

博士論文

統計系配位の深層学習による相転移温度推定

金沢大学大学院自然科学研究科
数物科学専攻

学籍番号	1624012006
氏名	藤田 達大
主任指導教員名	青木 健一
提出年月	2019年 1月

目次

第1章 導入	3
第2章 分類・推定問題	5
2.1 機械学習の基礎	5
2.1.1 関数近似	5
2.1.2 統計的推定	5
2.2 クラス分類	6
2.2.1 ロジスティック回帰	7
2.2.2 ソフトマックス回帰	8
2.3 深層学習	9
2.3.1 パーセプトロン	9
2.3.2 ニューラルネットワーク	9
2.4 畳み込みニューラルネットワーク (CNN)	10
2.4.1 チャンネル	11
2.4.2 畳み込み層	11
2.4.3 プーリング層	12
第3章 1次元イジングモデルの確率分布と配位生成	13
3.1 1次元最近接イジングモデルの配位生成と確率分布	13
3.2 1次元長距離イジングモデル	14
3.3 ボルツマンマシン	15
3.4 制限ボルツマンマシンとくりこみ群	15
3.5 厳密な制限ボルツマンマシン (Exact RBM)	17
第4章 CNNを用いた1次元長距離イジングモデルの温度推定	20
4.1 訓練データ	20
4.2 機械の構成	20
第5章 1次元長距離イジングモデルの温度推定	24
5.1 最尤推定	24
5.2 1次元イジングモデル	24
5.3 1次元長距離イジングモデル	25
5.4 $N\alpha$ Configuration	26
5.5 混合サンプルにおける温度推定	27

第 6 章	機械は何を学んでいるか？	30
6.1	ハミルトニアン認識率 (HRR)	30
6.2	共分散不等式	31
第 7 章	相転移の特異性	32
第 8 章	まとめ	36
第 9 章	Appendix	37
Appendix		37
9.1	peierls の議論	37
9.2	Block Decimation くりこみ群	39
9.2.1	1次元イジングモデル	39
9.2.2	1次元長距離イジングモデル	40
9.2.3	有限レンジスケーリング仮説	44
9.3	勾配法について	46
9.3.1	誤差逆伝播法	46
9.3.2	勾配消失問題・勾配爆発問題	47
9.3.3	勾配降下法	49
9.3.4	勾配降下法の問題点	50
9.3.5	モーメントム法	51
9.3.6	AdaGrad	51
9.3.7	Adam	51
9.4	Kadanoff 形式	52

第1章 導入

深層学習は機械学習（人工知能）の分野で大きく発展している。その技術は様々な分野で注目を集め、物理学の分野でもその影響を受けており、素粒子・物性物理問わず様々な応用がなされている [1][2][3][4][5][6][7][8][9][10]。

深層学習は成功は収めているが、なぜ上手くいくか十分に理解されていないところがある。その一方では、深層学習の学習過程は大量の入力データから、重要なデータを残し、不要な情報は捨てることで情報縮約を行っていると考えられる。物理学者にとってどこか見覚えのある光景であり、言い換えると、ミクロな情報から重要なマクロの情報を抽出する「くりこみ群」である。深層学習とくりこみ群に関しては、[2][3][11][12]の先行研究で言及されている。また深層学習を用いて相転移温度を検出する議論がなされている [2][6][12][13]。

制限ボルツマンマシンと呼ばれる深層学習の手法とくりこみ群の対応関係を考察している先行研究 [11] がある。この手法は、現実の物理量などのデータが入力される可視変数 v_i からなる可視層、モデルの表現能力を向上させる隠れ変数 h_i からなる隠れ層の2種類の層がある。これらの層の変数同士はパラメータ ω_{ij} で結合される。この可視変数に対し画素データやスピン配位などを入力データとして与え、パラメータ ω_{ij} を最適化することにより、入力データに対応する確率分布を与える。現実の物理量 v_i からより少ない隠れ変数 h_i への書き換えは、重要な情報の抽出に対応しており、くりこみ群に対応していると考えられる。

他にくりこみ群に類似している深層学習の方法として、画像認識で用いられる畳み込みニューラルネットワークと呼ばれる手法がある。この手法では、画像データなどを入力とし、それを各ブロックに分けてブロックごとに畳み込みという情報縮約を行う。出力側で元の画像データが何であったかを識別する操作を行い、その結果が入力と一致するように畳み込みに用いたパラメータを最適化していく。この手法は情報縮約という点でブロックスピン変換にとっても類似しており、畳み込みの操作によって相転移点が推定されると予想される。先行研究 [12] では、画像データを2次元イジングモデルのスピン配位に置き換え、相転移温度の検出が行われている。ただし機械に教えていることは配位とその温度のみであ

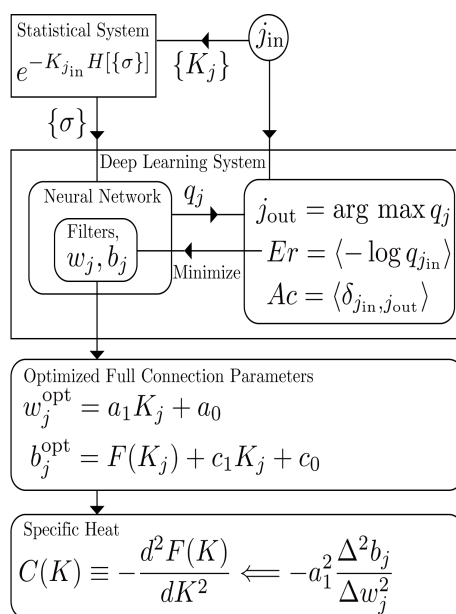


図 1.1: 概要

り、機械はパラメーターに相転移温度および秩序変数の情報を残していると主張している。くりこみ群の類似性はあるが、なぜ相転移温度が検出されるのか疑問が残ったままである。

本論文では深層学習の方法を用いて統計系の配位を入力データとし温度推定を行い、相転移温度の推定について議論した。結論を述べると最適化された機械は、統計系の配位から温度を推定する過程において機械内部にエネルギー関数を構成している。つまりエネルギーを学んでいる。温度の分類（推定）器であるソフトマックス関数の重みは温度の情報をもち、バイアスは自由エネルギーの情報を持つ。相転移の情報は比熱の特異性から得られるとすると、バイアスを重みで2階差分をとったものである。よって統計系の配位から温度推定を行い、相転移温度を検出したいならば、最適化された機械から比熱の情報で判断する必要がある。

本論文の構成について述べる。2章では本論文で必要となる機械学習の手法について簡単に説明する。3章では1次元長距離イジングモデルの確率分布および入力データとなる配位集合の生成法について述べる。1次元最近接イジングモデルは厳密解やpeierlsの議論からわかるように有限温度では相転移が起こらないが、長距離相互作用がある場合は、相転移が起こるとされている。また1次元長距離イジングモデルは有限サイズ（有限体積）・有限レンジでは解析的に解け分配関数を評価でき、また厳密なくりこみ群を実行することができる。その結果は機械学習の結果を比べる際に重要な役割を果たし、また5章で説明する温度推定の理論を考える際に、厳密くりこみ群から得られた確率分布の情報が必要となる。4章で畳み込みニューラルネットワークを用いた温度推定の実験結果を述べる。5章では温度推定の理論について述べる。6章では機械学習は何を学んで温度推定しているのか、7章では相転移温度の推定について述べる。8章は本論文のまとめを行う。

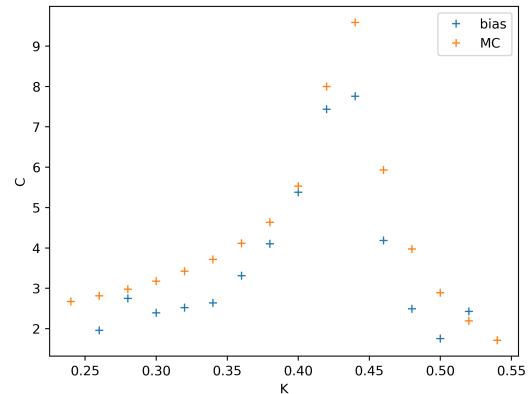


図 1.2: 比熱の特異性 (2次元イジングモデル)

第2章 分類・推定問題

2.1 機械学習の基礎

機械学習は大きく分けて教師あり学習と教師なし学習の2つの学習方法がある。分類・推定問題を考える場合は教師あり学習がよく使われ、ほとんどの教師あり学習は関数近似と統計的推定の2つを組み合わせられて使われる。

2.1.1 関数近似

教師あり学習について考えると教師ありのデータの集合 D は、入力データ x と出力データ y の N セットとなる。

$$D = \{(x_n, y_n)\}_{n=1 \dots N} \quad (2.1.1)$$

機械学習の目的は、未知のデータ x に対して適切な y を推定することを目的とする。適切な y を推定するため、関数モデル（関数近似）を仮定して、関数に含まれるパラメータを訓練データから最適化する。また入力 x から出力 y を決定するために、関数 $y = y(x)$ を求めることを回帰と呼ばれる。本論文では関数モデルはニューラルネットワークを用いて構成する。

