混沌系工学特論 配布資料 #8

担当：井上 純一 (情報科学研究科棟 8-13)
URL：http://chaosweb.complex.eng.hokudai.ac.jp/˜j_inoue/

平成 17 年 1 月 24 日

目 次

3.5 オンライン学習 111
3.6 実現不可能な規則のオンライン学習 112
3.6.1 教師機械：$K = 3$ パリティ・マシン 112
3.6.2 マクロな量の導入と汎化誤差の一般形 113
3.7 学習過程のダイナミックス 115
3.7.1 学習方程式の導出 115
3.8 Hebb 学習と過学習 117
3.9 学習における「質問」の効果 118
3.9.1 Hebb 学習における質問の構築と過学習の消失 119

3.5 オンライン学習

前節では学習機械の性能を評価するための指標として汎化という概念を導入し, パーセプトロンの学習曲線 — 汎化誤差が例題数とともにどのように振舞うか — を統計力学の方法により調べた. まずは, それを簡単にまとめておこう.

与えられた例題に対して, 正しい入出力を実現するような解空間の体積を評価すると, 解空間は例題数とともに

$$\Omega_P(\epsilon, \alpha) \quad = \quad \exp[S(\epsilon, \alpha)] \tag{48}$$

のように振舞うことがわかった. 与えられた例題数 $\alpha = (P/N)$ 無限大で解が 1 点 $\boldsymbol{J} = \boldsymbol{T}$ に収縮しないかぎり, 有限の例題数では解空間も有限の体積を有するので, この空間の中のどの結合を選ぶか, により, 異なるパフォーマンスの学習機械ができあがることになる. 前節ではこの結合選択の戦略としては, 解空間からランダムに結合 $\boldsymbol{J}$ を選ぶ, ギブス学習を採用した. すると, 解空間のエントロピー S を最大にする汎化誤差を持つ学習機械が生き残ることになる[1] . そして, その生き残った学習機械の学習曲線が

$$\epsilon(\alpha) \quad = \quad \arg\max_{\epsilon} S(\epsilon, \alpha) \tag{49}$$

[1] 解空間のエントロピーは $S(\epsilon, \alpha)$ のように, 与えられた例題数 α に対して (マクロには) 汎化誤差 ϵ で特徴つけられるわけであるから, 解空間にはこれら異なる ϵ の値でラベル付けされた「サブ・シェル」が多数存在する (全解空間は複数のサブ・シェルで分割されている). しかし, そうしたサブ・シェルの中では圧倒的にエントロピーを最大化する ϵ_* で特徴つけられたものの占める割合が大きく, 従って, ランダムに解空間の中から 1 点 $\boldsymbol{J}$ をとってくるとなれば, おのずとその解 $\boldsymbol{J}$ が汎化誤差 ϵ_* を持つものである確率が圧倒的に高い. 当然, そのようにしてランダムに選ばれる $\boldsymbol{J}$ はミクロに見ればその都度違うが, いずれも同じ ϵ_* でマクロに特徴つけられた $\boldsymbol{J}$ である.

で与えられたわけである.

ところで, ある例題数 $P = \alpha N$ での解空間はこの P 個の例題全てに対して正しい答えを出すものとして定義されているわけであるから, この解空間の中から 1 点選ぶような学習則は前回学んだ学習様式で言えば**バッチ学習**のカテゴリーに入ることになる. 一方で, 学習様式にはもう一つ, **オンライン学習**があることも既に見ているが, ここではそれについて詳しく調べていくことにしょう.

3.6　実現不可能な規則のオンライン学習

前節でも述べたが, 生徒機械はブラックボックスからの入出力からその機構を推測するわけだが, その課題が学習可能であるものとは限らない. 従って, 現実的にはいくら沢山の例題数を与えても, 汎化誤差がゼロにならないという意味で「実現不可能」な規則を学習する場合の方が多いであろう. ここではそのような場合の中で, 教師機械の性能が生徒機械の性能よりはるかに優っている場合を考えてみることにする. つまり, 世の中にはいくら努力しても敵わない存在がいるわけで, そのような師匠から「よく学ぶ」にはどうしたら良いのか, を考えていこうというわけである.

3.6.1　教師機械： $K = 3$ パリティ・マシン

教師機械として次のような学習機械を考えてみよう. この学習機械は出力がそれぞれ次のように与えられる 3 つ ($K = 3$) のパーセプトロン：

$$T^{(1)}(v) = \mathrm{sgn}[v] \tag{50}$$

$$T^{(2)}(v) = \mathrm{sgn}[a - v] \tag{51}$$

$$T^{(3)}(v) = \mathrm{sgn}[a + v] \tag{52}$$

から成る. ここで, v は結合を $\boldsymbol{J}^0$, 大きさが 1 である入力ベクトルを $\boldsymbol{x}$ とした場合の内部ポテンシャルであり, $v = \sqrt{N}(\boldsymbol{J}^0 \cdot \boldsymbol{J})/|\boldsymbol{J}^0|$ で与えられる. また, a はある実数の閾値であるとしよう. このとき, 3 つの学習機械は自分の出力を持ち寄り, 3 人の出力の中で -1 の数が奇数ならば最終結論として出力 -1 を, 偶数であるならば $+1$ を出力するものとする. つまり, 個々のパーセプトロンは独立に入力を受け取り, 最終結果は「談合」により決める. こうした意味では, この学習機械は 3 つのエージェントからなる「集団学習」をしているとみなすことができるであろう. この機械の概念図を図 9 に載せた.

