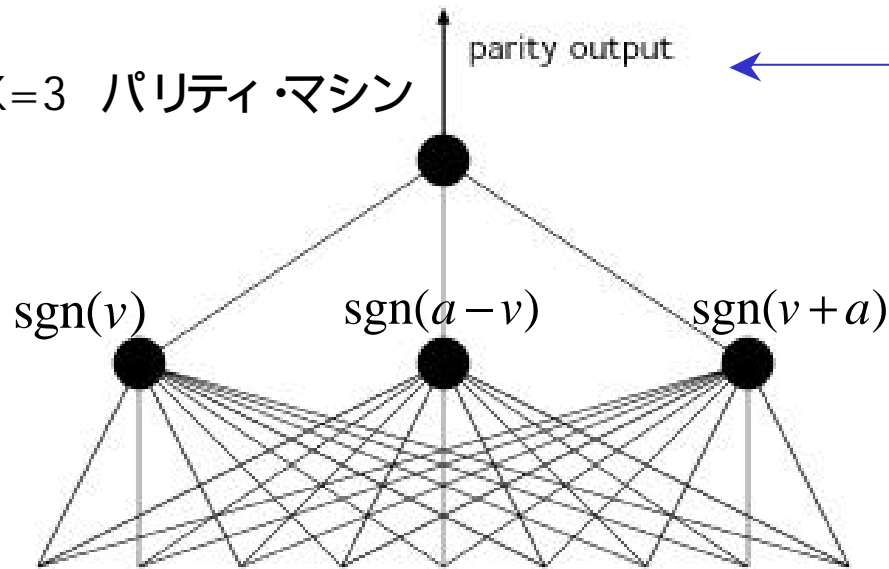


処理時間の少ない
オンライン式の
学習を考えると
どうなるか?

実現不可能な規則のオンライン学習

教師機械

K=3 パリティ・マシン



$$T_a(v) = \text{sgn}[v(a-v)(a+v)]$$

$$v = \sqrt{N} (\mathbf{J}^0 \cdot \mathbf{x}) / |\mathbf{J}^0|$$

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)

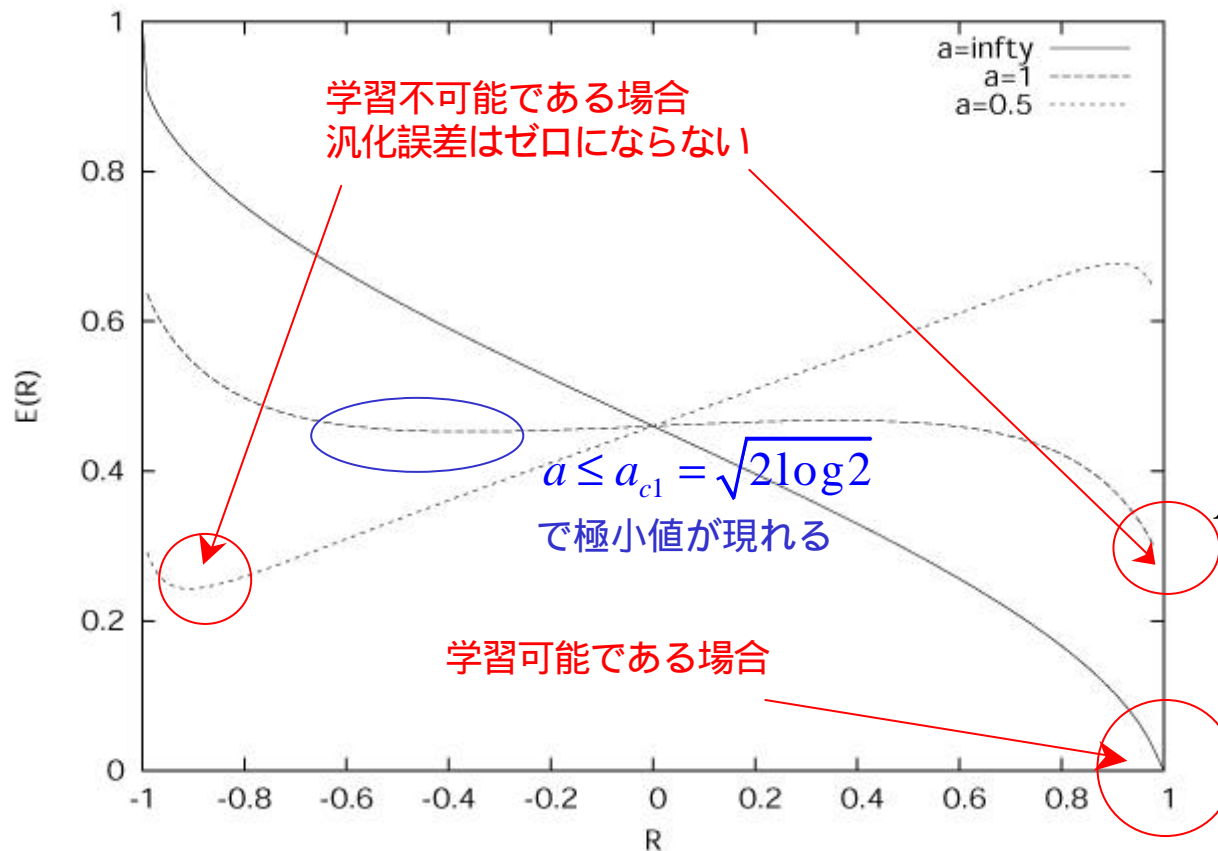
(単純パーセプトロンは $2N$)

