ただし

$$\tilde{F}_{ij}=\tilde{f}(\xi_i, \xi_j), \qquad i, j=1, \ldots, k,$$

$$\tilde{f}(\xi, \eta)=\begin{bmatrix} \tilde{S}(\xi, \eta) & \tilde{I}(\xi, \eta) \\ \tilde{D}(\xi, \eta) & \tilde{S}(\eta, \xi) \end{bmatrix},$$

$$\tilde{S}(\xi, \xi)=1, \qquad \tilde{D}(\xi, \xi)=\tilde{I}(\xi, \xi)=0,$$

$$\tilde{S}(\xi, \eta)=\frac{\sin[\pi(\xi-\eta)]}{\pi(\xi-\eta)},$$

$$\tilde{D}(\xi, \eta)=\frac{d}{d(\xi-\eta)}\left[\frac{\sin[\pi(\xi-\eta)]}{\pi(\xi-\eta)}\right],$$

$$\tilde{I}(\xi, \eta)=-\frac{1}{2}+\int_0^{\xi-\eta}\frac{\sin[\pi(\xi-\eta')]}{\pi(\xi-\eta')}d(\xi-\eta').$$

これらの場合については Balian のアンサンブルの局所準位相関の普遍性が証明された．このように歪直交多項式の漸近形の性質と Balian のアンサンブルの局所準位相関の普遍性が密接に結びついていることがわかる．

学習の統計的理論

電子技術総合研究所　麻　生　英　樹

Abstract. Learning process can be formalized as obtaining an efficient description of the structure or relation in the data from given training data. If the training data are given randomly according to some probabilistic distribution, the result of the learning becomes random variables. In this case, the learning process should be investigated as a statistical process.

Recently, in the field of computational learning theory, a statistical formalization of learning called "PAC (Probably Approximately Correct) Learning" was proposed and investigated. Here, techniques of statistical inference and statistical mechanics are playing important roles.

In this report, we give an overview of the statistical formalizations of learning from a rather general point of view. Some results on the relation between the size of training samples and the goodness of the result of learning which are obtained by uniform convergence method are also given as an example.

要　旨

学習用の事例が確率分布に従って与えられる条件の下での学習過程は，統計的な研究の対象である．計算論的学習理論の分野では，近年，PAC（Probably Approximately Correct）学習の枠組みなどが提案され，盛んに研究が行なわれている．

本報告では，統計的な学習の理論の一般的な定式化について概観し，一つの具体例として，Uniform convergence method による学習用データの数と学習結果の評価との関係についての結果を紹介する．

1. はじめに

情報処理の課題が従来の閉じた世界における完全情報の下での処理から開いた世界における処理へと拡張され，個別の状況に即した処理を行なうための情報表現やアルゴリズムが複雑化し，あらかじめ確定しにくいものになるにつれて，また，ハードウェア技術の進歩によって，ある程度の学習・自己組織化能力のハードウェア的な実現も現実的になってくるにつれて，「学習・適応」能力を持つシステムへの期待が高まっている．これに伴い，「学習」の問題を数理的にきちんと定義し，どのような対象が「学習可能」であるのか，そのための「学習のアルゴリズム」としてはどのようなものがよいのか，といった事柄を理論的に明らかにしようとする研究の重要性が増している．

学習過程の数理的な研究は，これまでに，計算複雑さの理論，統計的推測の理論，人工知能やニューラルネットワークにおける学習の理論，など複数の研究の流れの中で，それぞれの文脈に応じた形で行なわれてきているが，明らかに，相互に非常に強い関連を持っている．以下では，学習過程に関する統計的理論について概観する．

2. 学習の統計的理論

学習過程の定式化に含まれる要素は学習環境と学習者である．学習環境はある種の情報源であり，学習者に対して，学習用のデータを提供する．学習者はそのデータを用いて，学習環境の構造を反映した行動ルール（仮説）の自分自身の持つ内部表現による表現を獲得する．これが学習結果である．