この機械の最終出力を式で書けば

$$T_a(v) = \mathrm{sgn}[v(a - v)(v + a)] \tag{53}$$

となる. $a \to \infty$ の極限では

$$\lim_{a\to\infty} T_a(v) = \mathrm{sgn}[v] \tag{54}$$

となり, 単純パーセプトロンの出力に一致することに注意しよう. このような構造を持つ学習機械を**パリティ・マシン**と呼ぶ. この機械は通常のパーセプトロンと比べて優っているのであろうか？ 実際, 様々な観点からこの機械の性能を議論することができるが, 「全ての可能な入出力 2^N 個の中のいくつを実現できるのか」という指標で両者を評価するのであれば, 単純パーセプトロンは前回も少し触れたように, 入力次元 N に対し, $2N$ 個のパターンを実現することができる[2]. 一方, このパリティ・マシンの場合にはレプリ

[2] もし, パーセプトロンの結合を $J_{ij} = (1/N)\sum_\mu \xi_i^\mu \xi_j^\mu$ に選び, 素子を全結合した場合の実現可能なパターン数は $0.138N$ であったことを思い出そう. 結合にこのような Hebb 則の制約を付けずに最良な選び方ができると考えた場合, 実現できるパターン数は $2N$ まで増える, ということをここでは言っている.

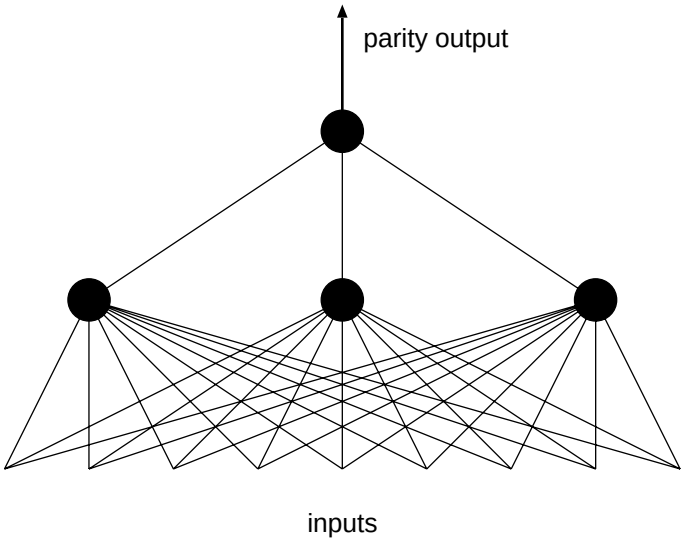

図 9: $K = 3$ パリティ・マシン. 3 つのパーセプトロンの出力の「パリティ」を最終出力とする.

カ法に基づく解析により, 閾値 a を適切に選べば $\sim 10N$ 程度まで上昇し, それ以外の全ての有限な a の値に対しても, 実現可能なパターン数は $2N$ 以上であることがわかっている. 従って, このパリティ・マシンはパーセプトロンと比べてはるかに表現能力が高く, その意味で, 生徒機械であるパーセプトロンにとってパリティ・マシンの規則は実現不可能なものとなるのである.

ところで, このパリティ・マシンは解析上は非常に便利である. というのも, この機械の与える入出力関係は閾値関数が $v = \pm a$ で反転した**非単調パーセプトロン**と同じであり, このパーセプトロンの解析を通じてパリティ・マシンの様々な側面を計算機シミュレーションによらずに明らかにすることができる. 従って, ここではその中で, 単純パーセプトロンがパリティ・マシンの持つ規則をオンライン式に学習する場合の学習曲線を詳しく見ていくことにしょう.

3.6.2 マクロな量の導入と汎化誤差の一般形

統計力学によれば, システムの性質はマクロな量を介して明らかになる. 従って, ここでもそのような量を導入しょう. そこで, 今後, 教師の結合を $\boldsymbol{J}^0$, 生徒の結合を $\boldsymbol{J}$, 入力ベクトルを 1 に規格化された $\boldsymbol{x}$ で表記することにする. このとき, 我々がここで導入するマクロな量は教師, 生徒機械の結合の重なり：

$$R = \frac{\boldsymbol{J}^0 \cdot \boldsymbol{J}}{|\boldsymbol{J}^0||\boldsymbol{J}|} \tag{55}$$

及び, 生徒機械のノルム：

$$l = \frac{|\boldsymbol{J}|}{\sqrt{N}} \tag{56}$$

である. 従って, ある学習則を与えて, 「オンライン的に」生徒の結合ベクトルが変化して行けば, その変化はその都度 R と l にも反映されることになる.

