混沌系工学特論 #8

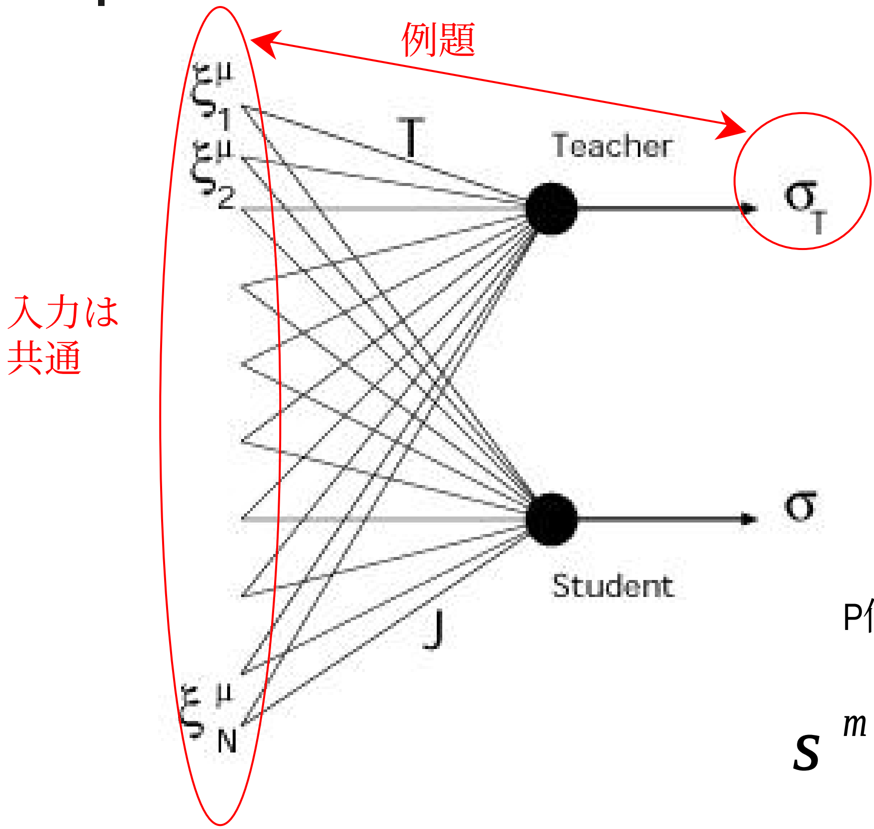
情報科学研究科　井上純一

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



$$s_T^m = \mathrm{sgn}\left(\mathbf{T}\cdot x^m\right) = \mathrm{sgn}\left(\sum_{i=1}^{N} T_i x_i^m\right)$$

