深層ニューラルネットワークの解剖
――統計力学によるアプローチ

吉野　元　〈大阪大学サイバーメディアセンター　yoshino@cmc.osaka-u.ac.jp〉

深層ニューラルネットワーク（Deep Neural Network, DNN）を用いた機械学習は，深層学習とよばれ，画像認識，機械翻訳などで身近なものとなった．しかしその高い学習能力のメカニズムはよくわかっておらず，ブラックボックスとして使われている面が無視できない．最先端の応用では様々なノウハウが駆使されるが，単純化した状況設定から考える物理学の発想がこのブラックボックスにメスを入れるのに役立つであろう．ニューラルネットワークを用いた機械学習はスピングラスに端を発するランダム系の統計力学，情報統計力学において伝統的に重要なテーマである．

Nビットの入力を，Nビットの出力に変換する「関数」を，DNNでデザインすることを考えてみよう．このNをDNNの「幅」とよぶことにする．入出力を含めて，ネットワークには多数のニューロンがある．あるニューロンの状態を変数S_iで表そう．これが入力信号$h=\sum_j J_{ij}S_j$の関数として$S_i=f(h)$で決まるとする．ここでS_jは隣接する，上流側，すなわち入力層に近い方の層にあるニューロンの状態でJ_{ij}はシナプス結合とよばれる．$f(h)$は活性化関数とよばれる．このDNN（このさき機械とよぶ）は多くの調節可能なシナプス結合J_{ij}をもち，これを調節してデザインできる機械の全体集合をΩ_0としよう．

統計力学的には次のような問いが立つ．M個の異なる入出力データの組が訓練データ（境界条件）として与えられたとして，これに完全に適合する機械は，シナプス結合J_{ij}を色々変えて，何通り作ることができるか？　この「正解の集合」をΩとし，その統計力学を考えるのである．

学習の問題で重要なのは，訓練データである．人工的だがシンプルなシナリオとして，(1) ランダムな入出力データ，(2) Ω_0から無作為に選んだ一つの「教師機械」にランダムなデータを入力し，対応する出力を取り出し，この組を「生徒機械」の訓練データとする，というものがある．(1)はガラス・ジャミング系の統計力学に深く関係する．他方，(2)はいわば結晶（隠された「教師機械」）を推定する統計力学である．

DNNの構成要素として最も単純なのは，符号を取り出す関数$f(h)=\mathrm{sgn}(h)$を活性化関数とするもので，ニューロンの状態はイジング変数$S_i=\pm 1$になる．これはいわゆるパーセプトロンの一つである．単体の場合は(1)(2)のシナリオともに深く理解されている．しかしこれを多数組み合わせたDNNの理論解析は困難とされてきた．

この困難は次のように克服できる．まず，全パーセプトロンの入出力関係が満足されることを拘束条件として導入することにより，シナプス結合J_{ij}のほかにニューロンS_iも力学変数に加えることができる．これによって，入力と出力を多段階の非線形写像で結ぶ問題が，局所的な相互作用をもつ多体系の統計力学として捉え直される．

得られた系には入出力層以外にランダムネスはない．ここで重要なヒントとなるのは，無限大次元の剛体球ガラスなど，近年急速に発展したガラス・ジャミング系の平均場理論である．そこではハミルトニアンにランダムネスがない系に対してもスピングラスなどランダム系で用いられたレプリカ法が強力なツールとなることが明らかになっている．

レプリカ法で理論を構成して解析した結果，熱力学極限N（幅），M（データ数）$\to\infty$で，比$\alpha=M/N$の増大とともに(1)レプリカ対称性の破れを伴うガラス転移，(2)結晶化が，ネットワークの両端から逐次的に起こって解空間Ωが狭くなること，ネットワークが十分深ければ中央部に「遊び」（液体領域）が残されることがわかった．これはある種の濡れ転移とみなせる．現実的には幅Nは有限であり，転移はクロスオーバーとなり，系は深さ方向にダイナミックスが変化する複雑な液体となる．

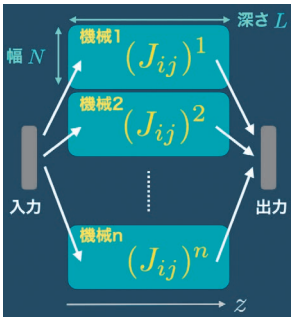
用語解説

深層ニューラルネットワーク（DNN）：
多数の層からなる層状のニューラルネットワーク．様々なタイプがあるが，ここでは多くの層からなる中間層（隠れ層）を通して，入力層から出力層まで信号が逆戻りせずに伝搬するネットワークを考える．

スピングラス：
強磁性と反強磁性相互作用がランダムに混在した磁性体．

ガラス・ジャミング系：
ガラス系とはガラス状態，すなわち乱れたパターンを（準）安定な固体状態としてもつ系である．例えば，ある箱にN個の球を入れるとして，球同士を重なりを許さない「剛体の制約を充足する」配置の位相空間Ωでの統計力学が考えられる．これは剛体球系の液体–結晶転移，結晶を押しかためた最密充填とともに，過冷却状態でのガラス転移，ガラスを押し固めたジャミング（ランダム充填）を研究する舞台である．

レプリカ法：
ガラス系の統計力学における代表的な理論手法．DNNの問題では，下図のように，同じ訓練データのもとで学習している複数の機械をレプリカとする．これらを互いに比べることによって，どの程度似通った機械になっているのか，またその深さ方向zでの変化を探ることができる．



レプリカ法の概念図．

濡れ転移：
例えば気相と液層が共存するとき，温度の低下とともに壁の表面から液相が厚みを増していく現象．