ここで, 問題を一旦整理しておく. 我々が考える状況は内部ポテンシャルが $u = \sqrt{N}(\boldsymbol{J} \cdot \boldsymbol{x})/|\boldsymbol{J}|$ で与えられ, 出力が

$$S(u) = \mathrm{sgn}[u] \tag{57}$$

で決まる生徒機械が, 内部ポテンシャルが $v = \sqrt{N}(\boldsymbol{J}^0 \cdot \boldsymbol{x})/|\boldsymbol{J}^0|$ で与えられ, その出力が

$$T_a(v) \quad = \quad \mathrm{sgn}[v(a-v)(v+a)] \tag{58}$$

で与えられる教師機械からオンライン学習をするというものであった. 入力は $|\boldsymbol{x}| = 1$ を満たす空間からランダムに取ってくるわけであるから, 中心極限定理より, u, v は正規分布に従い, その同時分布は相関：$\langle uv \rangle = R$ を用いて

$$P_R(u,v) \quad = \quad \frac{1}{2\pi\sqrt{1-R^2}} \exp\left[-\frac{u^2+v^2-2Ruv}{2(1-R^2)}\right] \tag{59}$$

と書ける. このとき, 汎化誤差 ϵ_g は全ての例題に対して生徒, 教師の出力が食い違う, つまり, $T_a(v) \neq S(u)$ となる確率で与えられるわけであるから,

$$\begin{aligned} \epsilon_g &= = \langle \Theta(-T_a(v)S(u)) \rangle \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} dudv\, \Theta(-T_a(u)S(v)) P_R(u,v) \\ &= 2\int_a^{\infty} Dv\, H\left(\frac{-Rv}{\sqrt{1-R^2}}\right) + 2\int_0^a Dv\, H\left(\frac{Rv}{\sqrt{1-R^2}}\right) \equiv E_a(R) \end{aligned} \tag{60}$$

のように, パラメータ a を与えると, 教師, 生徒機械の重なり R で与えられることになる. ここで, $H(x)$ は誤差関数：$H(x) = (1/\sqrt{2\pi})\int_x^{\infty} dz \mathrm{e}^{-z^2/2}$ である.

図 10 にいくつかの a の値に対して汎化誤差 $\epsilon_g = E_a(R)$ を R の関数としてプロットしたものを載せる. この図より, $a = \infty$ の学習可能な場合には $R = 1$ で汎化誤差がゼロになる. つまり, $\boldsymbol{J} = \boldsymbol{J}^0$ となるように

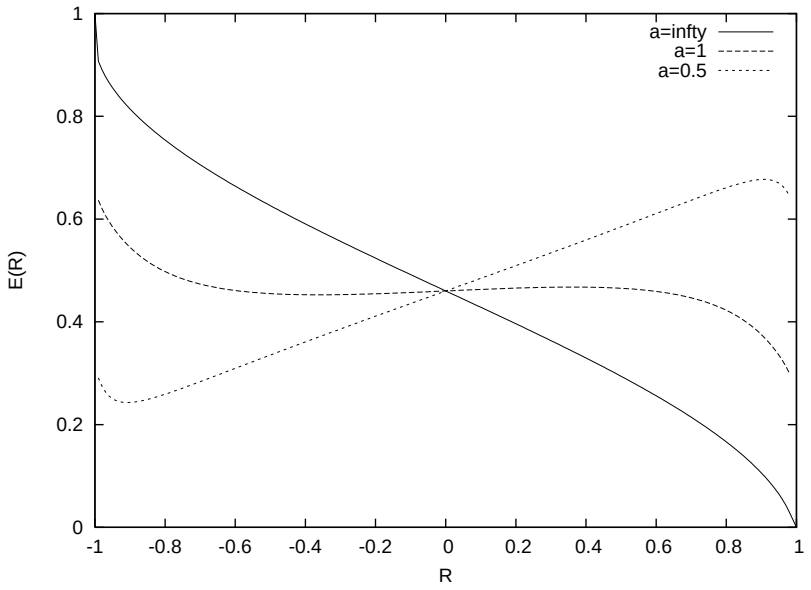

図 10: 重なり R の関数としての汎化誤差 ϵ_g. 教師機械の持つマクロなパラメータ a を $a = \infty, 1, 0.5$ に選んである.

学習が完結した段階で, 両者の重なりが 1 になり, 誤差も完全にゼロになる. 一方, $a = 1.0, 0.5$ のときには様子が違い, $R < 0$ の領域で極小値が現れる. この極小値は $a \leq a_{c1} = \sqrt{2\log 2}$ のときに現れ, その極小値を与える R の値は

$$R_* \quad = \quad -\sqrt{\frac{2\log 2 - a^2}{2\log 2}} \tag{61}$$

である. この極小値 $E(R_*)$ が最小値になる条件は $a \leq a_{c2} = 0.80$ であり, $a_{c2} \leq a < a_{c1}$ では極小 $E(R_*)$ は存在するが, 最大値は $R = 1$ のときの $E(1) = 2H(a)$ となる. 従って, 以上をまとめると, 最適な学習則に対して, 教師機械の規則の実現不可能性のために残留する汎化誤差 ϵ_{min} はパラメータ a の値に対して

$$\epsilon_g \quad = \quad \begin{cases} E\left(-\sqrt{\frac{2\log 2 - a^2}{2\log 2}}\right) & (a < a_{c2}) \\ 2H(a) & (a \geq a_{c2}) \end{cases} \tag{62}$$

となる.

ここで, 重要なのは, **以上の結論, とりわけ (60) 式は生徒機械の採用する学習則には依らない**ということである. 与えられた学習則が R をこの最小値 ϵ_{min} を与えるような値に導くのか, また, そのときの収束スピードはどの程度であるか, という我々の知りたい事柄は具体的に $\boldsymbol{J}$ の更新式を与えなければわからない. そこで, 具体的にいくつかの学習則を与えて, R-l の流れ図, 及び, 汎化誤差の発展式を見ていくことにしょう.