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

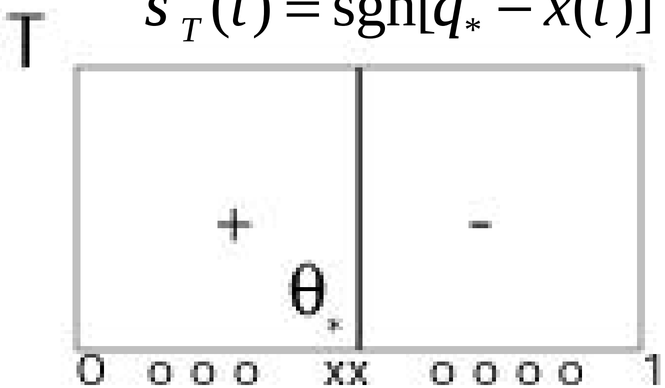
P個の例題全てに正しい答えをだす結合を選べれば

$$s^m = s_T^m \qquad (m = 1, \ldots, P)$$

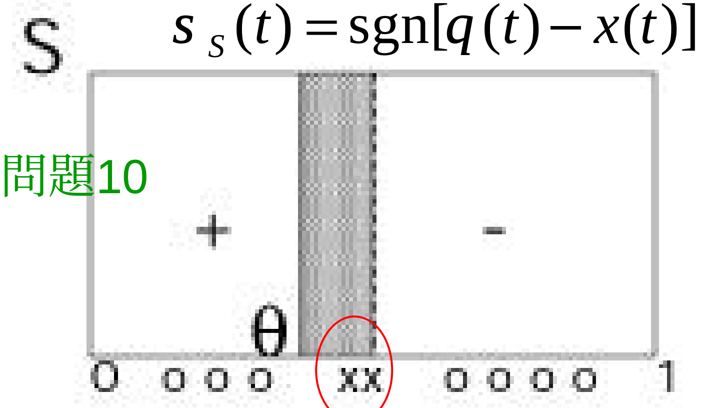
⇒ 例題以外の入出力を正しく与える保障は汎化誤差によりわかる

訓練誤差と汎化誤差

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



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



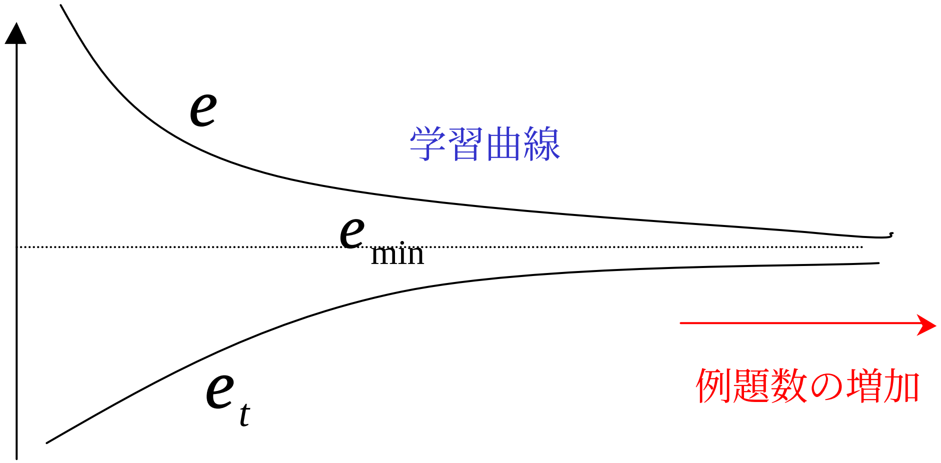
問題10

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

訓練誤差　：　$e_t = \dfrac{1}{P}\sum_{m=1}^{P} \Theta(-s_T s)$

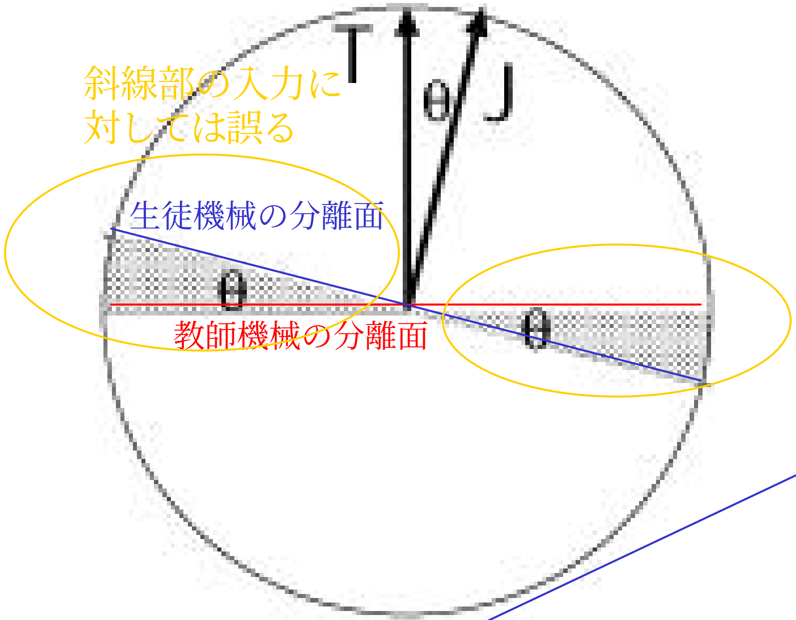
汎化誤差　：　$e = \sum_{\{S\}} P_S(\mathbf{S})\Theta(-s_T s)$

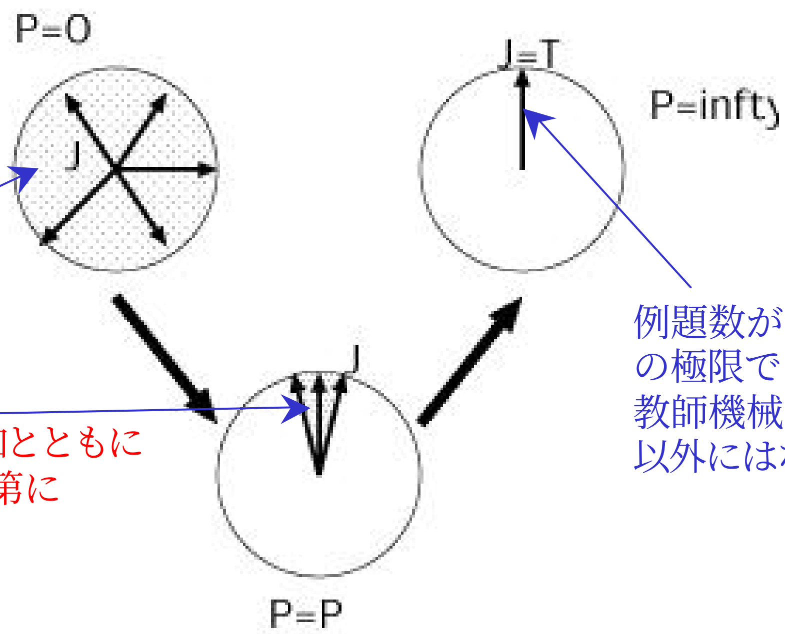


解空間の体積の例題数依存性

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

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





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

$$P = a\,N$$

P個の例題数に対する解空間の体積

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

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

$$= \exp\left(N s(e,a)\right)$$

(導出の詳細は講義ノート参照)