2.1.2 統計的推定

教師あり学習および教師なし学習どちらの場合でも統計的推定は用いられる。データ（標本）とは、母集団から抽出されたとする。母集団はデータの生成分布 $P_{\text{data}}(x)$ によって特徴付けられているとする。データを解析することにより母集団の知識を得ることを目標とする。生成分布 $P_{\text{data}}(x)$ がわかれば母集団の知識を知ることができるが、生成分布 $P_{\text{data}}(x)$ は一般的に知られていない。よって生成分布をパラメータ ω を用いて

$$P_{\text{data}}(x) \approx P(x; \omega) \quad (2.1.2)$$

と近似する。データから最適なパラメータ ω^* を推定し、モデル分布 $P(x; \omega^*)$ を求める方法を統計的推定と呼ぶ。

データの集合 $D = \{x_1, x_2, \dots, x_n\}$ を用いて尤度関数を以下のように定義する。

$$L(\omega) = \prod_{n=1}^N P(x_n; \omega) \quad (2.1.3)$$

データ D が生成された理由は、パラメータ ω が確率を最大にするようなものであったとする。つまり尤度関数が最大となるようなパラメータ ω が、データ D を実現されやすいと考える。尤もらしいパラメータ ω^* は尤度関数を最大化することで求めることができる。

$$\omega^* = \arg \max_{\omega} L(\omega) \quad (2.1.4)$$

尤度関数を最大化する際、便利上対数をとることがある。

パラメータ ω を決める他の方法として、カルバック・ライブラー情報量がある。カルバック・ライブラー情報量は、私たちが知りたい確率分布を $p(x)$ とすると、確率分布 $p(x)$ と $q(x; \omega)$ の離れ具合を示す量として以下のように定義される。

$$\begin{aligned} D_{\text{KL}}(p||q) &= \int dx p(x) \log \left(\frac{p(x)}{q(x; \omega)} \right) \\ &= - \int dx p(x) \log q(x; \omega) + \int dx p(x) \log p(x) \end{aligned} \quad (2.1.5)$$

ω は学習によって定めるパラメータであり、確率分布 $q(x; \omega)$ を $p(x)$ に近づけるためには、カルバック・ライブラー情報量または交差エントロピー

$$S_{\text{cross}} = - \int dx p(x) \log q(x; \omega) \quad (2.1.6)$$

を最小にするようにパラメータ ω を定める。

2.2 クラス分類

クラス分類とは、入力 x を K 個の種類（クラス）に分類することである。 K 個のクラス C_1, C_2, \dots, C_K に関して、出力 y を離散変数として扱う。 y の値によって決まる K 成分のベクトルを

$$t(y) = (t(y)_1 \ t(y)_2 \ \dots \ t(y)_K) \quad (2.2.1)$$

と表現する。この表現を 1-of-K 符号化と言う。one-hot 表現は、例えば 1 番目のクラスに y が属する場合

$$t(y=1) = (1 \ 0 \ \dots \ 0) \quad (2.2.2)$$

クロネッカーデルタを用いることで

$$t(y)_k = \delta_{y,k} \quad (2.2.3)$$

と書くことができる。

2.2.1 ロジスティック回帰

最初に2クラスの分類について考える。入力 x に関してクラス C に分類される確率 $P(C|x)$ とかく。ここでシグモイド関数を以下のように定義する。

$$\sigma(u) \equiv \frac{1}{1 + e^{-u}} \quad (2.2.4)$$

クラス C に分類される確率 $P(C|x)$ はシグモイド関数を用いて

$$P(C|x) = \sigma(u) \quad (2.2.5)$$

と書くことができる。 u は対数オッズ比とよばれる。

$$u = \log \left(\frac{\sigma(u)}{1 - \sigma(u)} \right) = \log \left(\frac{P(C_1|x)}{P(C_2|x)} \right) \quad (2.2.6)$$

対数オッズ比が1を越えると C_1 に分類される可能性が C_2 より高いことが言える。つまり対数オッズ比 u が1を越えると C_1 に分類され、それ以外だと C_2 に分類される。

対数オッズ比が入力 x に対して線形関数であることを仮定した確率モデル

$$u = \omega x + b \quad (2.2.7)$$

をロジスティック回帰と呼ばれ ω は重み、 b はバイアスと呼ばれている。

また、クラス C に分類される確率 $P(C|x)$ は

$$P(C_1|x) + P(C_2|x) = 1 \quad (2.2.8)$$

の条件を満たし、さらにベルヌーイ分布なので確率分布は

$$P(y|x) = (P(C_1|x))^y (1 - P(C_1|x))^{1-y} \quad (2.2.9)$$

とかける。 ω を最尤法を用いて決める。データ $D = \{(x_l, y_l)\}_{l=1 \dots n}$ を用いて、尤度関数を定義する。

$$L(\omega) = \prod_{n=1}^N (P(C_1|x_n))^{y_n} (1 - P(C_1|x_n))^{1-y_n} \quad (2.2.10)$$

クロスエントロピーは尤度関数を用いて

$$E(\omega) = -\log L(\omega) = -\sum_{n=1}^N \{y_n \log P(C_1|x_n) + (1 - y_n) \log (1 - P(C_1|x_n))\} \quad (2.2.11)$$

と定義する。クロスエントロピーを最小化すること、つまり尤度関数を最大化することで、最適なパラメータ ω および b を求めることができる。また実際の計算ではサンプル数で平均をとる。

2.2.2 ソフトマックス回帰

次に多クラス分類について考える。ロジスティック回帰を一般化することができ、one-hot 表現を用いて確率分布を

$$P(y|x) = \prod_{k=1}^K (P(C_k|x))^{t(y)_k} \quad (2.2.12)$$

と書く。これは2クラスの場合の確率分布であるベルヌーイ分布の、多クラスへの一般化でありマルチヌーイ分布と呼ばれる。 $\sum_k^K t(y)_k = 1$ を用いて式(2.2.12)を変形すると

$$\begin{aligned} P(y|x) &= \prod_{k=1}^{K-1} (P(C_k|x))^{t(y)_k} (P(C_K|x))^{1-\sum_k^{K-1} t(y)_k} \\ &= (P(C_K|x)) \left(\frac{P(C_k|x)}{P(C_K|x)} \right)^{\sum_k^{K-1} t(y)_k} \\ &= (P(C_K|x)) (e^{u_k})^{\sum_k^{K-1} t(y)_k} \end{aligned} \quad (2.2.13)$$

ここで対数オッズ比 u_k を

$$u_k \equiv \log \left(\frac{P(C_k|x)}{P(C_K|x)} \right) \quad (2.2.14)$$

と定義した。 $\sum_k P(C_k|x) = 1$ の条件を用いると

$$\sum_k e^{u_k} = \frac{1}{P(C_K|x)} \quad (2.2.15)$$

である。ソフトマックス関数を

$$\text{softmax}_k(u_1, u_2, \dots, u_K) \equiv \frac{e^{u_k}}{\sum_l e^{u_l}} \quad (2.2.16)$$

と定義するとクラス C_k に分類される確率 $P(C_k|x)$ は

$$P(C_k|x) = \text{softmax}_k(u_1, u_2, \dots, u_K) \quad (2.2.17)$$

とかける。対数オッズ比が入力 x に対して線形関数であることを仮定した確率モデル

$$u_k = \omega_k x + b_k \quad (2.2.18)$$

をソフトマックス回帰と呼ばれる。パラメータ ω_k および b_k は、クロスエントロピーを最小化することで求めることができる。クロスエントロピーは

$$\begin{aligned} E(\omega) &= -\log L(\omega) \\ &= -\log \left\{ \prod_{n=1}^N \prod_{k=1}^K P(C_k|x_n)^{t(y_n)_k} \right\} \\ &= -\sum_{n=1}^N \sum_{k=1}^K \left\{ t(y_n)_k \log \left(\text{softmax}_k(u_1, u_2, \dots, u_K) \right) \right\} \end{aligned} \quad (2.2.19)$$

である。

2.3 深層学習

人間の脳は1000億個以上のニューロン（神経細胞）が集まって構成される。ニューロンの末端にはシナプスと呼ばれる部位があり、他のニューロンから情報を受け取っている。ニューロンの核の部分には、細胞体と呼ばれる部位がある。シナプスから受け取った情報は、この細胞体に流れていく。細胞体はある一定以上の値（閾値）を超えると反応することが知られており、反応すると他のニューロンに情報を流す仕組みとなっている [14]。機械学習および深層学習では、人間の脳を真似て作られた関数モデルつまりニューラルネットワークを用いた解析がよく行われる。

2.3.1 パーセプトロン

以上の説明を元に、脳の構造に類似させた模型を説明する。ニューロンの情報は、onとoffのみとする。ニューロン同士の結合の強さの度合いは異なっていることから、シナプスの結合を表す重み ω を導入して

$$u = \sum_i \omega_i x_i \quad (2.3.1)$$

と書く。 u は細胞体に入る情報に相当する。ニューロンの出力 z は、活性化関数

$$\theta(x+b) = \begin{cases} 1 & (x+b > 0) \\ 0 & (x+b \leq 0) \end{cases}$$

を用いることで、閾値を b として

$$z = \theta(u+b) \quad (2.3.2)$$

と書くことができる。活性化関数は細胞体の反応に相当する。このモデルを形式ニューロンと言う。形式ニューロンを複数組み合わせさせた模型をパーセプトロンと言い、層構造をなすように構成できる。

2.3.2 ニューラルネットワーク

形式ニューロンまたパーセプトロンは、離散値しかとれないが実数値を取れるようにしたもののがニューラルネットワークである。活性化関数を連続値が取れるよう、また階段関数のように振る舞う関数を採用する。

ここで層構造をなすニューラルネットワークの関数系を書き下す。入力データ $\mathbf{x} = (x_1, x_2, \dots, x_n)$ を

$$\mathbf{z}^{(1)} = \mathbf{x}, z_i^{(1)} = x_i \quad (2.3.3)$$

と表すことにする。 l 層目の出力は

$$z_j^{(l)} = f^{(l)}\left(u_j^{(l)} + b_j^{(l)}\right), u_j^{(l)} = \sum_i \omega_{ji}^{(l)} z_i^{(l-1)} \quad (2.3.4)$$

と書くことができる。 $f^{(l)}$ 、 $\omega^{(l)}$ 、 $b^{(l)}$ は l 層目の活性化関数、重み、バイアスである。 $z^{(l-1)}$ は $l-1$ 層目の出力である。表記を簡単にするため

$$z_0^{(l)} = 1, \omega_{j0}^{(l)} = b_j^{(l)} \quad (2.3.5)$$

とし、重み、バイアスを行列、ベクトル表記にすると

$$\mathbf{z}^{(l)} = f^{(l)}(\mathbf{u}^{(l)}), \mathbf{u}^{(l)} = \mathbf{W}^{(l)}\mathbf{z}^{(l-1)} \quad (2.3.6)$$