3.7　学習過程のダイナミックス

汎化誤差の例題数依存性は学習則, つまり, $\boldsymbol{J}$ の更新則を与えなければわからないわけだから, ここでは具体的に次のような更新式を考えてみよう.

$$\boldsymbol{J}^{m+1} \quad = \quad \boldsymbol{J}^m - \Theta(-T_a(v)S(u))S(u)\boldsymbol{x} \tag{63}$$

ここで, $\Theta(\cdots)$ は階段関数であり, この学習則は教師, 生徒の出力が食い違う場合のみ生徒の結合ベクトルの更新が行われ, 自分の答えは間違ったわけであるから, そのときの出力値と逆符号で結合を入力ベクトル方向に移動させよう, という意味を持っている. 前回の講義でみた, (**講義ノート #7** の図 3, 及び, **問い 10** 参照) 1 次元のパラメータ θ の学習の高次元への拡張版となっている. この学習則を**パーセプトロン学習**と呼んでいる.

この (63) 式で入力数を例えば $N = 10000$ と具体的に与えて計算機上でシミュレートし, 重なり R を更新ステップ数の関数として求め, それを (60) に逐次代入すれば, 汎化誤差の時間発展が実際に観測できる. しかし, 入力次元 N, 例題数 P がともに無限大に取れ, その比 $\alpha = (P/N)$ が有限値に取れる極限を考えると, 解析的にこの学習のダイナミックスを議論することができる.

3.7.1　学習方程式の導出

まずは, (63) 式の両辺を自乗して, それを $l^m = |\boldsymbol{J}^m|/\sqrt{N}, R^m = (\boldsymbol{J}^0 \cdot \boldsymbol{J}^m)/|\boldsymbol{J}^0||\boldsymbol{J}^m|$ を用いて書き直すと

$$2l\left(\frac{l^{m+1} - l^m}{\frac{1}{N}}\right) \quad = \quad -2l^m\Theta(-T_a(v)S(u))u + \Theta(-T_a(v)S(u)) \tag{64}$$

が得られるが, 結合ベクトル $\boldsymbol{J}$ の次元は N であることから, このベクトルの自乗で定義されるノルム l に $\mathcal{O}(1)$ の変化が見られるためには $\mathcal{O}(N)$ の例題数が必要となる. このとき, $(1/N) = d\alpha$ は十分小さな量であり, これで上式左辺を展開し, P, N が十分大きければ, 右辺はその平均値に一致する (自己平均) ことを用いれば

$$2l\frac{dl}{d\alpha} \quad = \quad -2l\langle\Theta(-T_a(v)S(u))u\rangle + \langle\Theta(-T_a(v)S(u))\rangle + \mathcal{O}\left(\frac{1}{N}\right) \tag{65}$$

が得られる. 同様に, (63) の両辺に $\boldsymbol{J}^0$ をかけて内積をとることにより,

$$l^2\frac{dR}{d\alpha} \quad = \quad -\frac{R}{2}\langle\Theta(-T_a(v)S(u))\rangle + l\left(R\langle\Theta(-T_a(v)S(u))u\rangle - \langle\Theta(-T_a(v)S(u))v\rangle\right) + \mathcal{O}\left(\frac{1}{N}\right) \tag{66}$$

が導かれ, 同時分布による平均 $\langle\cdots\rangle$ は直ちに計算できるので, 結局, 生徒機械の学習過程は次のようなマクロな量 l, R に関する閉じた微分方程式：

$$\frac{dl}{d\alpha} = \frac{1}{l}\left[\frac{E_a(R)}{2} - F_a(R)l\right] \tag{67}$$

$$\frac{dR}{d\alpha} = \frac{1}{l^2}\left[-\frac{R}{2}E_a(R) + (F_a(R)R - G_a(R))l\right] \tag{68}$$

で与えられることになる. ここに,

$$F_a(R) = \langle\langle\Theta(-T_a(v)S(u))u\rangle = -\frac{R}{\sqrt{2\pi}}(1-2\Delta) + \frac{1}{\sqrt{2\pi}} \tag{69}$$

$$G_a(R) = \langle\langle\Theta(-T_a(v)S(u))v\rangle = -\frac{1}{\sqrt{2\pi}}(1-2\Delta) + \frac{R}{\sqrt{2\pi}} \tag{70}$$

であり, $\Delta = \mathrm{e}^{-a^2/2}$ と置いている. 図 11 にいくつかの a の値に対する, R-l 空間の流れ図を載せよう. この