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

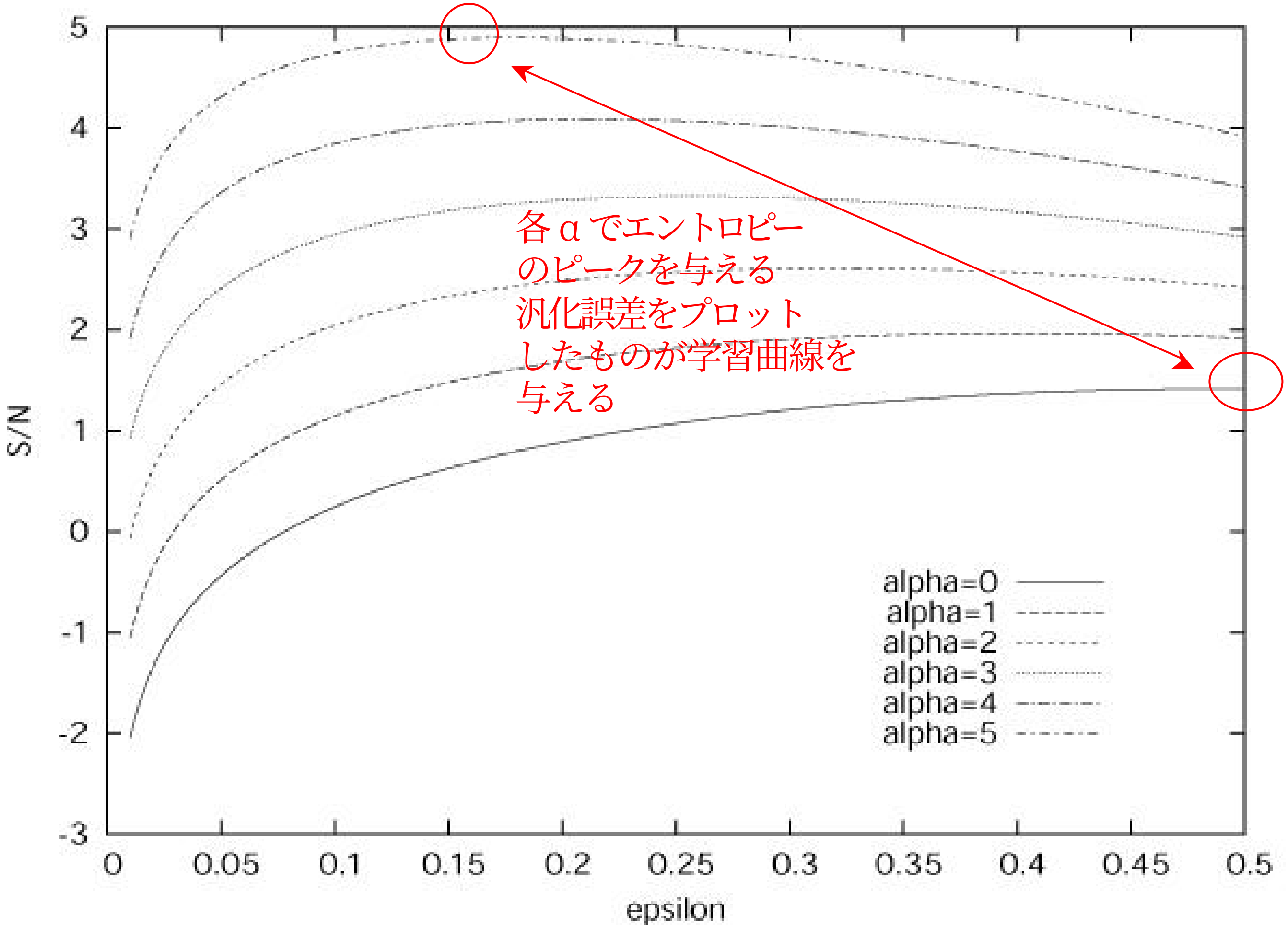
解空間のエントロピー

※　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

$$\boxed{e(a) = \arg\max_e\, s(e,a)}$$

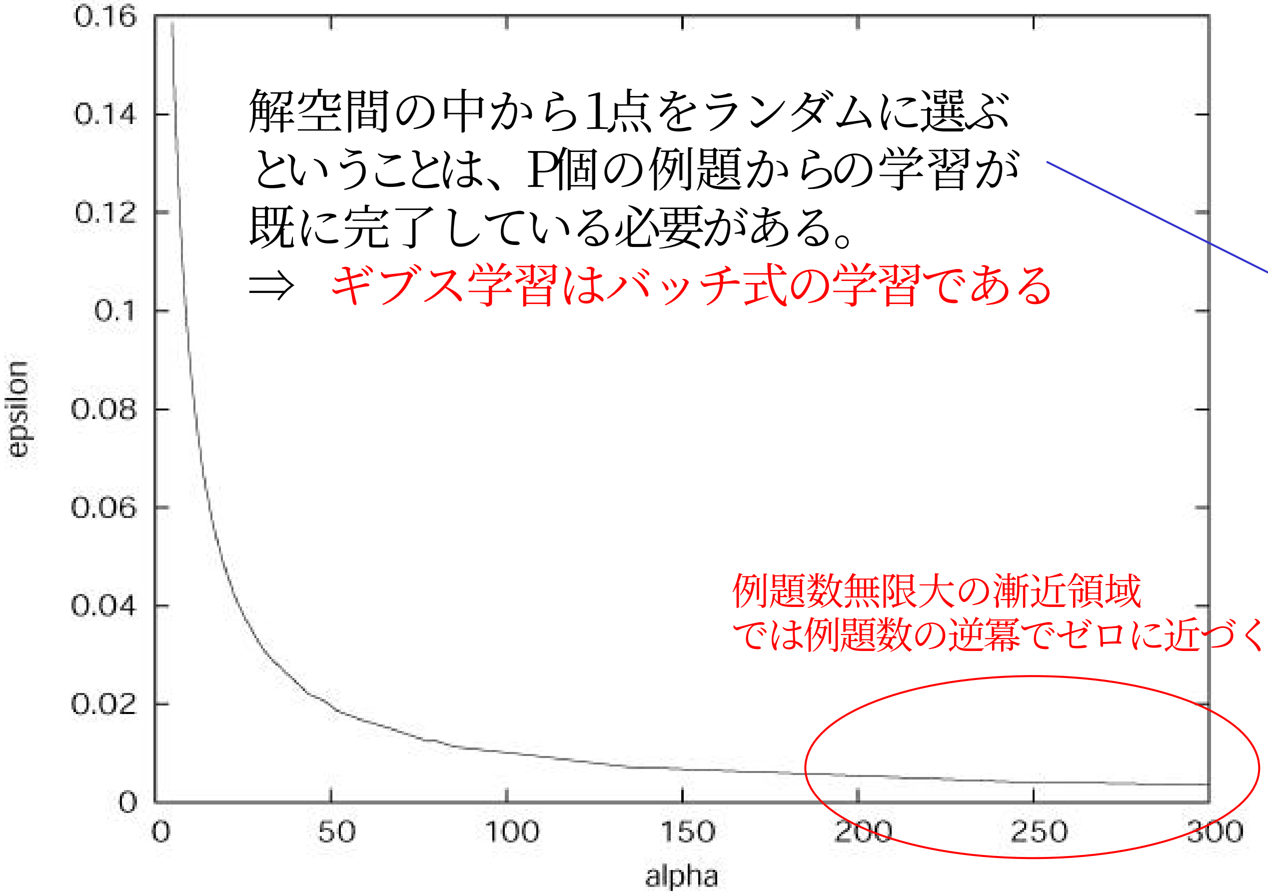
学習曲線

解空間のエントロピー



各αでエントロピー
のピークを与える
汎化誤差をプロット
したものが学習曲線を
与える
S/N
epsilon
5
4
3
2
1
0
-1
-2
-3
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
0.45
0.5
alpha=0
alpha=1
alpha=2
alpha=3
alpha=4
alpha=5