学習結果に基づいて，学習者はなんらかの行動を行ない，その行動の結果は，学習環境によって評価され，評価値が返される．これが学習結果の評価の基礎となる．

上のような漠然とした枠組みをより具体化したものは学習モデルと呼ばれる．学習モデルにはさまざまなものがあるが，本稿では，例として次のような単純なものをとりあげる．

学習環境からは，集合（確率変数）Z 上のある確率分布 $P(Z)$ に従って，データが独立にサンプルされるとする．さらに，Z は二つの集合 X, Y の直積であるとする．従って，Z の要素 z は (x, y) と書ける．以下では，最も簡単な場合として，X として R^k, Y として $\{0, 1\}$ を想定する．

学習者は，$P(Z)$ に従ってランダムに生成された m 個のデータ $\{(x_i, y_i)\}_{i=1}^m$ を用いて，X の値から Y の値を予測するための仮説を学習する．このデータを D^m と書く．D^m の従う分布 $\prod_{i=1}^m P(x_i, y_i)$ を便宜的に Q と書くことにする．以下では，予測のための仮説は決定的なもの，すなわち，X から Y への関数であるとし，学習によって得られた仮説を $h(x)$ と書く．D^m から $h(x)$ を得る手続きを学習のアルゴリズムと呼ぶ．

この問題の場合には，学習によって得られた仮説の一つの評価として，次のような予測の期待誤差を使うのは自然であろう．

$$R(h)=\int(y-h(x))^2P(x, y)dxdy$$

$Y=\{0,1\}$ の場合には，この値は予測の期待誤り率となる．

さて，$h(x)$ は，学習用のデータ D^m に依存して決まるが，D^m は Q に従って確率的に決まるため，$h(x)$ も確率的に定まる．したがって，$h(x)$ の評価値 $R(h)$ は確率変数である．そこで，学習アルゴリズムの評価，さらには，学習問題のむずかしさの評価には，この $R(h)$ の確率分布の性質を用いることになる．

この評価には，次の二つの量がしばしば用いられる．

$$U(\varepsilon)=\mathrm{Prob}_{D^m\sim Q}[R(h(D^m))\leq\varepsilon]$$
$$V=E_{D^m\sim Q}[R(h(D^m))]$$

$U(\varepsilon)$ は期待誤り率 $R(h)$ が区間 $(0,\varepsilon)$ の中にある確率であり，V は $R(h)$ の平均値である．$U(\varepsilon)$ を評価に用いる学習モデルは，計算論的な学習理論の分野において Valiant によって提案された PAC (Probably Approximately Correct) 学習モデルと呼ばれるものになる (Valiant (1984), Angluin (1988))．一方，V を評価とする理論は，平均学習曲線の理論などと呼ばれている (Levin et al. (1990), Haussler et al. (1991))．

以上のような枠組みの上で，問題に適した仮説の表現およびそれを用いた学習アルゴリズムの提案，一定の評価を達成するために必要な学習用データの数や学習アルゴリズムの計算量の理論的評価，などが研究課題となっている．以下では，一つの例として，Uniform convergence method を用いた $U(\varepsilon)$ の評価を紹介する．

3. Uniform convergence method

上のような課題を解決するために，さまざまな技法が用いられているが，中でも $U(\varepsilon)$ と学習用データの数との関係を評価するための比較的汎用性の高い技法の一つに，Uniform convergence method がある (Blumer et al. (1989))．

$U(\varepsilon)$ の評価のめんどうさは，$P(x,y)$ が未知であることによる．そこで，学習用データに対する誤り率

$$R_{emp}(h)=\frac{1}{m}\sum_{i=1}^{m}(y_i-h(x_i))^2$$

を $R(h)$ の推定量と考える．$R_{emp}(h)$ は簡単に評価できるから，この推定量が学習用のサンプル数 m の増加に伴ってどのように $R(h)$ に近付いてゆくかがわかれば，$R(h)$ を評価できる．