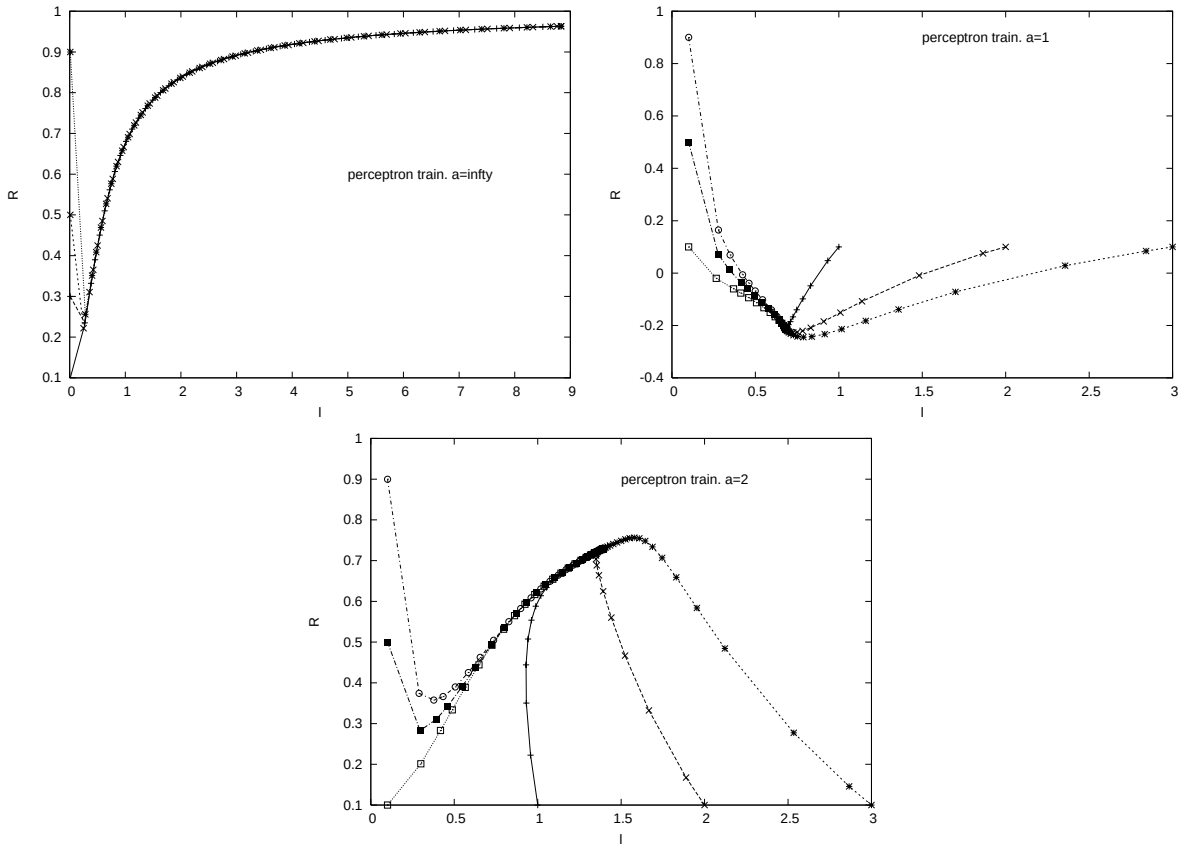

図 11: パーセプトロン学習での R-l 流れ図. 学習可能な場合 ($a=\infty$) の場合には $R=1$ へと漸近していくが, 実現不可能な場合 ($a=1,2$) にはある固定点に引き込まれてしまう. この固定点は ϵ_{min} を与えない. しかも, 引き込まれ具合は指数的に速い.

図より, 学習可能な場合については, ϵ_{min} を与える $R=1$ へ順調に漸近していくことがわかる. この場合の α が十分大きな漸近領域での振舞いは図 11 で得られた R の値を $E_a(R)$ に代入して, それを対数プロットしてみると, 図 12 のようになり, $\alpha^{-1/3}$ 則に従うことがわかる. これは数値的にではなく, 解析的にも汎化誤差を $\alpha\to\infty$ で展開することにより

$$\epsilon_g = k\,\alpha^{-1/3}, \quad k = \sqrt{2}(3\sqrt{2})^{-1/3}\pi \tag{71}$$

のように振舞うことが確かめられる．この結果より, 同じ例題を与えた場合に汎化誤差を通じて評価される精度は前回のギブス・バッチ学習 (α^{-1} 則) に比べて落ちることがわかる．これをどのように改善するか (加速するか) は後に (次回最終回：#**9 講義ノート**) 述べることにしょう．

さて, 学習が不可能な場合, R は ϵ_{min} を与える値に収束しない．しかも, 悪いことにこの非最適値への収束は指数関数的に速い．a が無限大ではないが比較的に大きな場合の漸近解析によれば汎化誤差はこの領域で

$$\epsilon_g \quad = \quad \frac{1}{\pi}\Gamma\left(\frac{1}{4}\right)\Delta^{3/4} + \frac{\sqrt{2}}{\pi}\exp\left(-\frac{2\Delta^{2/3}}{\pi}\alpha\right) \tag{72}$$

のように振舞うことがわかる．ここで, $\Gamma(\cdots)$ はガンマ関数であり, 上の方程式 (72) の右辺第 1 項は ϵ_{min} とは異なる値である．ちなみに, $\Delta = \mathrm{e}^{-a^2/2}$ であるから, 学習可能となる極限で (72) 式の指数部の「緩和時間」は発散し, 収束性は指数則から冪則に落ちて (71) 式に従うようになる．