ギブス学習の学習曲線



解空間の中から1点をランダムに選ぶ
ということは、P個の例題からの学習が
既に完了している必要がある。
⇒　ギブス学習はバッチ式の学習である

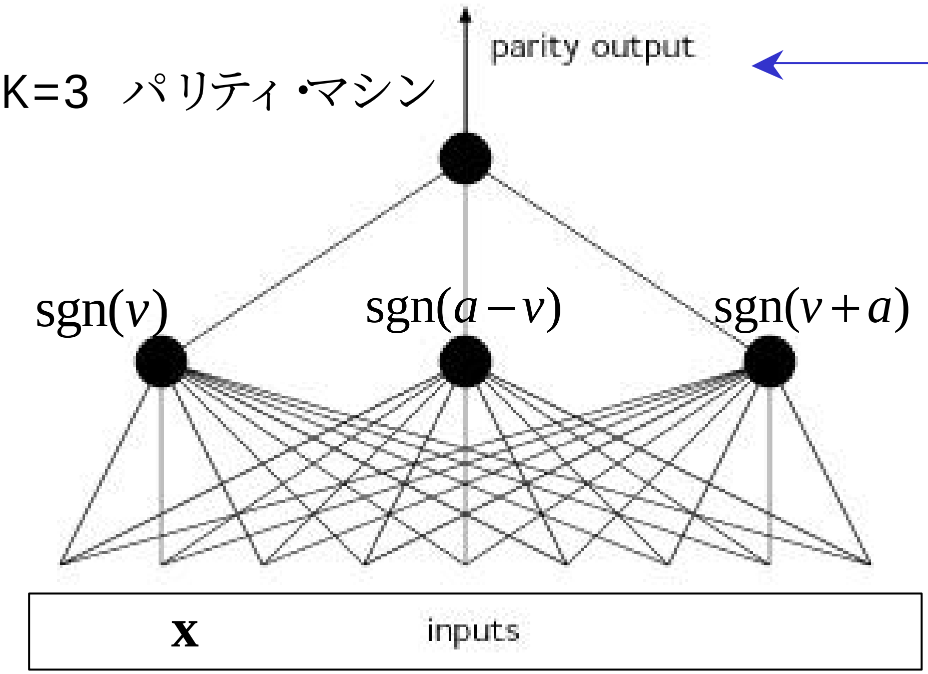
例題数無限大の漸近領域
では例題数の逆冪でゼロに近づく

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



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

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

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

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

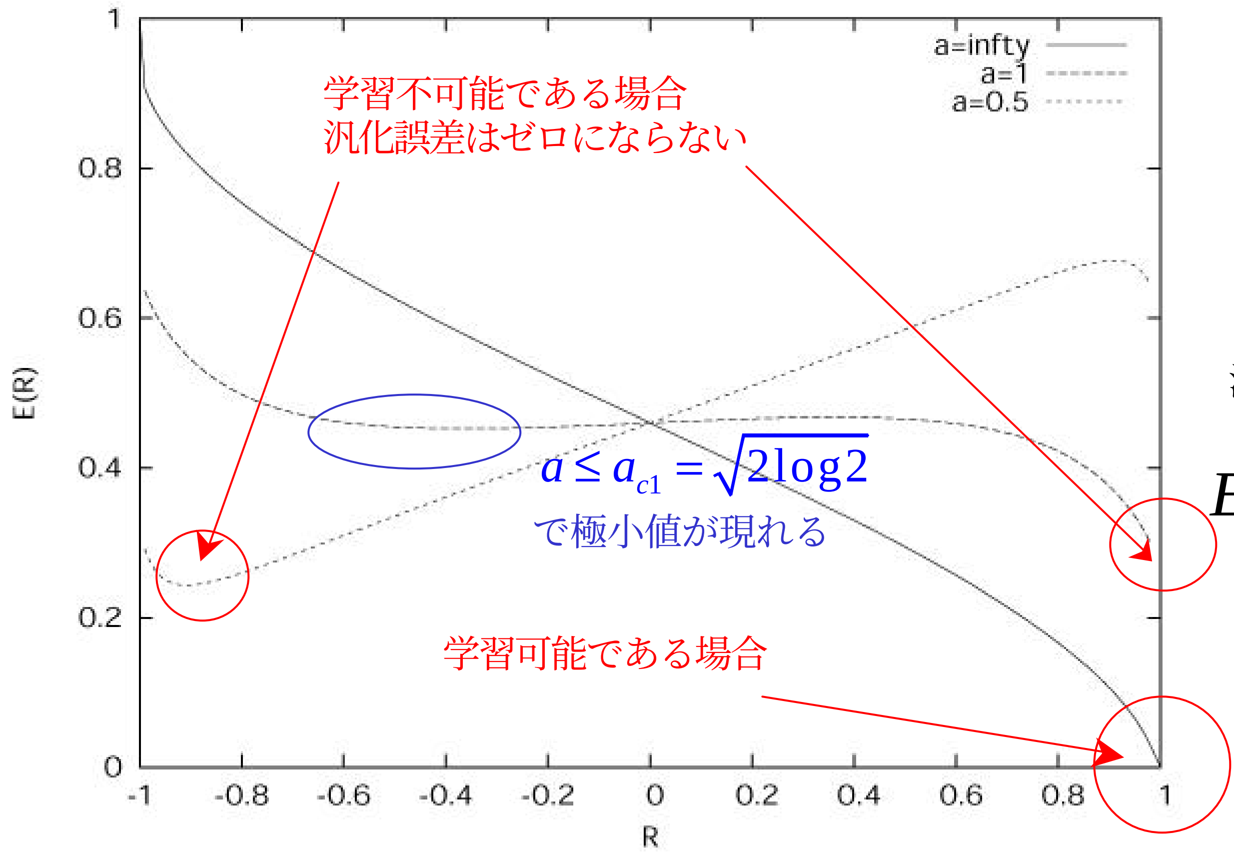
性能　：実現することのできる
入出力関係の数は

$\sim 10N$ （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

マクロな量の導入と汎化誤差



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