今，簡単のために，学習アルゴリズムは $R_{emp}(h)=0$ とするような h を出力するとする．すなわち，学習アルゴリズムは，学習用のデータに関しては完全に正解を与える仮説を出力するとする．学習者が用いている表現によって表現できる仮説の全体を H とすると，このような $h\in H$ は一般に複数存在するが，学習アルゴリズムがそのうちのどれを出力するかはわからないとする．

このとき，次の一連の定理が成り立つ (Vapnik (1982))：

定理 3.1. (Vapnik and Chervonenkis (1971))

$$\mathrm{Prob}_{D^m\sim Q}\left[\sup_h R(h)>\varepsilon\right]<8\Pi_H(2m)\exp\left\{-\frac{\varepsilon m}{4}\right\}.$$

ここで，Π_H は，仮説の集合 H の成長関数 (growth functions) と呼ばれる関数で，以下のように定義される：

成長関数の定義

1. X の中の m 点の集合を S とする.
2. H の要素 h を一つとると, S は $h(x), x\in S$ の値によって二分割されるが, $\Pi_H(S)$ を, H の要素全体によって実現される S の異なる二分割の仕方の全体の数とする (したがって, 明らかに $\Pi_H(S)\leq 2^m$).
3. $\Pi_H(m)$ を, $\Pi_H(S)$ の S のとり方に関する最大値によって定義する.

成長関数については, 次のような性質が一般に成り立つ:

VC 次元の定義と成長関数の性質

1. 成長関数 $\Pi_H(m)$ の値が 2^m になるような m の最大値 (最大値がない場合は ∞) を仮説の集合 H の VC 次元と呼び, d と書くことにする.
2. このとき, 任意の $m\geq d\geq 1$ に対して $\Pi_H(m)\leq 1.5(m^d/d!)$ が成り立つ.

この性質を用いると, 次の定理が示せる.

定理 3.2. H の VC 次元を $d<\infty$ とするとき,

$$[\mathrm{Prob}]_{D^m\sim Q}\left[\sup_h R(h)>\varepsilon\right]<12\frac{(2m)^d}{d!}\exp\left\{-\frac{\varepsilon m}{4}\right\}.$$

このことから, さらに, 次の定理が示せる.

定理 3.3. $0<\delta<1$ として, 学習用データの数 m が,

$$m\geq m(\varepsilon,\delta)=\max\left(\frac{32}{\varepsilon}\ln\frac{8}{\delta},\frac{64d}{\varepsilon}\ln\frac{64}{\varepsilon}\right)$$

であれば, $R_{emp}(h)=0$ となるような h に対して, h および Q によらずに, 確率 $1-\delta$ 以上で「$0\leq R(h)\leq\varepsilon$ が成り立つ」といえる.

定理 3.1 は, m が増加する場合に, 期待誤り率 $R(h)$ が経験的誤り率 $R_{emp}(h)$ にどのような速度で h に関して一様に収束してゆくかを与えるものである. H の VC 次元とは, 直観的には, 仮説の集合 H の豊かさ, 広さ, 複雑さの一つの評価である. 直観的に考えると, H が広い場合には, 学習用のデータに対して完全に正解する $h\in H$ も多様になり, 従って, 学習用データについては正解しているのに, 期待誤り率は悪いというものが存在する可能性が増える. 定理 3.2 はこの直観に対応するものである.