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

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

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

マクロな量の導入と汎化誤差



両機械の結合の重なり

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

生徒機械の結合の長さ

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

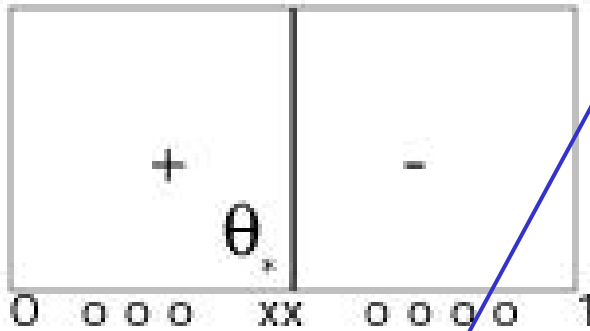
汎化誤差

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

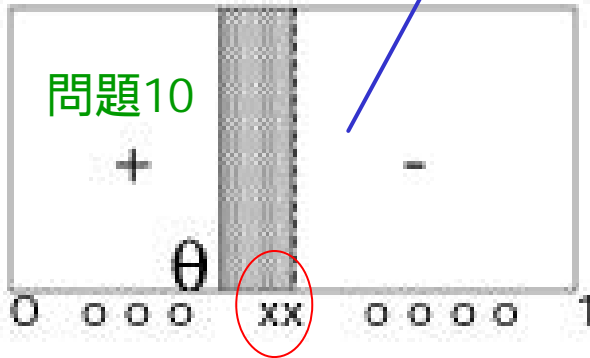
汎化誤差の例題数の増加にともなう振る舞いは具体的に学習則を与えることにより明らかとなる