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

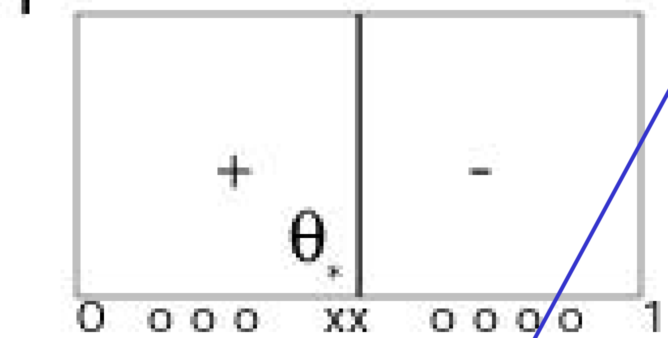
汎化誤差

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

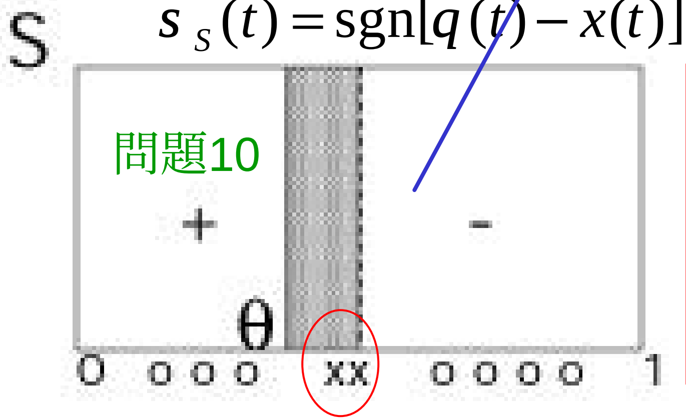
汎化誤差の例題数の増加に
ともなう振る舞いは具体的に
学習則を与えることにより明らか
となる

パーセプトロン学習とそのダイナミックス

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



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



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　$P = a\,N$

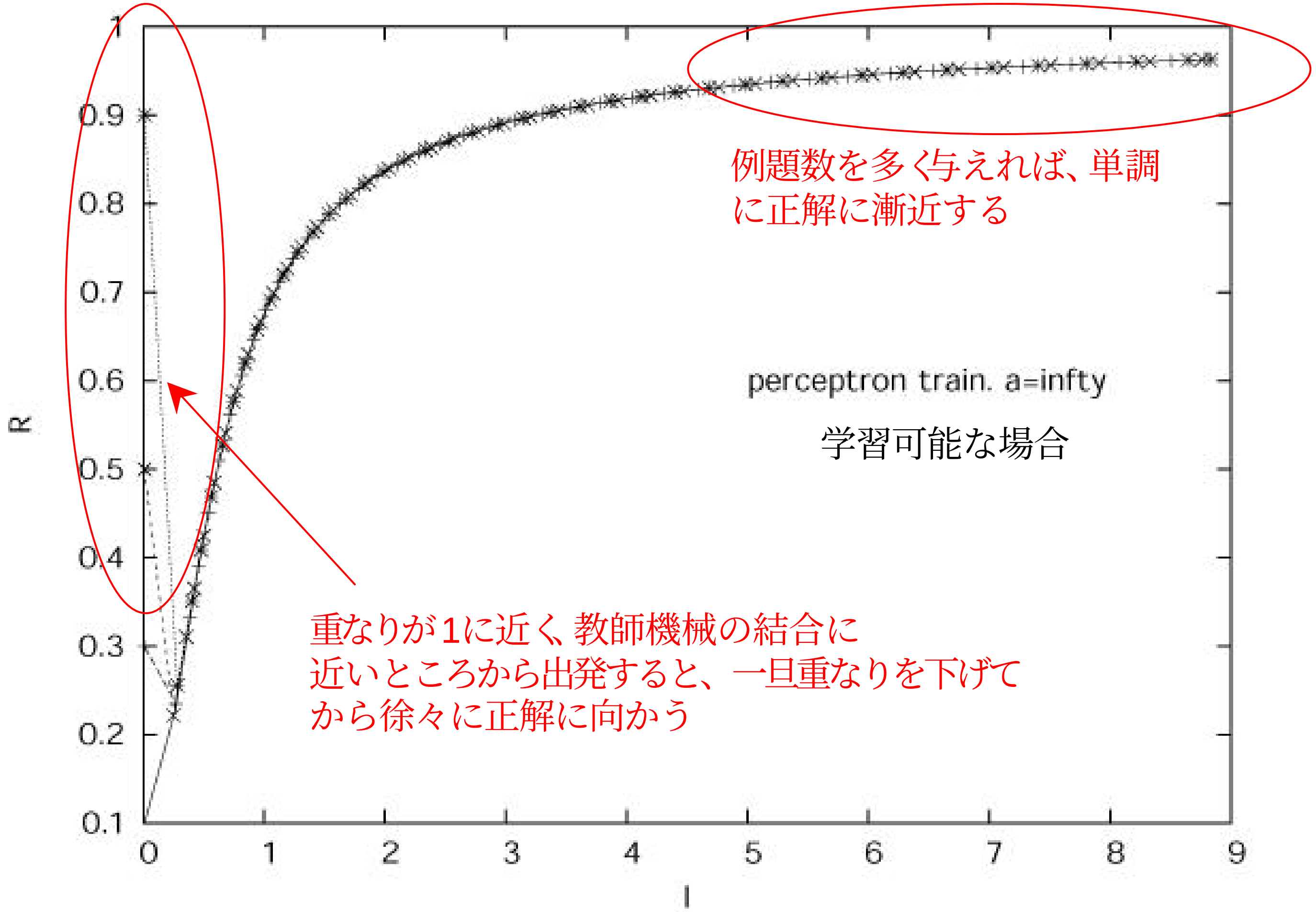
$$2l\frac{dl}{da} = -2l\langle\Theta(-T_a(v)S(u))u\rangle + \langle\Theta(-T_a(v)S(u))\rangle$$