とかける。 L 層目を出力層とすると推定量 \hat{y} とすると

$$\hat{y} = \mathbf{z}^{(L)} = f^{(L)}(\mathbf{u}^{(L)}), \mathbf{u}^{(L)} = \mathbf{W}^{(L)}\mathbf{z}^{(L-1)} \quad (2.3.7)$$

となる。よって L 層のニューラルネットワークにおける関数近似モデルは

$$\hat{y}(\mathbf{W}, \mathbf{x}) = f^{(L)}(\mathbf{W}^{(L)}f^{(L-1)}(\mathbf{W}^{(L-1)}f^{(L-2)}(\dots\mathbf{W}^{(2)}f^{(1)}(\mathbf{W}^{(1)}\mathbf{x}))) \quad (2.3.8)$$

となる。パラメータ ω の最適化について考える。訓練データ $D = \{(x_n, y_n)\}_{n=1, \dots, n}$ を用意する。入力データ \mathbf{x}_n を入れた際の出力 $\hat{y}(\mathbf{W}, \mathbf{x}_n)$ が、出力データ \mathbf{y}_n との誤差が小さくなるように最適化する。誤差を表す適切な関数 $E(\omega)$ を選び、誤差関数を最小化することで

$$\omega^* = \arg \min_{\omega} E(\omega) \quad (2.3.9)$$

最適なパラメータ ω^* が求まる。

次にニューラルネットワークを用いた多クラス分類については最終出力をソフトマックス関数に書き換えれば多クラス問題を考えることができる。入力データ x が $y = k$ 番目のクラスに属する確率 $P(y = k|x)$ を推定するものとする。つまり \hat{y}_k が最大となる k を判定する。

$$\hat{y}_k(\mathbf{W}, \mathbf{x}) = P(y = k|\mathbf{W}, \mathbf{x}) = \text{softmax}_k(u_1^{(L)}, u_2^{(L)}, \dots, u_K^{(L)}) \quad (2.3.10)$$

ニューラルネットワークを用いた場合も同様に誤差関数は、クロスエントロピー

$$\begin{aligned} E(\omega) &= - \sum_{n=1}^N \sum_{k=1}^K \{t(y_n)_k \log(\hat{y}_k(\mathbf{W}, \mathbf{x}_n))\} \\ &= - \sum_{n=1}^N \sum_{k=1}^K \left\{ t(y_n)_k \log \left(\text{softmax}_k(u_1, u_2, \dots, u_K) \right) \right\} \end{aligned} \quad (2.3.11)$$

を用いる。

2.4 畳み込みニューラルネットワーク (CNN)

人間は視覚情報からパターンを認識する機能が備わっており、パターンを認識する際、平行移動・拡大縮小・回転などズレを考慮してパターンを認識する。つまりパターン認識の機械学習は、このような変換のもと不変であるような機械を構成する必要が

ある。また画像などからパターンを認識する際、隣同士の画素は遠くに離れている画素よりも強い相関を持っており、局所的でなければならない [15]。

人間は視覚情報を得る視覚野には単純型細胞と複雑型細胞の大きく2つの種類に分かれることが知られている。受容野と呼ばれる部位があり、受容野はニューロンが反応するような領域である。単純型細胞は受容野が狭く、局所的である。複雑型細胞は受容野が広く、パターンのズレに対して頑健性をもつことが知られている。この2つの細胞により、特定の位置にあるパターンと、位置にかかわらず存在するパターンを認識することができる。

ヒューベルトウィーゼルの階層仮説があり、複数の単純型細胞が束ねられ複雑型細胞に結合しているとする。この構造により、局所的なパターンは単純型細胞が検知し、その情報を統合する事で複雑型細胞が位置のズレに対してパターンを認識できるとされる。この仮説に則り、構成された機械学習（深層学習）を畳み込みニューラルネットワークと呼ぶ [14][16][17]。

2.4.1 チャネル

画像データの読み込む事を前提に、畳み込みニューラルネットワークについて説明する。画像データを1次元の配列に書き直すのではなく、2次元の配列を用いて表す。画像の位置 (i, j) とした時の画素値 x_{ij} と表す。画像データは色、または奥行き方向の情報を持っており、その自由度をチャネルと呼ぶ。入力データは $x_{ijk} = z_{ijk}^{(1)}$ となる。つまり1つの画像に対してチャネル数を K 個とすると、 K 枚の画像が出てくることとなる。

2.4.2 畳み込み層

畳み込み層は単純型細胞に対応している箇所である。画像サイズより小さいサイズの $H \times H$ フィルターを用いて畳み込みを行う。フィルターの画素値を h_{pq} とすると畳み込みの操作は、入力 z に対して

$$u_{ij} = \sum_{p,q=0}^{H-1} h_{pq} z_{i+p,j+q} \quad (2.4.1)$$

と書くことができる。入力の画像サイズ $W \times W$ とすると、畳み込みによって $(W - H + 1) \times (W - H + 1)$ のサイズに畳み込まれる。 K 個のチャネルにおける l 層目の出力は、活性化関数を用いて

$$\begin{aligned} z_{ij}^{(l)} &= f(u_{ij}^{(l)}) \\ u_{ij}^{(l)} &= \sum_{k=0}^{K-1} \sum_{p,q=0}^{H-1} h_{pqk} z_{i+p,j+q,k}^{(l-1)} + b_{ij}^{(l)} \end{aligned} \quad (2.4.2)$$

となる。畳み込み層からの出力を特徴マップと呼ばれる。

以上の畳み込みの方法はフィルターを画素1つ分ずらしていく方法である。ストライドの方法を用いることで、画素を S 個分ずらしていくことができる。

$$u_{ij}^{(l)} = \sum_{k=0}^K \sum_{p,q=0}^{H-1} h_{pqk} z_{Si+p, Sj+q, k}^{(l-1)} + b_{ij}^{(l)} \quad (2.4.3)$$

畳み込みによって $(\lfloor \frac{W-H}{S} \rfloor + 1) \times (\lfloor \frac{W-H}{S} \rfloor + 1)$ のサイズになる。

2.4.3 プーリング層

プーリング層は、複雑型細胞に対応している箇所である。畳み込みによって得られた局所的な情報を、入力的位置がずれても情報を捉えることが目的である。不変でありたい領域の添え字を縮約すれば、位置に対して不変であると言える。不変でありたい領域 P_{ij} とすると、最大プーリング

$$u_{ij}^{(l)} = \max_{(p,q) \in P_{ij}} z_{p,q,k}^{(l)} \quad (2.4.4)$$

平均プーリング

$$u_{ij}^{(l)} = \text{average}_{(p,q) \in P_{ij}} z_{p,q,k}^{(l)} \quad (2.4.5)$$

と呼ばれるプーリングの方法がある。

第3章 1次元イジングモデルの確率分布と配位生成

3.1 1次元最近接イジングモデルの配位生成と確率分布

1次元最近接イジングモデルのハミルトニアンを

$$H = -K \sum_i d_i, \quad d_i = \sigma_i \sigma_{i+1} \quad (3.1.1)$$

と書き換える。変数 d_i が 1 の場合は $\sigma_i = \sigma_{i+1}$ なのでドメインウォールがなく、 d_i が -1 の場合はドメインウォールがある。よってドメインウォールが現れる確率は

$$p(K) = \frac{\exp(-K)}{\exp(-K) + \exp(K)} \quad (3.1.2)$$

である。つまり配位の端のスピン向きが決まれば、確率分布 (3.1.2) に従って隣のスピン向きを決めることができる。これを体積の大きさ N とすると、 N 回くりかえせば 1次元イジングモデルの配位が求まる。マルコフ連鎖モンテカルロ法ではマルコフ連鎖の特性により、一個前に生成された配位と強い相関（自己相関）を持ってしまう。ドメインウォール生成確率を用いた配位生成法では自己相関がなく高速な配位集合を生成できる。

1次元最近接イジングモデルの確率分布はドメインウォールを用いると、具体的に書き下すことができ、 n 個のドメインウォールが生成される配位の確率は

$$P(n) = {}_N C_n p(K)^n (1 - p(K))^{(N-n)} \quad (3.1.3)$$

となる。 N は体積である。確率分布は図 3.1 となる。

また、オペレータ部分はドメインウォールを用いて計算することができ

$$\hat{O} = \sum_n \sum_i \sigma_i \sigma_{i+1} = (N - n) - n = N - 2n \quad (3.1.4)$$

その期待値は、

$$\langle \hat{O} \rangle |_{K_i} = N - 2\langle n \rangle |_{K_i} \quad (3.1.5)$$

である。また n 個のドメインウォールが生成される配位の確率は、二項分布となっており平均と分散は以下の通りである。

$$\mu = Np(K), \quad \sigma^2 = Np(K)(1 - p(K)) \quad (3.1.6)$$

しかし外場がある場合また 1次元長距離イジングモデルの場合には、確率分布を (3.1.2) のように書くことができないため適用できない。

3.2 1次元長距離イジングモデル

次に1次元長距離イジングモデルの場合を考える。1次元長距離イジングモデルのハミルトニアンは以下の通りである。

$$H = -K \sum_n \sum_i \left(\frac{1}{n^p} \right) \sigma_i \sigma_{i+n} \quad (3.2.1)$$

今後の議論では $p = 1.8$ 、 $n = 8$ とする。最近接相互作用の場合ではドメインウォールを考えることで、確率分布を求めることができたが、1次元長距離イジングモデルは長距離の相互作用があるため上記の方法で確率分布を求めることはできない。しかしBDRGの手法を用いれば自由エネルギーを求めることができ、その結果を用いて平均と分散を求め、正規分布で近似した確率分布を求めることはできる。

自由エネルギーを微分すれば平均と分散がもとまり

$$\mu_i = \frac{\partial F}{\partial K} \Big|_{K_i} = \frac{1}{N} \langle \hat{O} \rangle \Big|_i \quad (3.2.2)$$

$$\sigma_i^2 = \frac{1}{N} \frac{\partial^2 F}{\partial K^2} \Big|_{K_i} = \frac{1}{N^2} \left\langle \left(\hat{O} - \langle \hat{O} \rangle \right)^2 \right\rangle \Big|_i \quad (3.2.3)$$