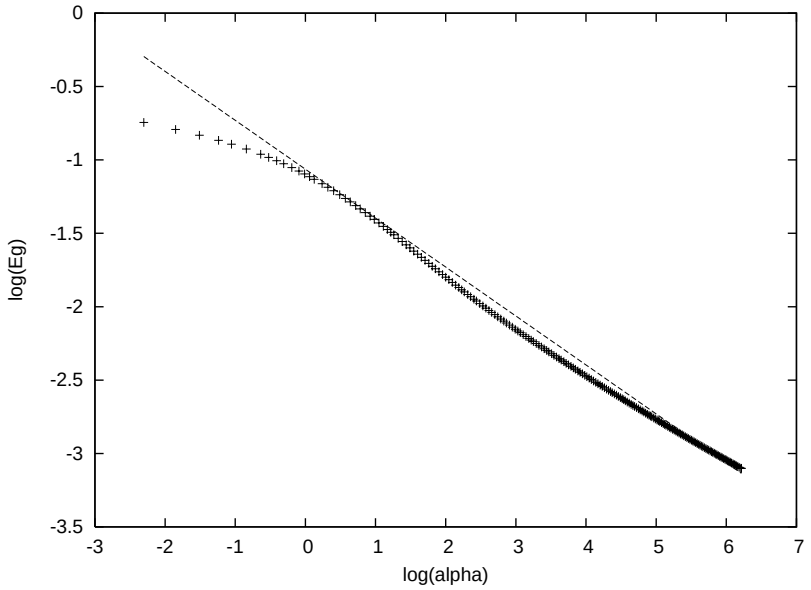

図 12: 学習可能な場合のパーセプトロン学習 (63) に対する学習曲線の漸近領域でのスケーリング則. $\alpha^{-1/3}$ 則に従う.

3.8 Hebb 学習と過学習

前節ではパーセプトロン学習の学習曲線に関して議論したが, 学習則としてはこれよりも素朴に

$$\boldsymbol{J}^{m+1} \quad = \quad \boldsymbol{J}^m + T_a(v)\boldsymbol{x} \tag{73}$$

を採用することもできる．これは教師機械の出力符号に応じて, 入力ベクトル方向に自分の結合を向けていく学習であり, ここでは **Hebb 学習**と呼ぶことにしょう．これも同様に, このミクロな成分の発展方程式 (73) から l と R のマクロな量に関する微分方程式を導くことができる．結果は

$$\frac{dl}{d\alpha} \quad = \quad \frac{1}{l}\left[\frac{1}{2} + \frac{2R}{\sqrt{2\pi}}(1-2\Delta)l\right] \tag{74}$$

$$\frac{dR}{d\alpha} \quad = \quad \frac{1}{l^2}\left[-\frac{R}{2} + \frac{2}{\sqrt{2\pi}}(1-2\Delta)(1-R^2)l\right] \tag{75}$$

に従う．この微分方程式を数値的に解き, $\epsilon_g = E_a(R)$ の発展をいくつかのパラメータ a に対してプロットしたものを図 13 に載せる．この図より, $a = \infty$ の学習可能な場合, 汎化誤差は単調にゼロへと向かい, そ

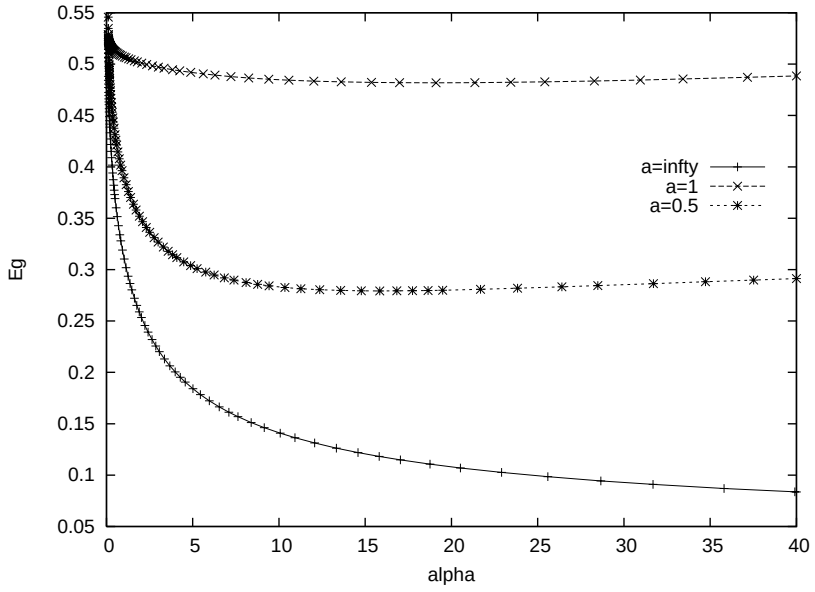

図 13: Hebb 学習の学習曲線．学習可能な場合 ($a = \infty$), 汎化誤差は単調にゼロへと収束するが, $a = 0.5, 1$ の学習不可能な場合には例題数を与えれば与えるほど汎化能力が減少する「過学習」が見られる．

の減少の仕方は漸近的に $\alpha^{-1/2}$ 則であることがわかる．これは数値的にも図 14 のように確認できる．

ところで, 学習不可能である, 例えば, $a = 1, 0.5$ の場合には図 13 からわかるように, 例題をもらえばもらうほど汎化能力が落ちていくという意味で**過学習**が生じている．つまり, 皮肉なことに「良い加減」(けっして「いい加減」ではない．念のため) のところで学習をストップしないと, それ以上の学習は返って生徒機械の害になるというわけである．Hebb 学習の $\alpha \to \infty$ での漸近形をもう少し詳しく調べても見ると

$$\epsilon_g = \begin{cases} \frac{1}{\sqrt{2\pi}(1-2\Delta)} \frac{1}{\sqrt{\alpha}} + 2H(a) & (a > a_{c1}) \\ \frac{1}{\sqrt{6\pi}(1-2\Delta)} \frac{1}{\sqrt{\alpha}} + 1 - 2H(a) & (a < a_{c1}) \end{cases} \tag{76}$$