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

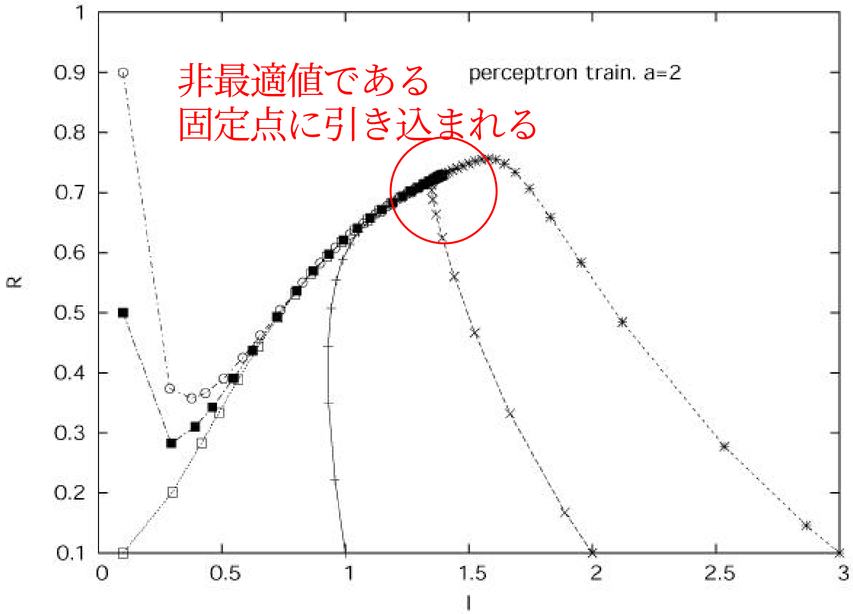
状態更新式

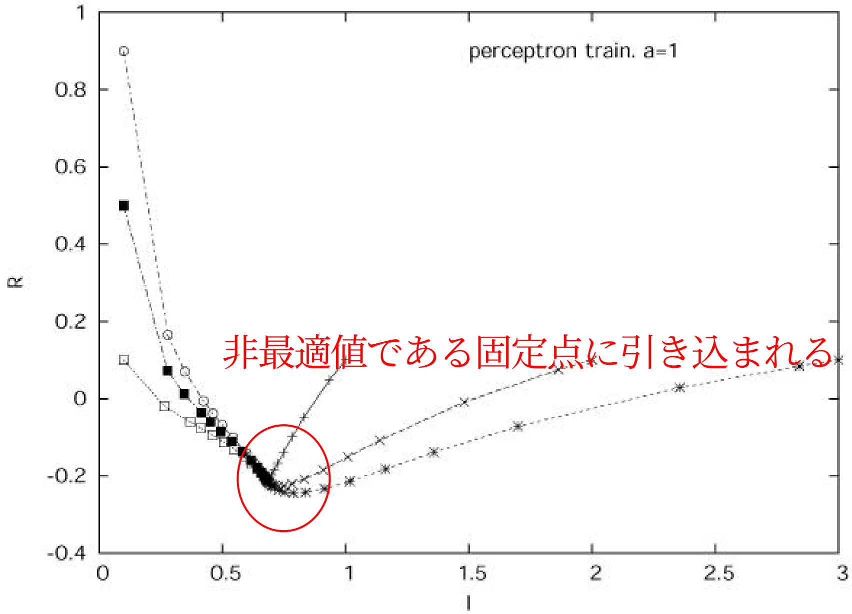
導出の詳細は講義ノート参照

マクロな量のフロー図#1



マクロな量のフロー図#2

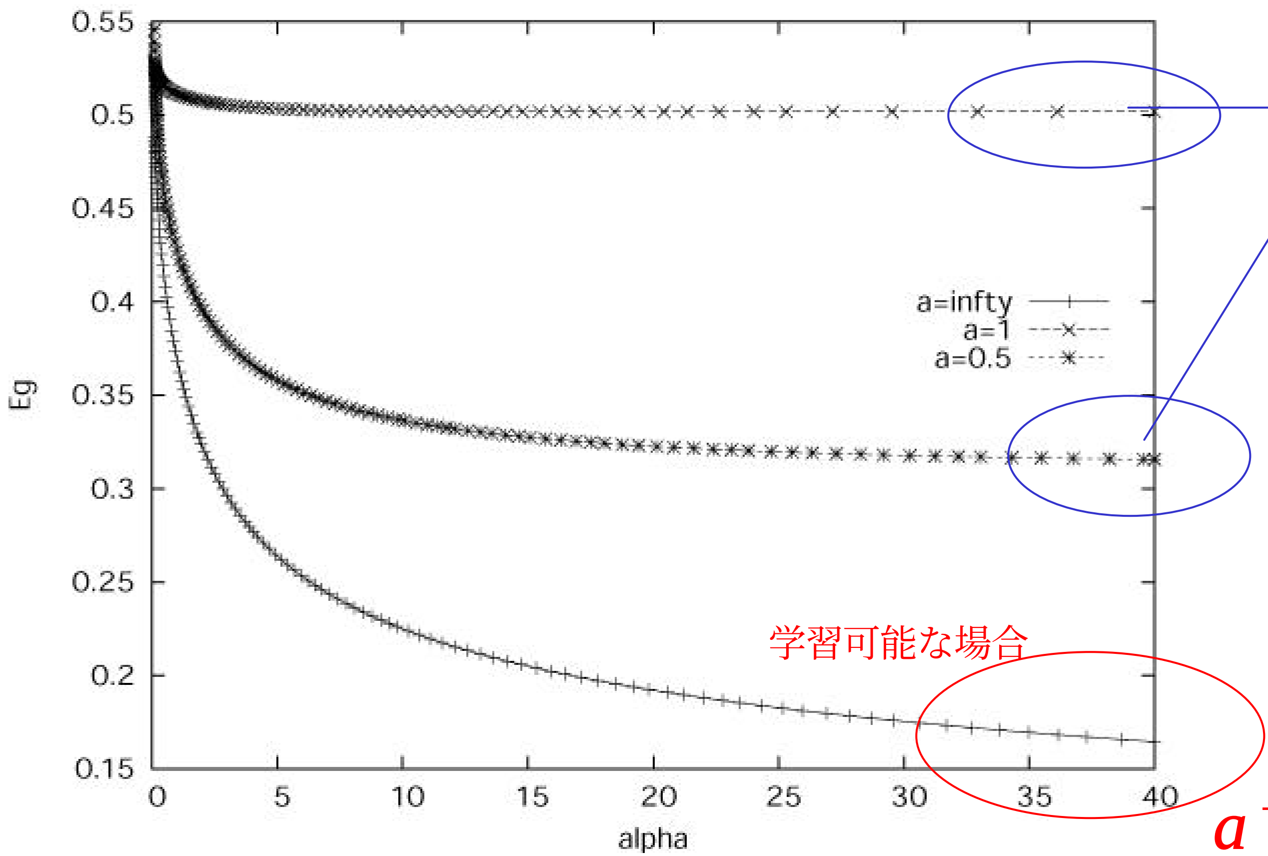




汎化誤差は

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