温度 K_i における確率分布は

$$P(E, K_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} \int_{-\infty}^{\infty} \exp \left\{ -\frac{(E - \mu_i)^2}{2\sigma_i^2} \right\} dE \quad (3.2.4)$$

となる。確率分布は図 3.2 となる。

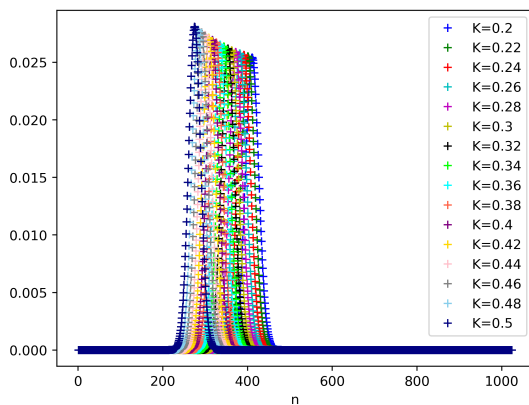


図 3.1: 二項分布

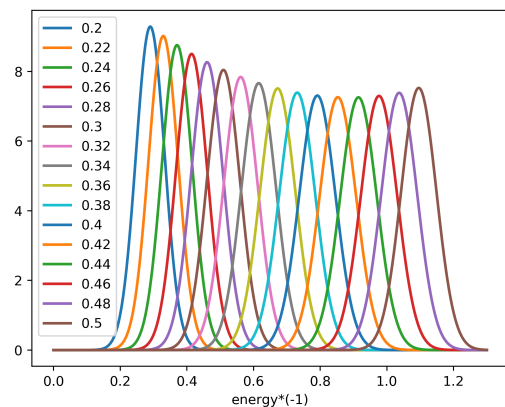


図 3.2: 正規分布近似

3.3 ボルツマンマシン

ボルツマンマシンは教師なし学習でよく使われ、ボルツマンマシンの目的としてはデータから未知の確率分布である生成分布を推論することである。ボルツマンマシンのエネルギー関数は、重みおよびバイアスを $\theta = \{b_j, \omega_{ij}\}$ とすると

$$E(\mathbf{v}, \theta) = - \sum_{i,j} \omega_{ij} v_i v_j - \sum_i b_i v_i \quad (3.3.1)$$

と表せ、確率分布は

$$p(\mathbf{v}; \theta) = \frac{1}{Z(\theta)} e^{-E(\mathbf{v}, \theta)}, \quad Z(\theta) = \sum_{\mathbf{v}} e^{-E(\mathbf{v}, \theta)} \quad (3.3.2)$$

である [14]。

一般的に機械学習のモデルでは、目的関数つまり誤差関数・対数尤度関数は非凸関数のため、最小値または最大値に収束することは保証されていない。しかしボルツマンマシンはヘルダーの不等式を用いることで対数尤度関数は凸であることが証明できる。よって理論的には勾配上昇法は必ず最大値に収束することが保証されているモデルである。

ボルツマンマシンの表現を向上させるために隠れ変数を導入したモデルを制限ボルツマンマシンという [14]。制限ボルツマンマシンのエネルギー関数は、可視変数 v および隠れ変数 h およびパラメータ $\theta = \{b_i, c_j, \omega_{ij}\}$ とすると

$$E(\mathbf{v}, \mathbf{h}, \theta) = - \sum_{i,j} \omega_{ij} v_i h_j - \sum_i b_i v_i - \sum_j c_j h_j \quad (3.3.3)$$

と表せ、確率分布は

$$p(\mathbf{v}, \mathbf{h}; \theta) = \frac{1}{Z(\theta)} e^{-E(\mathbf{v}, \mathbf{h}, \theta)}, \quad Z(\theta) = \sum_{\mathbf{v}} \sum_{\mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h}, \theta)} \quad (3.3.4)$$

また隠れ変数 h と結合する隠れ変数 h_1 を導入して層を増やしたモデルを深層ボルツマンマシンと言う。制限ボルツマンマシンの特徴として、可視変数および隠れ変数同士の結合がないことである。このような制限を用いることで可視変数および隠れ変数の条件付き確率分布が求まり、コントラストダイバージェンス法を用いてパラメータ更新が可能である。ただし隠れ変数を導入したため対数尤度関数の凸性は保証されない。

3.4 制限ボルツマンマシンとくりこみ群

制限ボルツマンマシンとくりこみ群について議論されており、この節ではそれについて紹介する [11]。制限ボルツマンマシンのエネルギー関数は可視変数と隠れ変数を用いて

$$E(\mathbf{v}, \mathbf{h}, \theta) = - \sum_{i,j} \omega_{ij} v_i h_j - \sum_i c_i v_i - \sum_j b_j h_j \quad (3.4.1)$$

である。パラメータ $\theta = \{b_j, c_i, \omega_{ij}\}$ とすると、確率分布は

$$p(\mathbf{v}, \mathbf{h}; \theta) = \frac{1}{Z} e^{-E(\mathbf{v}, \mathbf{h})}, Z(\theta) = \sum_{\mathbf{v}, \mathbf{h}} e^{-E(\mathbf{v}, \mathbf{h})} \quad (3.4.2)$$

と表せる。ここで便利上、可視変数の確率分布と隠れ変数の確率分布を以下の形で表すことにする。

$$\begin{aligned} p(\mathbf{v}; \theta) &= \sum_{\mathbf{h}} \frac{1}{Z} e^{-E(\mathbf{v}, \mathbf{h}, \theta)} = \frac{1}{Z} e^{-H^{RBM}(\mathbf{v}, \theta)} \\ p(\mathbf{h}; \theta) &= \sum_{\mathbf{v}} \frac{1}{Z} e^{-E(\mathbf{v}, \mathbf{h}, \theta)} = \frac{1}{Z} e^{-H^{RBM}(\mathbf{h}, \theta)} \end{aligned} \quad (3.4.3)$$

パラメータ θ を最適化するため、Kullback-Leibler(KL) divergence を用いる。

$$D_{KL}(P(\mathbf{v})||p(\mathbf{v}, \theta)) = \sum_{\mathbf{v}} P(\mathbf{v}) \log \left(\frac{P(\mathbf{v})}{p(\mathbf{v}, \theta)} \right) \quad (3.4.4)$$

$P(\mathbf{v})$ は本来の確率分布（生成モデル）である。制限ボルツマンマシンが厳密に本来の確率分布に従う場合

$$D_{KL}(P(\mathbf{v})||p(\mathbf{v}, \theta)) = 0 \quad (3.4.5)$$

である。つまり KL divergence が最小となるようにパラメータ θ を決める。

制限ボルツマンマシンとくりこみ群との対応関係は、Kadanoff 形式を用いて議論されている。くりこみ変換関数を

$$T(\mathbf{v}, \mathbf{h}) = -E(\mathbf{v}, \mathbf{h}) + H(\mathbf{v}) \quad (3.4.6)$$

とする。ここで $H(\mathbf{v})$ は本来の確率分布のハミルトニアンである。くりこみ変換した新しいハミルトニアンは

$$\begin{aligned} \frac{1}{Z} e^{-H^{RG}(\mathbf{h})} &= \sum_{\mathbf{v}} \frac{1}{Z} e^{T(\mathbf{v}, \mathbf{h}) - H(\mathbf{v})} \\ &= \sum_{\mathbf{v}} \frac{1}{Z} e^{-E(\mathbf{v}, \mathbf{h})} = p(\mathbf{h}) \end{aligned} \quad (3.4.7)$$

と書き表すことができ、

$$H^{RG}(\mathbf{h}) = H^{RBM}(\mathbf{h}) \quad (3.4.8)$$

と結論づけることができる。 $H^{RBM}(\mathbf{h})$ は可視変数部分を積分した有効ハミルトニアンはくりこまれたハミルトニアンと解釈できる。Kadanoff 形式で課している Tr 条件について考えると、くりこみ変換関数は

$$\begin{aligned} e^{T(\mathbf{v}, \mathbf{h})} &= e^{-E(\mathbf{v}, \mathbf{h}) + H(\mathbf{v})} \\ &= \frac{p(\mathbf{v}, \mathbf{h})}{p(\mathbf{v})} e^{-H^{RBM}(\mathbf{v}) + H(\mathbf{v})} \end{aligned} \quad (3.4.9)$$

と変形でき、Tr 条件をとることで

$$H(\mathbf{v}) = H^{RBM}(\mathbf{v}) \quad (3.4.10)$$

となる。この結果は $p(\mathbf{v}) = P(\mathbf{v})$ 、つまり Tr 条件は $D_{KL}(P(\mathbf{v})||p(\mathbf{v})) = 0$ を再現していると言える。

3.5 厳密な制限ボルツマンマシン (Exact RBM)

ボルツマンマシンの学習パラメータ θ は学習によって定めるが、くりこみ群の結果を利用することで学習パラメータ θ を定めることができる。くりこみ群とは逆の（確率）過程となるスピン配位の生成法 (Exact RBM) について述べる。

周期境界条件を課し体積は 2^l とする。制限ボルツマンマシンのエネルギー関数を

$$H(\mathbf{v}, \mathbf{h}, \theta) = - \lim_{\omega \rightarrow \infty} \sum_{j:\text{even}} [K^{(0)} h_j (v_{j-1} + v_{j+1}) + \omega h_j v_j] \quad (3.5.1)$$

とする。 $\omega \rightarrow \infty$ は低温極限 $T \rightarrow 0$ に対応しており、これはスピンの揃うことつまり $v_j = h_j$ であることを意味する。可視層と1層目の隠れ層の確率分布関数は

$$P(\mathbf{v}, \mathbf{h}; \theta) = \frac{1}{Z(\theta)} \prod_{j:\text{even}} \delta(h_j, v_j) \exp \{ K^{(0)} h_j (v_{j-1} + v_{j+1}) \} \quad (3.5.2)$$

$$Z(\theta) = \sum_{\mathbf{v}} \sum_{\mathbf{h}} e^{-H(\mathbf{v}, \mathbf{h}, \theta)}$$