こうして, 学習用データが定理 3.3 の $m(\varepsilon,\delta)$ 以上あれば, それらについて完全に正解するような仮説 h のうちで, 期待誤り率が最悪のものを選んでしまったとしても, その期待誤り率が ε を越える確率は δ を越えないということが, D^m の分布 Q によらずに, H の VC 次元にのみよる形で示せたことになる. 計算論的に見た場合に, 定理 3.3 のポイントは, d が有限であれば, $m(\varepsilon,\delta)$ が $1/\varepsilon$ および $1/\delta$ に関して多項式オーダーである (指数オーダーではない) という点にある. このことは, たとえば, 期待誤り率の上限 ε を十分に小さくしようと思った場合に, 必要となる学習用データの数が $1/\varepsilon$ の多項式オーダーでしか増えないことを意味している. これは, 実際的なデータ数で安定した学習を得ることができるために重要な条件である.

ここで述べた技法は, より一般の X, Y あるいは, さらに一般の問題設定についても応用さ

れている（Blumer et al. (1989), 麻生 (1992)). また，最近 Haussler は，こうした VC 次元を用いた評価と Shanon の情報量を用いた評価との関連を明らかにしている（Haussler et al. (1991)).

4. おわりに

学習過程に関する統計的な理論的研究について概観した．学習過程は，情報論的には，符合化過程にさまざまな仮説の表現方式および学習結果の評価方式を組み合わせたものと考えることができる．この分野は，統計数理的なアプローチが非常に有効な分野であり，従来の統計数理における，確率分布推定，仮説検定，モデル選択技法などと非常に関連が深い．また，統計力学における各種の計算技法なども有効に活用されている．今後，この分野のいっそうの発展が期待される．

参考文献

Angluin, D. (1988). Queries and concept learning, *Machine Learning,* **2**, 319-342.

麻生英樹 (1992). 学習の数理的理論の展開，電子技術総合研究所彙報，**56**, 602-635.

Blumer, A., Ehrenfeucht, A., Haussler, D. and Warmth, K. (1989). Learnability and the Vapnik-Chervonenkis dimension, *Journal of the ACM,* **36**, 929-965.

Haussler, D., Kearns, M. and Shapire, R. (1991). Bounds on the sample complexity of Bayesian learning using information theory and the VC dimension, *Proceedings of the 4th Annual Workshop on Computational Learning Theory,* 61-74, Morgan-Kaufmann, Palo Alto, California.

Levin, E., Tishuby, N. and Solla, S.A. (1990). A statistical approach to learning and generalization in layered neural networks, *Proceedings of the IEEE,* **78**, 1568-1574.

Valiant, L.G. (1984). A theory of the learnable, *Communication of the ACM,* **27**, 1134-1142.

Vapnik, V. (1982). *Estimation of Dependences Based on Empirical Data*, Springer, New York.

Vapnik, V. and Chervonenkis, A. Ya (1971). On the uniform convergence of relative frequencies of events to their probabilities, *Theory Probab. Appl.,* **16**, 264-280.

Marginal Fermi Liquid と Ward-Takahashi 関係式
—— Fermi 面の消失とバーテックスの発散について ——

名古屋商科大学 商学部　豊田　正

統計物理学における具体的な結果は応答関数，Green 関数等の相関関数と速度分布，密度分布等の分布関数で表わされる場合が殆どである．通常，与えられたモデル・ハミルトニアンからこれらの関数を求めることになるが，厳密解が得られる場合以外は，なんらかの近似を導入することになる．ハミルトニアンから得られるのは，Green 関数等の運動方程式であり，それらは一般に積分方程式である．粒子間相互作用がある場合，例えば 1 体 Green 関数の積分方程式は積分核に 2 体以上の Green 関数を含む．2 体 Green 関数の方程式は積分核に 3 体以上の Green 関数を含む．このように，粒子間相互作用がある場合，Green 関数の方程式系は閉じない（BBGKY）．実際の計算では，2 体 Green 関数までしか考慮しない場合が普通である．この場合 3 体 Green 関数を 2 体及び 1 体 Green 関数で近似することがまず行なわれる．その例としては，Born-Kirkwood 近似などが有名である．その結果，1 体 Green 関数と 2 体 Green 関数からなる方程式系が近似として得られる．問題は，このようにして得られた近似方程式系が元の厳