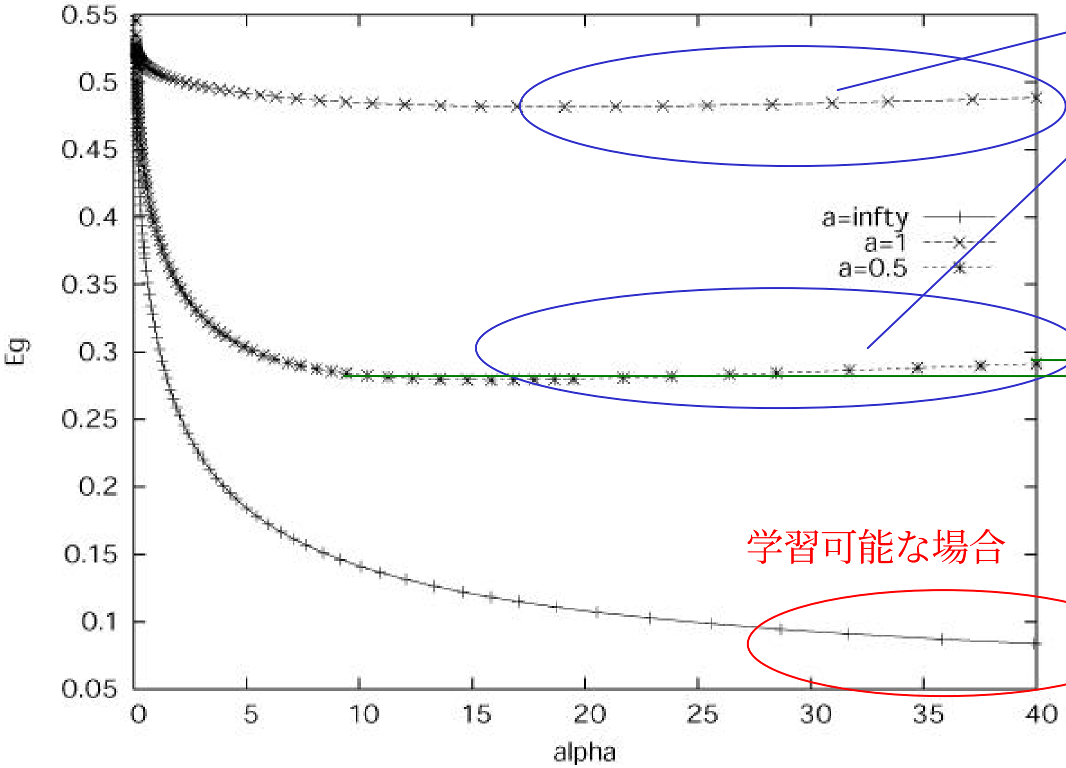
$a^{-1/3}$ 則に従う

Hebb学習のダイナミックスと過学習

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

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



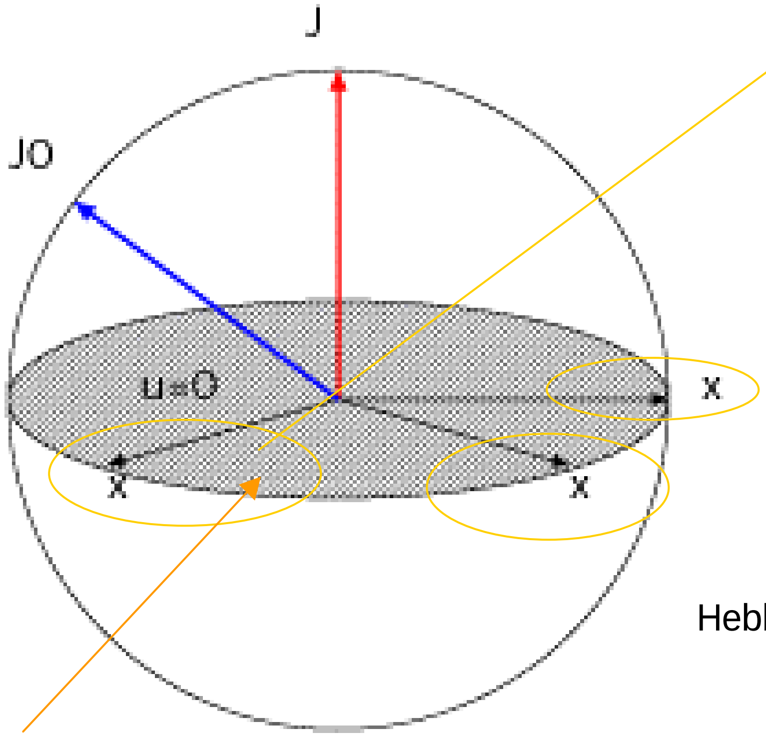
学習不可能な場合には
「過学習」が生じる

学べば学ぶほど
悪くなる

学習可能な場合

$a^{-1/2}$ 則に従う