となる。可視層の変数の数は 2^l 、隠れ層の変数の数は 2^{l-1} とする。 h について積分してしまうと

$$\sum_{\mathbf{h}} P(\mathbf{v}, \mathbf{h}; \theta) = \frac{1}{Z(\theta)} \exp \left\{ \sum_j K^{(0)} v_j v_{j+1} \right\}, \quad (3.5.3)$$

イジングモデルとなる。次に v について積分を行うと

$$\begin{aligned} \sum_{\mathbf{v}} P(\mathbf{v}, \mathbf{h}; \theta) &= \frac{1}{Z(\theta)} \sum_{v_k: k:\text{odd}} \exp \left\{ \sum_{j:\text{even}} K^{(0)} h_j (v_{j-1} + v_{j+1}) \right\} \\ &= \frac{1}{Z(\theta)} \prod_{j:\text{even}} 2 \cosh (K^{(0)} (h_j + h_{j+2})) \end{aligned} \quad (3.5.4)$$

この積分操作は decimation くりこみ群に対応している。 l 回くりこまれた結合定数 $K^{(l)}$ を用いると

$$\begin{aligned} \sum_{\mathbf{v}} P(\mathbf{v}, \mathbf{h}; \theta) &\propto \prod_{j:\text{even}} \exp (K^{(1)} h_j h_{j+1}) \\ K^{(1)} &= \frac{1}{2} \log (\cosh (2K^{(0)})) \end{aligned} \quad (3.5.5)$$

と書き直せる。ここでスピン v_k が生成される確率を考える。 k が偶数の場合

$$v_k = h_k \quad (3.5.6)$$

とする。 k が奇数の場合、スピン v_k が生成される相対確率は

$$\exp (K^{(0)} v_k (h_{k-1} + h_{k+1})) \quad (3.5.7)$$

であり規格化すると

$$P(v_k) = \frac{\exp (K^{(0)} v_k (h_{k-1} + h_{k+1}))}{2 \cosh \{ K^{(0)} (h_{k-1} + h_{k+1}) \}} \quad (3.5.8)$$

となる。 h_k を定めればスピン v_k が生成される確率 $P(v_k)$ が求まる。

スピン h_k が生成される確率は、ハミルトニアン (3.5.1) の隠れ層の変数を $h_k \rightarrow v_k$ と可視層の変数に置き換えて、結合定数を $K^{(0)} \rightarrow K^{(1)}$ とし、可視層の変数の数は 2^{l-1} 、隠れ層の変数の数は 2^{l-2} として同様の計算を行う。以上の計算を $l-1$ 回繰り返すことにより両端以外のスピンが生成される確率が求まる。

一般の v_k において

$$P(v_k) = \frac{\exp(K^{(n)}v_k(h_{k-2^n} + h_{k+2^n}))}{2 \cosh\{K^{(n)}(h_{k-2^n} + h_{k+2^n})\}} \quad (3.5.9)$$

となる。 n は $\gcd(k, 2^n)$ となる数である。

これにより、スピンの端の向きを決めれば、他の全てのスピンの生成確率は知られているので、すべてのスピンの向きを決めることができる。スピンの生成過程はくりこみの逆変換となっている。端のスピンの決め方は、周期境界条件を課しているので

$$P(v_0) = \frac{\exp(v_0 K^{(l)})}{\exp(-K^{(l)}) + \exp(K^{(l)})} \quad (3.5.10)$$

である。この配位生成法はマルコフ連鎖ではないので自己相関はない。

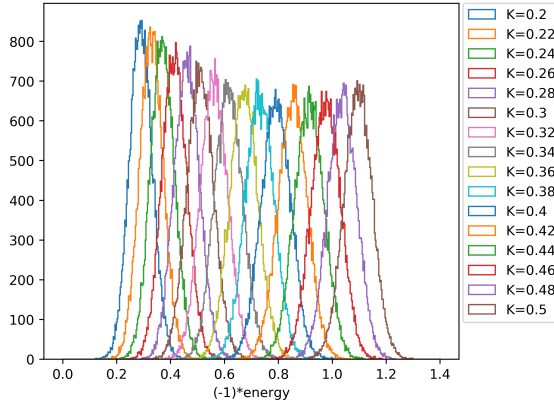


図 3.3: エネルギー分布

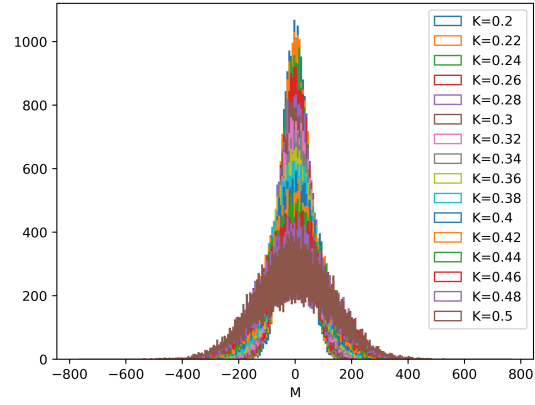


図 3.4: スピン分布

以上の議論を転送行列を用いて書き換えることを試みる。確率分布 (3.5.2) は転送行列を用いて

$$P(\mathbf{v}, \mathbf{h}; \theta) = \frac{1}{Z(\theta)} \prod_{j:\text{even}} \delta(h_j, v_j) T_{(h_j)(v_{j+1})}^{(0)} T_{(v_{j+1})(h_{j+2})}^{(0)} \quad (3.5.11)$$

となり、デルタ関数を積分すると

$$P(v) = \frac{1}{Z} \prod_j T_{(v_j)(v_{j+1})}^{(0)} \quad (3.5.12)$$

となり配位 $\{v\}$ が生成される確率を与える。よって k が奇数の場合、スピン v_k が生成される相対確率は

$$T_{(h_{j-1})(v_j)}^{(0)} T_{(v_j)(h_{j+1})}^{(0)} \quad (3.5.13)$$

であり規格化すると

$$P(v_k) = \frac{T_{(h_{k-1})(v_k)}^{(0)} T_{(v_k)(h_{k+1})}^{(0)}}{\sum_{v_j=\pm 1} T_{(h_{k-1})(v_j)}^{(0)} T_{(v_j)(h_{k+1})}^{(0)}} \quad (3.5.14)$$

と書き換えることができる。 n は $\text{gcd}(k, 2^n)$ とすると

$$P(v_k) = \frac{T_{(h_{k-2^n})(v_k)}^{(n)} T_{(v_k)(h_{k+2^n})}^{(n)}}{\sum_{v_j=\pm 1} T_{(h_{k-2^n})(v_j)}^{(n)} T_{(v_j)(h_{k+2^n})}^{(n)}} \quad (3.5.15)$$

一般の k で書き表わせる。端のスピンは

$$P(v_0) = \frac{T_{(v_0)(v_0)}^{(l)}}{\sum_{v_0=\pm 1} T_{(v_0)(v_0)}^{(l)}} \quad (3.5.16)$$

である。

この方法は、転送行列に書き直し decimation くりこみ群ができれば配位の生成が可能であり、長距離イジングモデルおよび外場がある場合でも適用ができる。図3.5および図3.6は $N = 1024$ 、 $p = 1.8$ における Exact RBM から生成された配位から計算されたエネルギーと比熱である。サンプル数は 128000 である。厳密解 (BDRG) の結果も同時にプロットしており、図3.5および図3.6からエラーバーの範囲内に収まっていることがわかる。

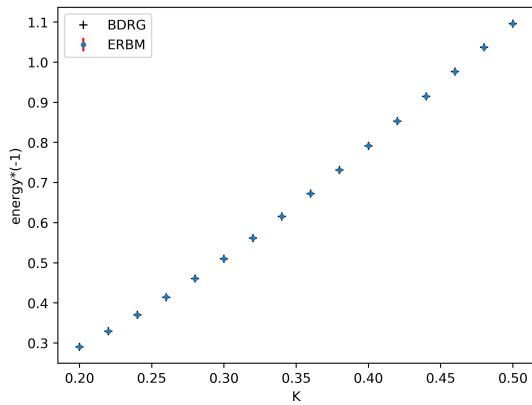


図 3.5: エネルギー

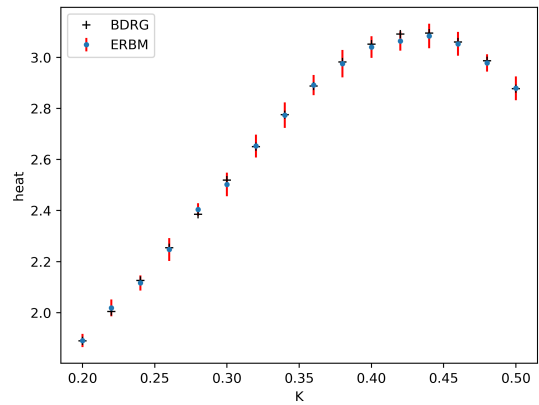


図 3.6: 比熱

第4章 CNNを用いた1次元長距離イジングモデルの温度推定

4.1 訓練データ

モンテカルロシミュレーション (Exact RBM) から1次元長距離イジングモデルの配位を、 $K = 0.2$ から $K = 0.5$ まで 0.02 刻みで生成する。サイズは $N = 2^{10} = 1024$ 、サンプル数 128000 、 $p = 1.8$ とする。ただし配位データをドメインウォール表示にしてある。 K_l を

$$K_l = 0.2 + 0.02l \quad (4.1.1)$$

とし K_l における配位を $\{\sigma\}_{K_l} = \{\sigma\}_l$ とする。 K_l における訓練データを

$$D_l = \{(\{\sigma\}_l, K_l)_n\}_{n=0, \dots, 128000} \quad (4.1.2)$$

として、畳み込みニューラルネットワークを用いて学習させる。そして新たな配位 $\{\sigma\}_l$ から K_l を推定する機械を構築する。全ての訓練データを

$$D = \{(\Sigma_n, K_n)\}_{n=0, \dots, 128000} \quad (4.1.3)$$

と記述することにする。 Σ が入力 x 、 K_n が正解となる出力 y に対応している。実際の機械学習では、データはランダムに入力する。サンプル数の 80% を訓練データとして使い、残りの 20% をテストデータとして過学習が起きていないか判断させる。つまりテストデータは未知のデータに対応している。