パーセプトロン学習とそのダイナミクス

T $s_T(t) = \text{sgn}[\mathbf{q}_* - x(t)]$



S $s_S(t) = \text{sgn}[\mathbf{q}(t) - x(t)]$



この領域に落ちた入力に対して
間違っただけの結果を出力する

これの高次元版を考える

パーセプトロン学習

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

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 $P = \mathbf{a}N$

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

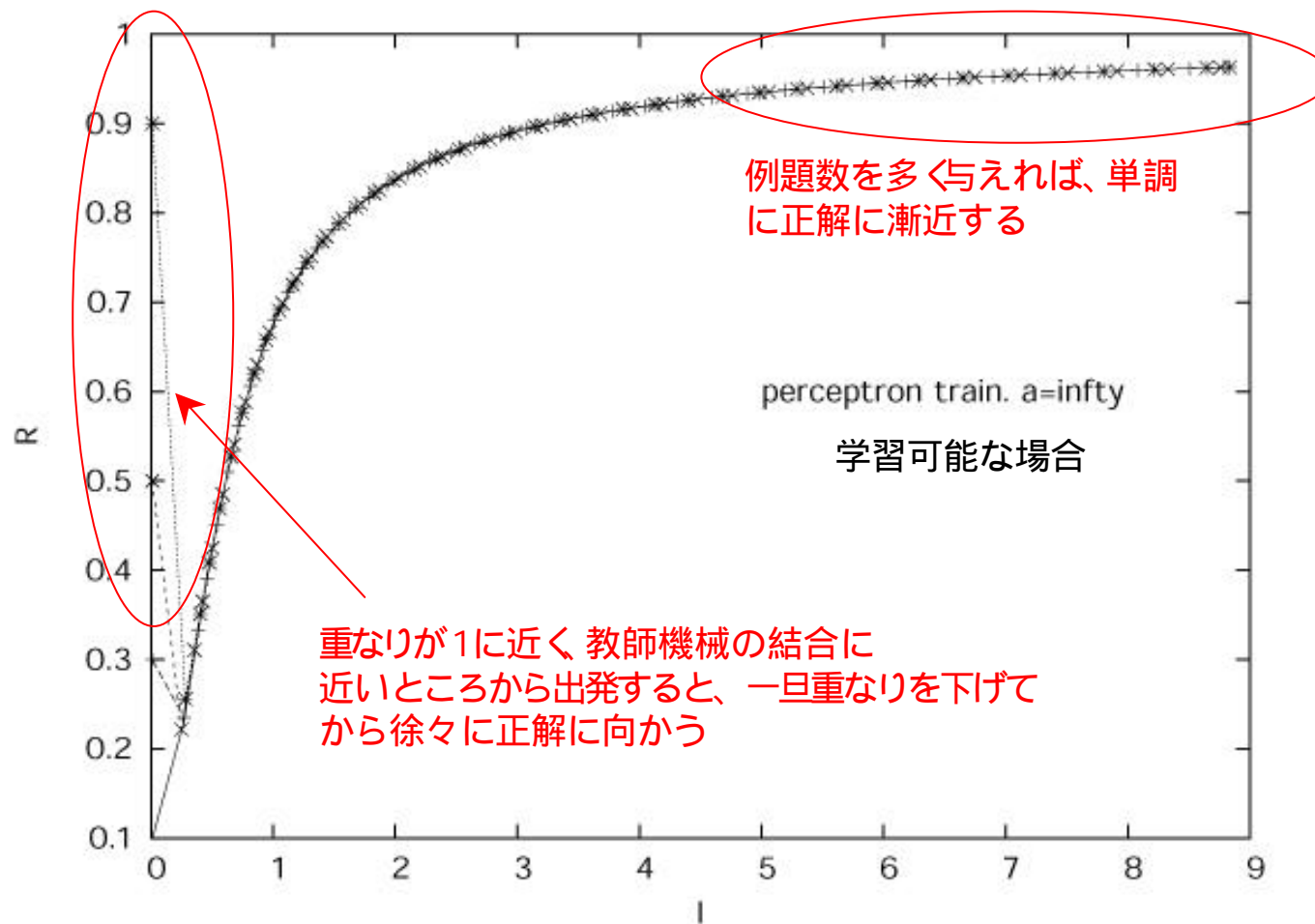
$$l^2 \frac{dR}{d\mathbf{a}} = -\frac{R}{2} \langle \Theta(-T_a(v)S(u))v \rangle + l(R \langle -\Theta(T_a(v)S(u))u \rangle$$