であることがわかる．従って, $a < a_{c1} = \sqrt{2\log 2}$ で過学習が生じることになる．

次節では生徒機械に「質問」をすることを許すことにより, この Hebb 学習の過学習がどのようになるのか, を見ていくことにしよう．

3.9　学習における「質問」の効果

我々は学生時代を通じて常に「わからなかったら質問してください」というアドバイスを教師から受けてきた．そしてその質問が良い質問であれば, その後の学習が進んだり, 教師に褒められたりしたものである[3]．そこで, そのような「質問」が, 我々の今まで調べてきた機械学習の文脈でも可能であるか, そして, どのような効果が見込めるのかを調べてみることは, とても意味のあることであろう．そこで, ここでは機

[3] 良い質問だけが歓迎されるかというと, そうでもないらしい．あの湯川秀樹は学会/研究会等でポイントのずれた質問を連発したそうだが, その質問がかえってその研究会を盛り上げて, 議論が活気づき, とても良かったということをどこかで聞いたことがある．しかし, 我々としては, 学会の中にはそうした質問が許容されるようなリベラルなものばかりがあるわけではない, ということは気にしておくべきであろう．ちなみに, この講義ではどの種の質問も歓迎してます．

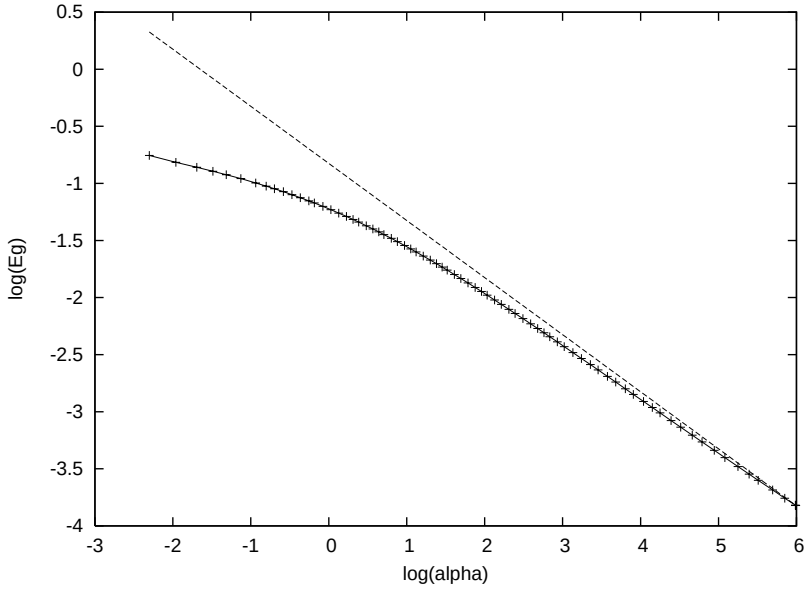

図 14: Hebb 学習の学習曲線の例題数無限大での漸近形. $\alpha^{-1/2}$ 則に従う.

械学習における質問をどのように構成し, その効果はどの程度のものなのか, をこれまでと同様に汎化誤差を通じて調べていくことにしょう.

まず, 「質問」と言った場合, 答えそのものズバリを聞いてはいけないことは明らかであろう. それが可能ならば何も苦労はない. 従って, ここでは結合 $\boldsymbol{J}^0$ の情報を直接的には質問の中に取り入れることはできない. それではどうするかと言うと, 入力ベクトル $\boldsymbol{x}$ の中で最も情報を担っていると思われるものをピックアップし, その入力と教師機械の出力を 1 セットとしてここでの「質問」とすることにしょう. 今までの議論では $|\boldsymbol{x}| = 1$ を満たす一様な分布から入力を取ってきていたわけで, その意味では質問による「バイアス」がかかっているわけではなかったのではあるが, ここでは**入力 $\boldsymbol{x}$ の空間に制約を入れる**ことによって質問としようというわけである.

それでは単純パーセプトロンである生徒機械にとって, どの入力が最も多くの情報を含むのか ? これは明らかに $u = \sqrt{N}(\boldsymbol{J}\cdot\boldsymbol{x})/|\boldsymbol{J}| = 0$ を満たす入力 $\boldsymbol{x}$, つまり, **学習の各ステップで生徒機械のパターン分離面上を向いた入力ベクトルを要求するのが最も効果的**であろう (図 15 参照). ここに落ちた入力に対する生徒機械のとる判断が最も難しく, 従って, それだけ多くの情報を含むことになる.

3.9.1 Hebb 学習における質問の構築と過学習の消失

このような入力を要求する場合, 教師の内部ポテンシャル v の分布は

$$P_R(v|u=0) \quad = \quad \sqrt{2\pi}\delta(u)P_R(u,v) \tag{77}$$

に従う. よって, Hebb 学習でこの質問を用いるのであれば[4] , その発展方程式は例によって (73) から R と l に関する方程式を組み立てることにより,

$$2l\frac{dl}{d\alpha} \quad = \quad \ll 1 \gg +2l \ll T_a(v)u \gg \tag{78}$$

[4] パーセプトロン学習にこの $u = 0$ での質問は導入できないことに注意しよう. パーセプトロン学習は $\mathrm{sgn}(u)$ のような生徒機械の出力を含むので, $u = 0$ とすることができないためである. なお, 生徒機械が a という階層の一つ高いパラメータの値を知っているのであれば, これを用いて $u = \pm a$ からの入力を取ってくることによって質問を構築することは可能である.