4.2 機械の構成

教師ありの機械学習の目的として入力データ \mathcal{X} から出力データ \mathcal{Y} を推定する関数 F

$$F : \mathcal{X} \rightarrow \mathcal{Y} \quad (4.2.1)$$

を構成することである。ニューラルネットワークの入力・出力の関係は

$$z_i(\mathbf{W}, \mathbf{b}, \mathbf{x}) = f(u_i), \quad u_i = \sum_j \omega_{ij} x_j + b_i \quad (4.2.2)$$

f は活性化関数であり ω は重み、 b はバイアスと呼ばれる。太字 \mathbf{W} 、 \mathbf{b} 、 \mathbf{x} は成分 ω 、 b 、 x を行列で表現したものである。活性化関数は正規線形関数 (ReLU) 関数を用いる。

$$f(x) = \max\{0, x\} = \begin{cases} x & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

層を厚くすると

$$z_i^{(l)}(\mathbf{W}^{(l)}, \mathbf{b}^{(l)}, \mathbf{x}) = f(u_i^{(l)}), \quad u_i^{(l)} = \sum_j \omega_{ij}^{(l)} z_j^{(l-1)} + b_i^{(l)} \quad (4.2.3)$$

と記述でき l は層の数であり、 $l=0$ のとき $z_j^{(0)} = x_j$ とした。CNN の場合 u_i は以下のようなになる。

$$u_{i,k} = \sum_{p=0}^{H-1} h_{p,k} z_{Si+p,k} + b_{i,k} \quad (4.2.4)$$

h はフィルターサイズ H であり、添字 k はチャンネル成分を表し S はストライドを表す。サイズ L のデータは畳み込みの操作によってサイズが $(\lfloor \frac{L-H}{S} \rfloor + 1)$ となる。モンテカルロのデータ $\{\sigma\}$ は周期境界条件を課しているので、平行移動不変性を持っている。それを考慮するためにソフトマックス関数に $z_{i,k}$ を代入する前に添字 i, k に対して縮約する。

$$z = \sum_i \sum_k z_{i,k} \quad (4.2.5)$$

縮約された z を用いて、入力データ x のクラス k に属する確率を求める。

$$\begin{aligned} P(C_k | \mathbf{W}, \mathbf{b}, \mathbf{x}) &= \operatorname{softmax}_k(u_1, u_2, \dots, u_K) \\ &\equiv \frac{e^{u_k}}{\sum_l e^{u_l}}, \quad u_k = \omega_k z + b_k \end{aligned} \quad (4.2.6)$$

最終的に入力データ x が推定する出力 y はソフトマックス関数が最大となる確率つまりもっともらしいクラスを出力とする。

$$y(\mathbf{W}, \mathbf{b}, \mathbf{x}) = \arg \max_C P(C | \mathbf{W}, \mathbf{b}, \mathbf{x}) \quad (4.2.7)$$

機械の構成を図4.1にまとめた。教師あり学習の場合、出力結果がわかっているので機械学習が推定した値と正解の値が一致するようにパラメータ ω, b を最適化する。これらのパラメータはクロスエントロピーが最小となるように決定する。

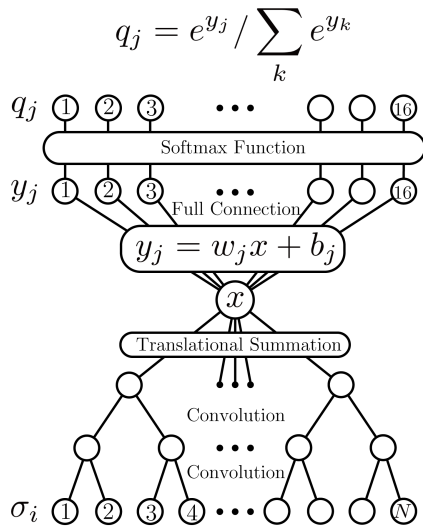
$$E(\mathbf{W}, \mathbf{b}) = -\frac{1}{N} \sum_{n=1}^N \sum_{k=1}^K \{\delta_{y_n k} \log(y(\mathbf{W}, \mathbf{b}, \mathbf{x}_n))\} \quad (4.2.8)$$

また機械の性能を評価するために正答率 A を

$$A = \frac{1}{N} \sum_{n=1}^N \delta(y(\mathbf{W}, \mathbf{b}, \mathbf{x}_n), y_n) \quad (4.2.9)$$

と定義する。本研究ではマシンパラメータ（ハイパーパラメータ）を表4.1とした。また畳み込み1層目のフィルターサイズ H_1 とストライド S_1 は8として変化させている。ハイパーパラメータは経験則に基づいて、またフィルターサイズやストライドはブロックスピン変換を意識して定めた。

正答率を上げるため入力データを $1024 \times Na$ とし2次元データを入力とした。つまり1次元長距離イジングモデルの配位を縦に Na 個並べる。しかし畳み込みの操作は



hyperparameter	value
Convolution layer l	4 layer
Filter size H_1	8
Stride S_1	8
Filter size H	2
Stride S	2
Mini batch size	120
Learning rate	0.0001
Channel	8
Na	1, 2, 4

表 4.1: hyperparameter

図 4.1: 機械の構成

横方向のみとする。ただしデータの数も固定してあるので Na を増やせばサンプル数も $128000/Na$ となる。 $Na = 1, 2, 4$ のとき正答率は 43.5%, 56.1%, 70.1% であり、誤差関数の値は 1.274, 0.985, 0.720 となる。

各 Na における入力データのヒストグラムを図 4.2、図 4.4、図 4.6、機械が出力（推定）したヒストグラムを図 4.3、図 4.5、図 4.7 にする。

Na の値を大きくすれば正答率は上がるが、機械が出力（推定）したヒストグラムを見ると分布が被らない端を除けば Na の値によらずほぼ同じ形をしている。これは機械は分布の交点で何かしらの操作をしていることがわかる。この理由は機械学習が配位から温度を推定する際に何を学んでいるのか考える際に重要となってくる。