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

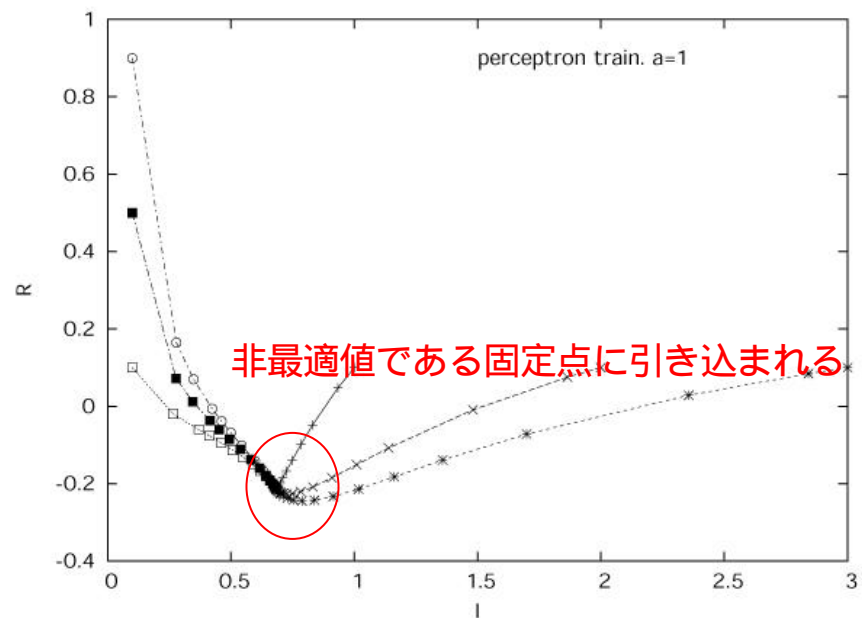
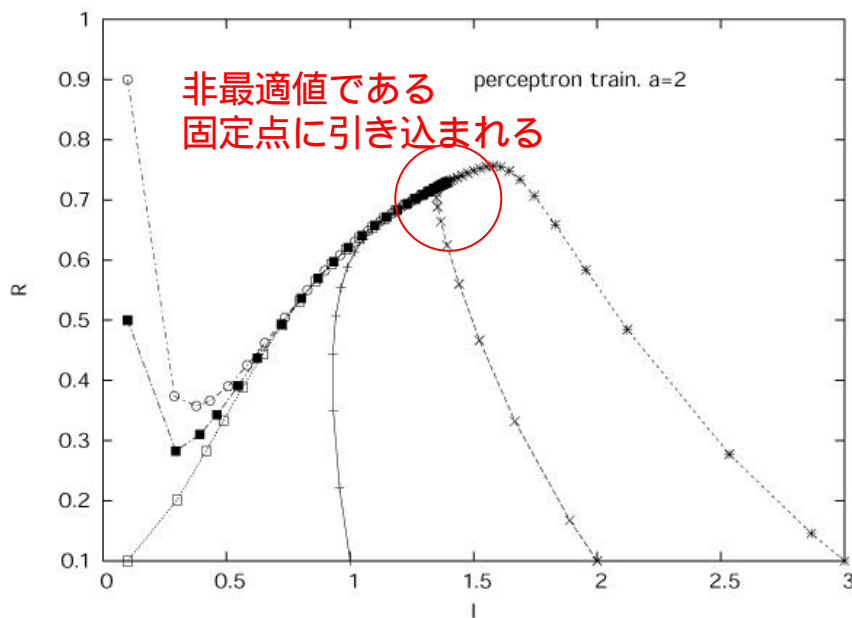
状態更新式