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

$$P_R(v \mid u=0) = \sqrt{2p}\, d(u) P_R(u,v)$$

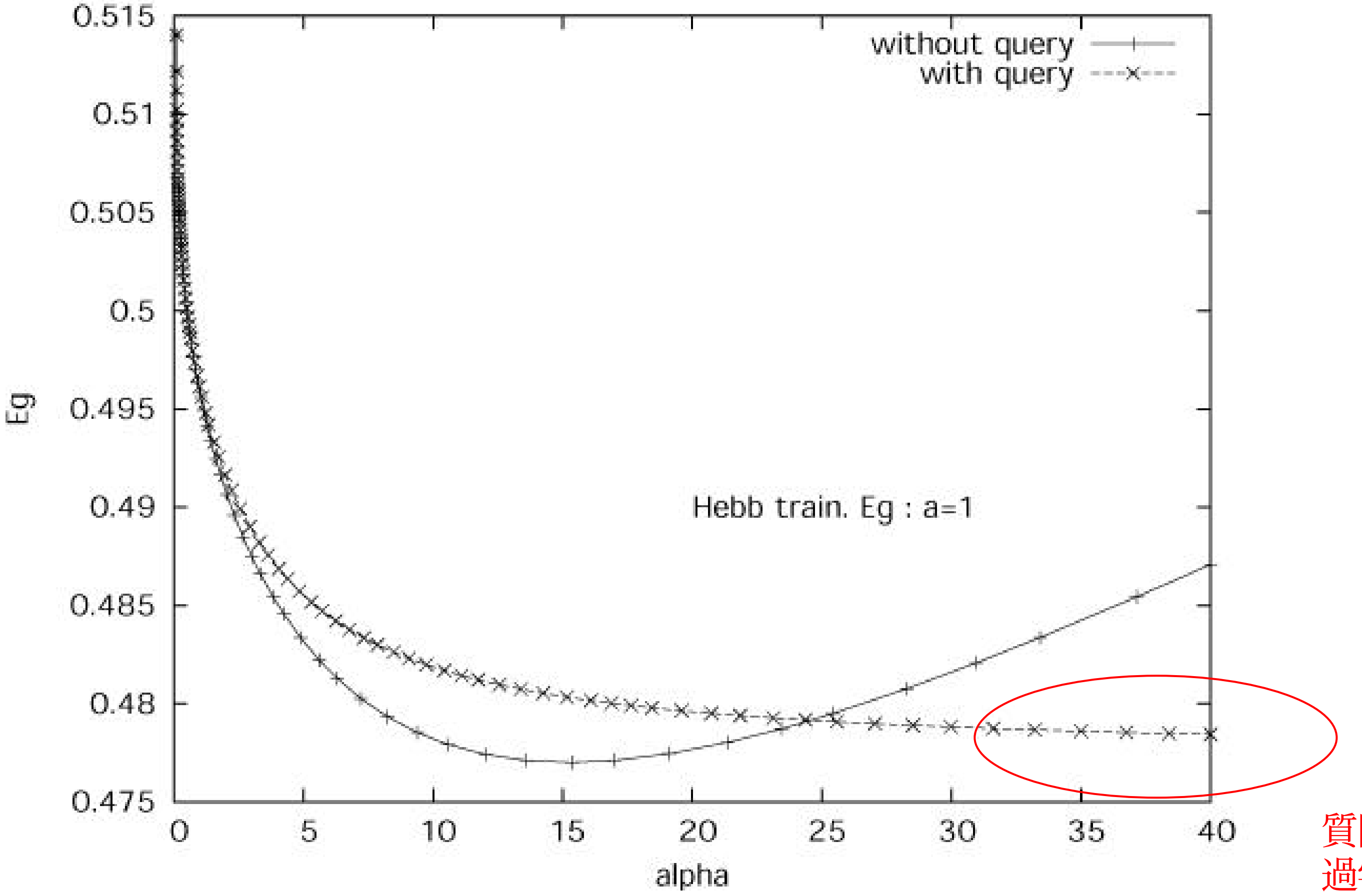
Hebb学習に適用すると　　状態更新式

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

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

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