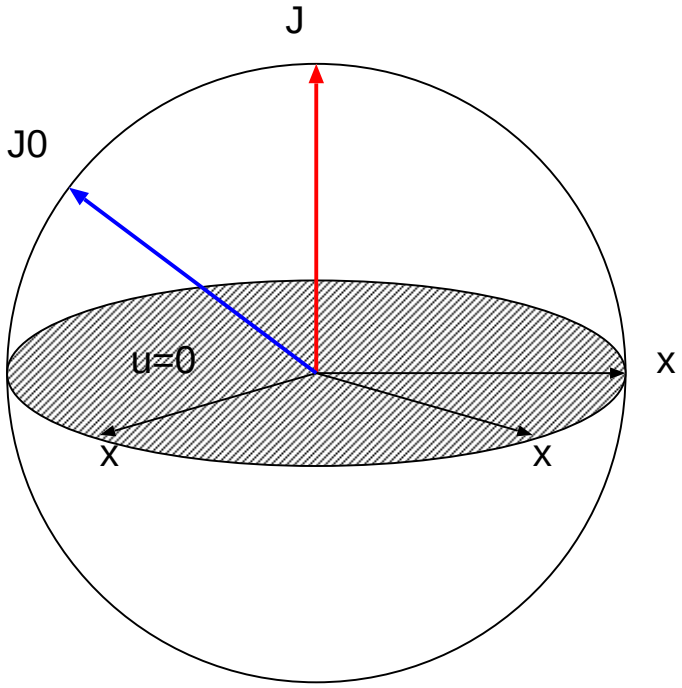

図 15: 入力として, 各ステップで分離面 $u = 0$ を満たすような入力ベクトル $\boldsymbol{x}$ を要求する. これが「質問」となる.

$$l\frac{dR}{d\alpha} + R\frac{dl}{d\alpha} \quad = \quad \ll T_a(v)v \gg \tag{79}$$

で与えられる. ここで, $\ll \cdots \gg$ は分布 (77) 式での平均を意味する. 従って, この平均を具体的に計算することにより, 学習方程式は R のみで書けて (l は $l = \sqrt{\alpha}$),

$$\frac{dR}{d\alpha} \quad = \quad \frac{1}{\sqrt{\alpha}}\left[\sqrt{\frac{2}{\pi}}\sqrt{1-R^2}\left\{1 - 2\,\mathrm{e}^{-\frac{a^2}{2(1-R^2)}}\right\} - \frac{R}{2\sqrt{\alpha}}\right] \tag{80}$$

が得られる. 図に前節でみた $a = 1.0$ の場合に過学習が起こる場合とそれに質問を加えた場合の学習曲線を載せる. この図より, 質問を考慮して, 各学習ステップで入力空間に制約を入れることにより, 過学習が消失し, 汎化誤差は理論上の最小値 ϵ_{min} に収束することがわかる.

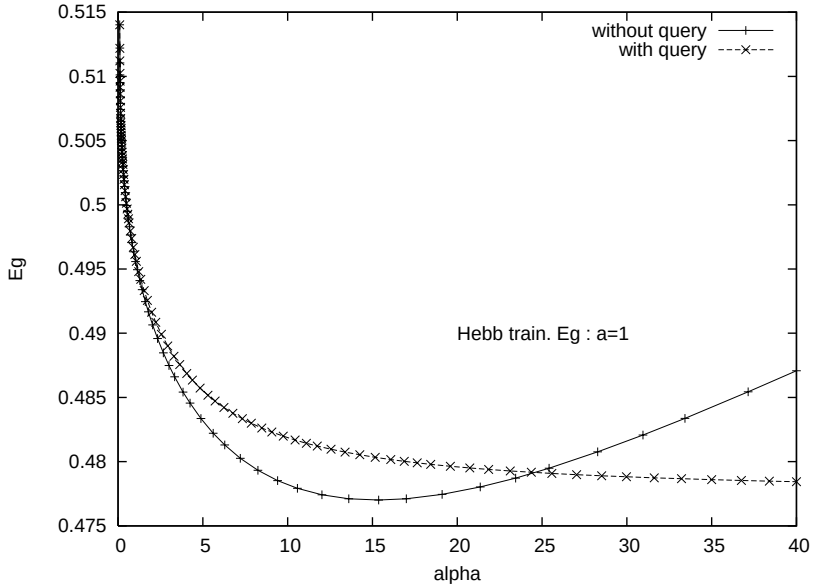

図 16: $a = 1.0$ の場合に過学習が生じる Hebb 学習の学習曲線 (without query) とそれに質問の効果を加えた場合の学習曲線 (with query).

問 12 :
ここで調べた Hebb 学習：

$$\boldsymbol{J}^{m+1} \quad = \quad \boldsymbol{J}^{m} + T_a(v)\boldsymbol{x}$$

を解析的にではなく，$N = 5000$ 程度のシステムサイズでの計算機シミュレーションを実行することにより調べよ．具体的には

(1) R**-**l **の流れ図を描く．**
(2) 汎化誤差 $E_a(R)$ **の発展を描く．**

の 2 点に関して調べること．特に，$a = 1$ の場合に過学習が確認できるかチェックせよ．入力ベクトル $\boldsymbol{x}$ は各成分を $[-1, 1]$ の一様乱数から取り出し，その大きさで 1 に規格化して用いると良い．余裕のある者は $u = 0$ を満たす入力ベクトルを作り，それを用いた場合に過学習が消失するか否かを確認せよ．