第4章 CNNを用いた1次元長距離イジングモデルの温度推定

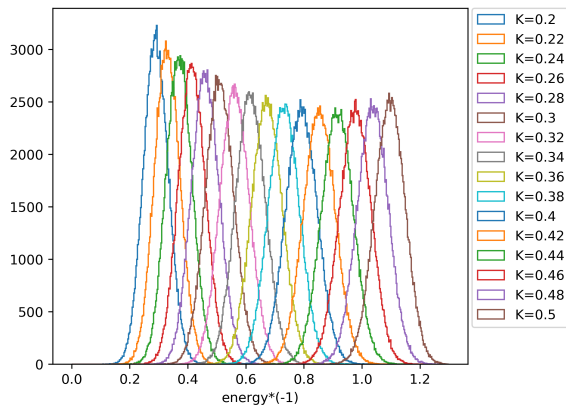


図 4.2: $N_a=1$ の入力ヒストグラム

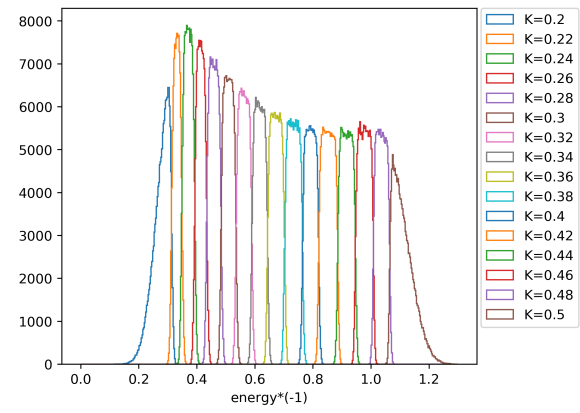


図 4.3: $N_a=1$ の出力ヒストグラム

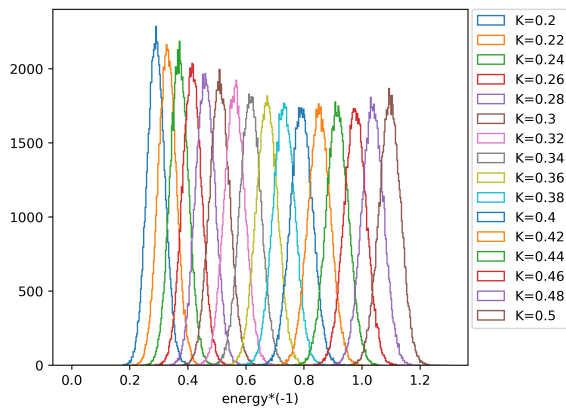


図 4.4: $N_a=2$ の入力ヒストグラム

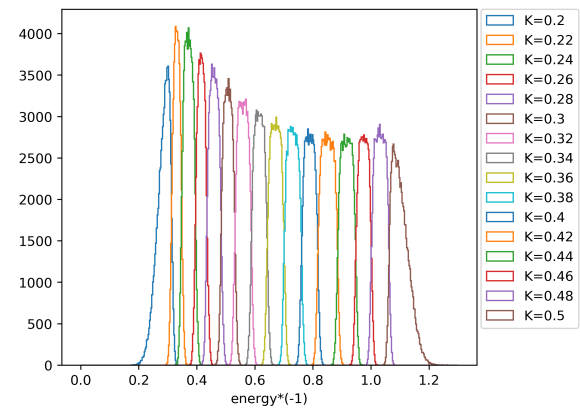


図 4.5: $N_a=2$ の出力ヒストグラム

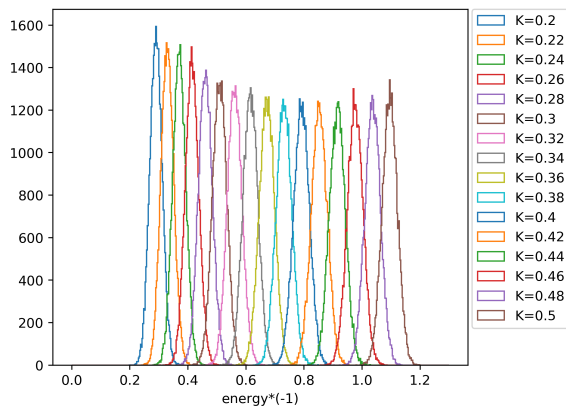


図 4.6: $N_a=4$ の入力ヒストグラム

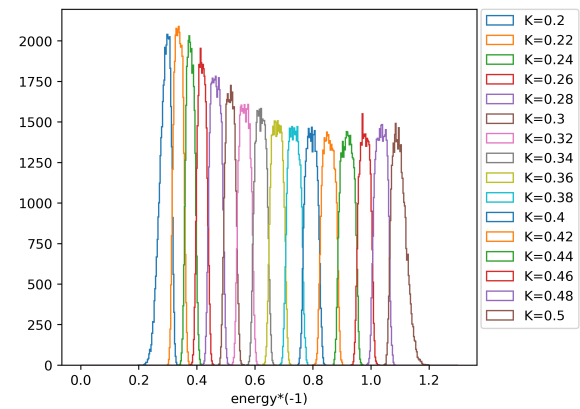


図 4.7: $N_a=4$ の出力ヒストグラム

第5章 1次元長距離イジングモデルの温度推定

5.1 最尤推定

最尤推定を用いて、1次元長距離イジングモデルの温度推定について考える。1次元長距離イジングモデルのハミルトニアンを温度とオペレーターの部分で分けて記述することにする。

$$H = -K \sum_n \sum_i \left(\frac{1}{n^p} \right) \sigma_i \sigma_{i+n} = -K \hat{O} \quad (5.1.1)$$

もし、ある温度における配位が与えられた場合、その生成される確率は

$$P[\sigma; K] = \frac{\exp(-H)}{Z(K)} \quad (5.1.2)$$

となる。自由エネルギー密度を以下のように定義する。

$$W(K) = \frac{1}{N} \log Z(K) \quad (5.1.3)$$

自由エネルギー密度は、単調増加関数である。(5.1.2)より

$$\frac{1}{N} \sum_i \log P[X_i; K] = \frac{1}{N} \log P[X; K] = K \frac{\hat{O}}{N} - W(K) \quad (5.1.4)$$

となる。(5.1.4)から配位の生成確率は、確率分布を最大にする K が存在することがわかる。確率分布を微分することにより

$$\frac{1}{N} \frac{\partial}{\partial K} \log P[X; K] \Big|_{K=K^*} = 0 \quad (5.1.5)$$

尤もらしい確率となる条件（最尤推定）がもとまる。

$$\hat{O}(X) = \langle \hat{O}(X) \rangle \Big|_{K^*} \quad (5.1.6)$$

しかし(5.1.6)を解くことは困難である。

5.2 1次元イジングモデル

1次元最近接相互作用の場合は、二項分布となっており式(3.1.5)、式(3.1.6)および式(5.1.6)の条件式から

$$p(K^*) = \frac{n}{N} \quad (5.2.1)$$

表 5.1: 正答率の上限値 (正規分布)

	n=1	n=4	n=8
K=0.20	0.623	0.663	0.670
K=0.22	0.245	0.331	0.349
K=0.24	0.244	0.337	0.359
K=0.26	0.243	0.343	0.369
K=0.28	0.242	0.348	0.379
K=0.30	0.240	0.353	0.388
K=0.32	0.239	0.357	0.398
K=0.34	0.238	0.361	0.406
K=0.36	0.236	0.364	0.414
K=0.38	0.235	0.366	0.420
K=0.40	0.233	0.367	0.424
K=0.42	0.231	0.367	0.427
K=0.44	0.229	0.367	0.427
K=0.46	0.227	0.365	0.425
K=0.48	0.225	0.362	0.420
K=0.50	0.612	0.679	0.707
上限値	0.284	0.396	0.436

表 5.2: 正答率の上限値の比較

	ERBM	正規分布
K=0.20	0.681	0.670
K=0.22	0.344	0.349
K=0.24	0.375	0.359
K=0.26	0.345	0.369
K=0.28	0.394	0.379
K=0.30	0.382	0.388
K=0.32	0.392	0.398
K=0.34	0.419	0.406
K=0.36	0.395	0.414
K=0.38	0.451	0.420
K=0.40	0.416	0.424
K=0.42	0.415	0.427
K=0.44	0.449	0.427
K=0.46	0.399	0.425
K=0.48	0.457	0.420
K=0.50	0.685	0.707
上限値	0.437	0.436

最尤推定が求まる。

ドメインウォール数 n の場合、各温度 K_i における確率を求め最大となる $P(K^*)$ が、ドメインウォール数 n における最適な温度 K^* である。温度 K_i における最大となる $P(K^*)$ を足し上げた値が、温度 K_i における正答率の上限値

$$A_i = \sum_{n \in \arg \max_{K_i} P(n, K_i)} P(n, K_i) \quad (5.2.2)$$

である。温度 K_i における正答率を足し上げた値を K_i の i の数 N_{class} だけ割ったものが、与えられた配位における正答率の上限値

$$A_{\text{max}} = \frac{1}{N_{\text{class}}} \sum_{i=1}^{N_{\text{class}}} A_i \quad (5.2.3)$$

となる。

5.3 1次元長距離イジングモデル

正規分布近似した1次元長距離イジングモデルの確率分布の場合、温度 K_i における正答率の上限値 A_i は

$$A_i = \frac{1}{\sqrt{2\pi\sigma_i^2}} \int_{b_i}^{a_i} \exp \left\{ -\frac{(x - \mu_i)^2}{2\sigma_i^2} \right\} dx \quad (5.3.1)$$

である。積分範囲の a_i, b_i は隣り合う正規分布の交点である。平均と分散は

$$\begin{aligned}\mu_i &= \left. \frac{\partial F}{\partial K} \right|_{K_i} = \frac{1}{N} \langle \hat{O} \rangle \Big|_i \\ \sigma_i^2 &= \left. \frac{1}{N} \frac{\partial^2 F}{\partial K^2} \right|_{K_i} = \frac{1}{N^2} \left\langle \left(\hat{O} - \langle \hat{O} \rangle \right)^2 \right\rangle \Big|_i\end{aligned}\quad (5.3.2)$$

である。 A_i は1次元最近接イジングモデルの議論において温度 K_i における最大となる $P(K^*)$ を足し上げた値に相当する。与えられた配位における正答率の上限値は

$$A_{\max} = \frac{1}{N_{\text{class}}} \sum_{i=1}^{N_{\text{class}}} A_i \quad (5.3.3)$$

である。表5.1に各温度における正答率の上限値と正答率の上限値の計算結果である。

Exact RBM の場合は、各エネルギーにおいて確率が最大となる $P(K_i)$ を求めれば、 K_i における正答率の上限値は $P(K_i)$ の足し上げとなる。

$$A_i = \sum_{E \in \arg \max_{K_i} P(E, K_i)} P(E, K_i) \quad (5.3.4)$$

与えられた配位における正答率の上限値は

$$A_{\max} = \frac{1}{N_{\text{class}}} \sum_{i=1}^{N_{\text{class}}} A_i \quad (5.3.5)$$

である。表5.2に Exact RBM の結果と正規分布近似の計算結果である。

5.4 N_a Configuration

ある配位 σ_i におけるエネルギー $E_k^{(i)}$ とし添字 k を N_a の数とすると N_a 個の配位における確率分布は

$$\begin{aligned}\mathcal{P}(\{\sigma\}_i, K) &= \prod_{k=1}^{N_a} \frac{\exp\{-K E_k^{(i)}\}}{Z(K)} = \frac{\exp\{-K \sum_{k=1}^{N_a} E_k^{(i)}\}}{(Z(K))^{N_a}} \\ &= \left(\frac{\exp\{-K E_a^{(i)}\}}{Z(K)} \right)^{N_a} = (P(E_a^{(i)}; K))^{N_a}\end{aligned}\quad (5.4.1)$$

となる。途中で平均エネルギーを

$$E_a^{(i)} = \frac{1}{N_a} \sum_{k=1}^{N_a} E_k^{(i)} \quad (5.4.2)$$

とした。式(5.4.1)から

$$P(E_a^*; K_j) = P(E_a^*; K_{j+1}) \quad (5.4.3)$$

表 5.3: n=8 における各 N_a における正答率

$N_a=1$	$N_a=2$	$N_a=4$
0.436	0.569	0.722

が得られる。 E_a^* は 確率分布 $P(E; K_j)$ と $P(E; K_{j+1})$ の交点であり、また確率分布の交点は N_a の数によらない。エネルギーの平均と分散は $\mu \rightarrow \mu$, $\Sigma^2 \rightarrow \Sigma^2/N_a$ と変化する。 N_a 個の配位の正規分布 (近似) は

$$P(E_a; K_i) \approx \sqrt{\frac{N_a}{2\pi\Sigma_i^2}} \exp\left\{-\frac{N_a(E_a - \mu_i)^2}{2\Sigma_i^2}\right\} \quad (5.4.4)$$

となる。図 5.1、図 5.2 は $N_a = 2, 4$ の場合の正規分布を表したものである。 $N_a = 1, 2, 4$ のとき正答率の上限値は 43.6%, 56.9%, 72.2%であった。機械学習から得られた結果とかなり近いことがわかり、かなり良い精度で温度推定が行われていることがわかる。