導出の詳細は講義ノート参照

マクロな量のフロー図#1



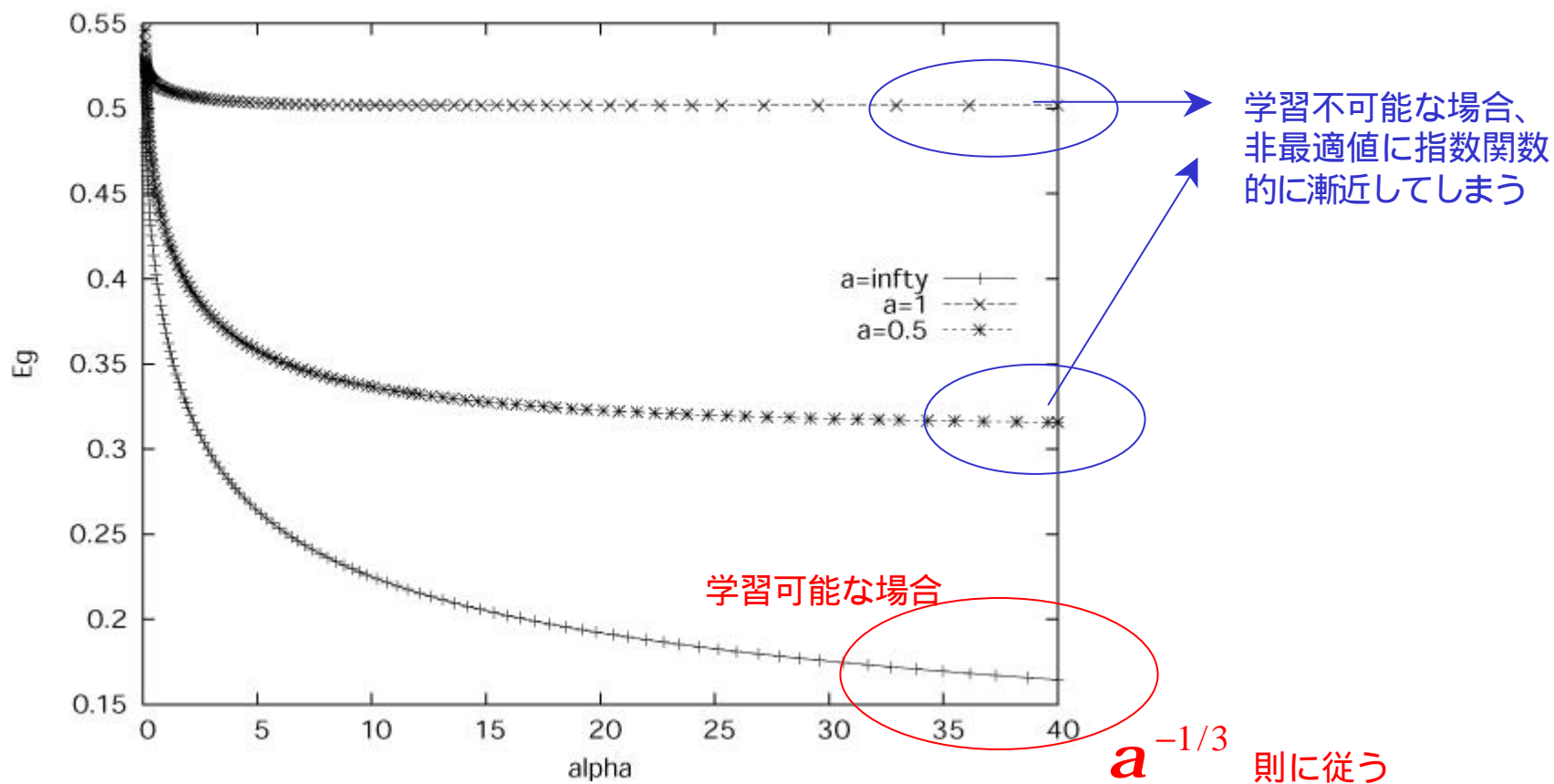
マクロな量のフロー図#2



汎化誤差は

$$E_a(R) = \langle \Theta(-T_a(v)S(u)) \rangle \quad \text{で算出される}$$

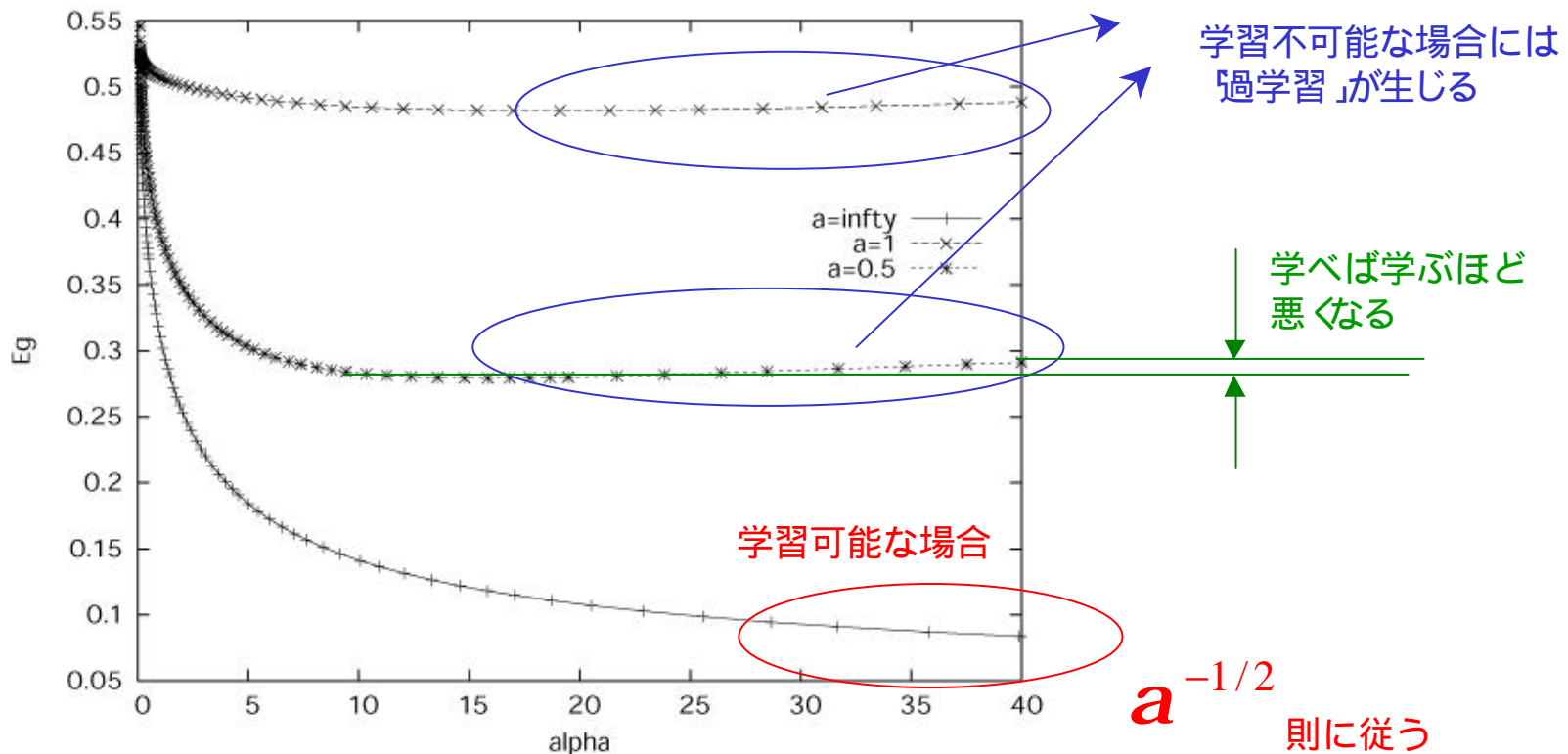
パーセプトロン学習の学習曲線



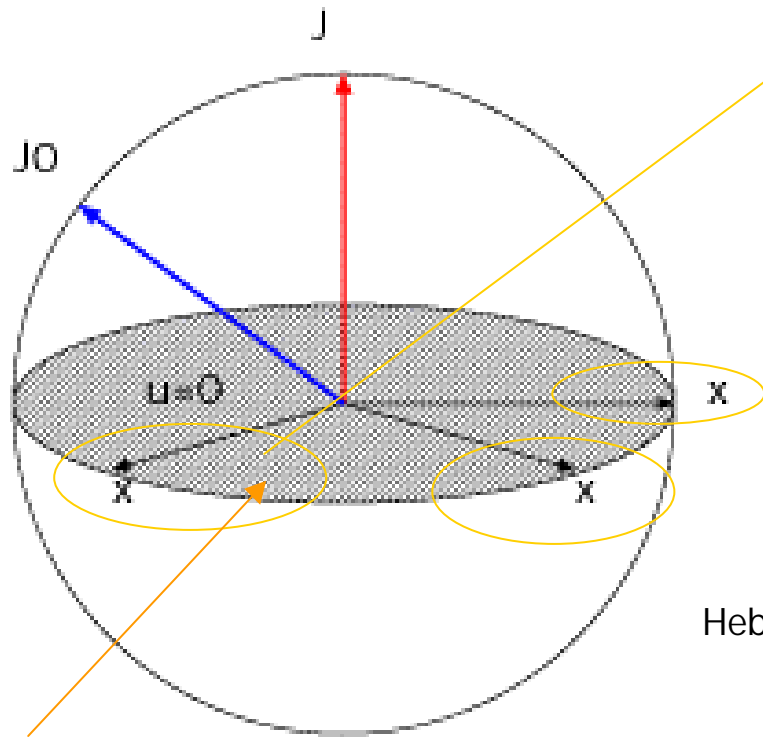
Hebb学習のダイナミクスと過学習

$$\mathbf{J}^{m+1} = \mathbf{J}^m + T_a(v)\mathbf{X} \quad \text{Hebb学習}$$

教師の出力に応じて入力ベクトルを結合に足しこむ



質問付き学習



入力ベクトルを拾ってくる空間に制限を加える

パーセプトロンの場合、その分離面：

$$u = \mathbf{J} \cdot \mathbf{x} = 0$$

から取ってくるのが効果的

$$P_R(v | u = 0) = \sqrt{2pd}(u)P_R(u, v)$$

Hebb学習に適用すると

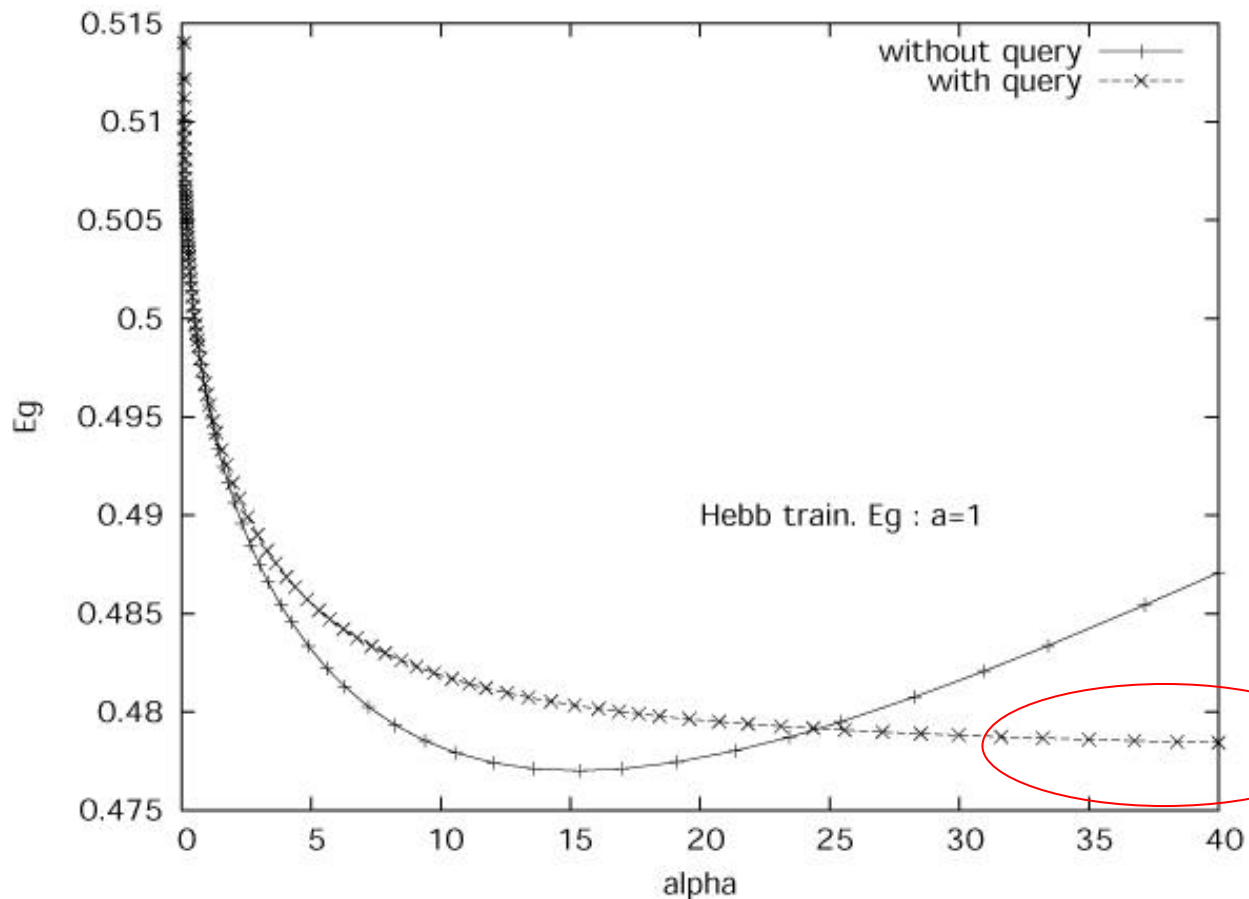
状態更新式

$$2l \frac{dl}{da} = \langle 1 \rangle + 2l \langle T_a(v)u \rangle$$

$$l \frac{dR}{da} + R \frac{dl}{da} = \langle T_a(v)v \rangle$$

生徒機械にとって分離面上に落ちた例題に対する判断が最も難しいので、率先してこの部分から例題を要求することにより、学習を進める

質問の効果による過学習の消失



質問の効果により
過学習が消失している