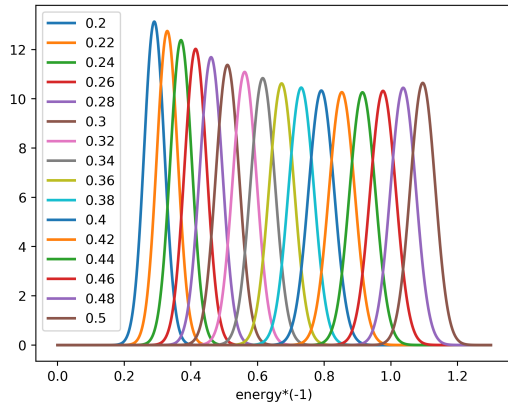


図 5.1: $N_a=2$ の場合の正規分布近似

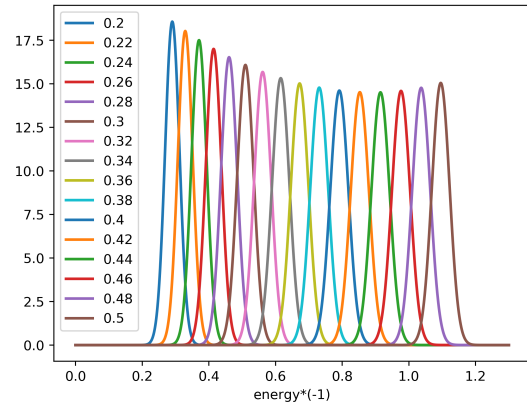


図 5.2: $N_a=4$ の場合の正規分布近似

図 5.3, 5.4 に各温度クラスにおける正答率の理論的上限值と畳み込みニューラルネットワークの正答率を示す。 $n = 1$ の場合、正答率の理論的上限值は 28.3%であり、CNN の結果は 28.1%であった。

5.5 混合サンプルにおける温度推定

実際の機械学習では、各 K_i の配位の集合、つまり混合サンプルから学習している。この混合サンプルにおける配位 σ の生成確率は

$$P(\{\sigma\}) = \frac{1}{N_{\text{class}}} \sum_j^{N_{\text{class}}} P(\{\sigma\}; K_j) = \frac{1}{N_{\text{class}}} \sum_j^{N_{\text{class}}} \frac{\exp(-K_j H(\{\sigma\}))}{Z(K_j)} \quad (5.5.1)$$

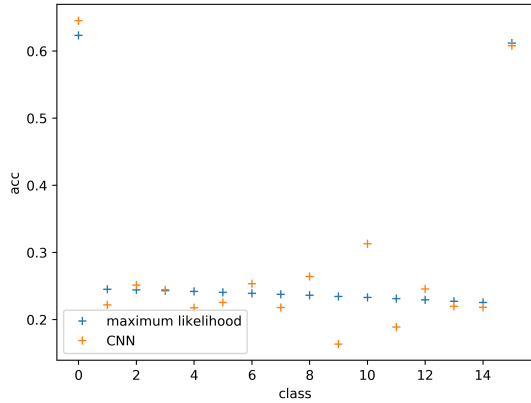


図 5.3: 機械の推定と最尤推定 (n=1)

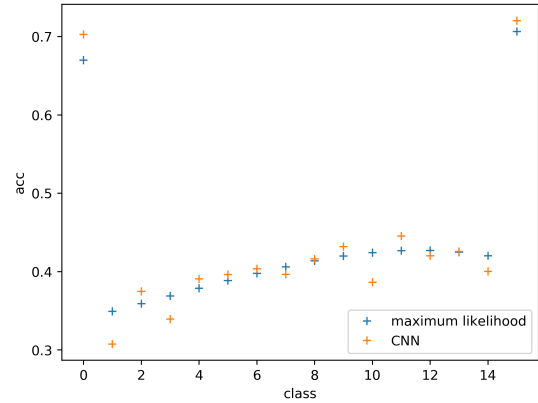


図 5.4: 機械の推定と最尤推定 (n=8)

と表せる。ある配位が与えられた際、その配位が K_j である確率は

$$Q(\{\sigma\}; j = K_j) = \frac{P(\{\sigma\}; K_j)}{\sum_l P(\{\sigma\}; K_l)} \approx \frac{\mathcal{N}(E; \mu_j, \Sigma_j)}{\sum_l \mathcal{N}(E; \mu_l, \Sigma_l)} \quad (5.5.2)$$

と表せ、2行目は正規分布近似の場合である。図 5.6、図 5.8、図 5.10 は正規分布近似した $Q(\{\sigma\}; j = K_j)$ を表す。したがってその配位の尤もらしい $K_{j \max}$ は

$$K_{j \max}(\{\sigma\}) = \arg \max_j Q(\{\sigma\}; j) \quad (5.5.3)$$

となる。最大正答率 $A(\{\sigma\})$ は

$$A(\{\sigma\}) = Q(\{\sigma\}; K_{j \max}) \quad (5.5.4)$$

となる。よって正答率の理論的上限值は式 (5.5.4) のサンプル平均で表せ

$$A_{\max} = \frac{1}{N_{\text{class}}} \sum_j \sum_{\{\sigma\}} P(\{\sigma\}; K_j) Q(\{\sigma\}; K_{j \max}) \quad (5.5.5)$$

となる。式 (5.5.5) は式 (5.2.3)、式 (5.3.3)、式 (5.3.5) と同じ結果となる。

最適化されたパラメータを用いてソフトマックス関数を

$$q(K_j; x) = P(C_j | \mathbf{W}, \mathbf{b}, \mathbf{x}) = \text{softmax}_j(u_1, u_2, \dots, u_{N_{\text{class}}}) \quad (5.5.6)$$

と表すことにする。図 5.5、図 5.7、図 5.9 は正規分布近似した $q(K_j; x)$ を表す。最適化されたパラメータは4章で説明した結果を用いた。

第5章 1次元長距離イジングモデルの温度推定

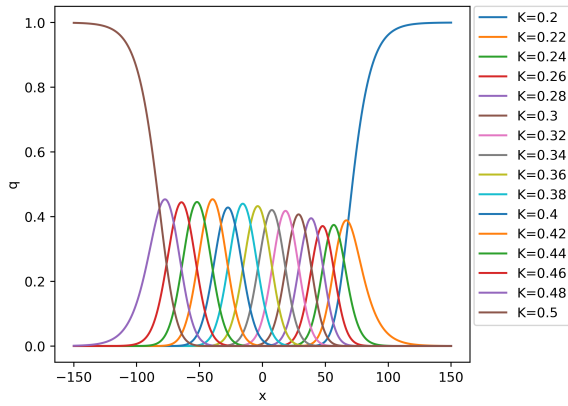


図 5.5: $N_a=1$ におけるソフトマックス関数

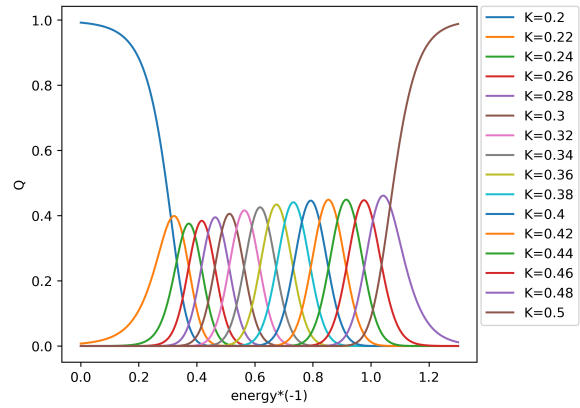


図 5.6: $N_a=1$ における確率

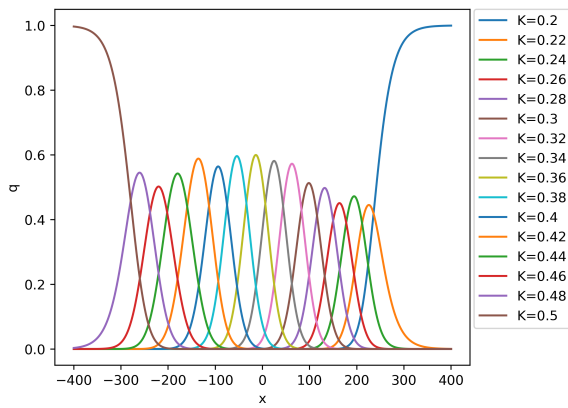


図 5.7: $N_a=2$ におけるソフトマックス関数

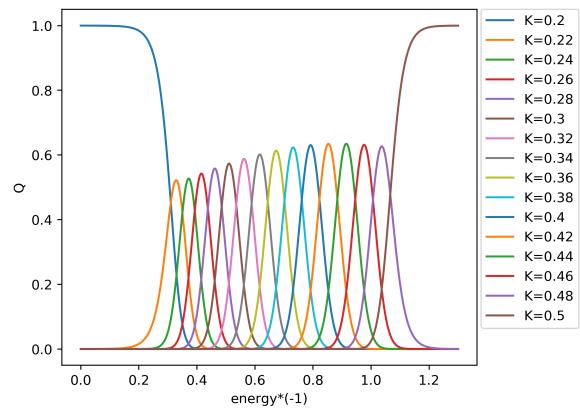


図 5.8: $N_a=2$ における確率

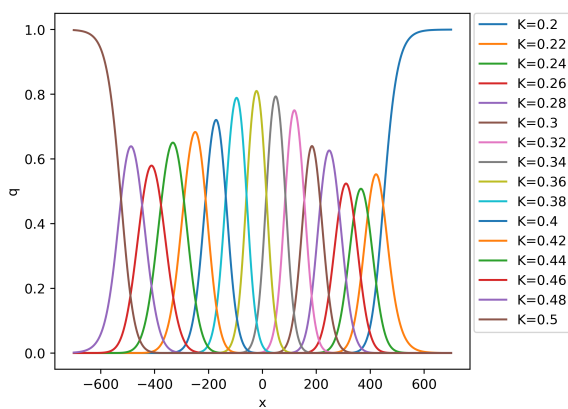


図 5.9: $N_a=4$ におけるソフトマックス関数

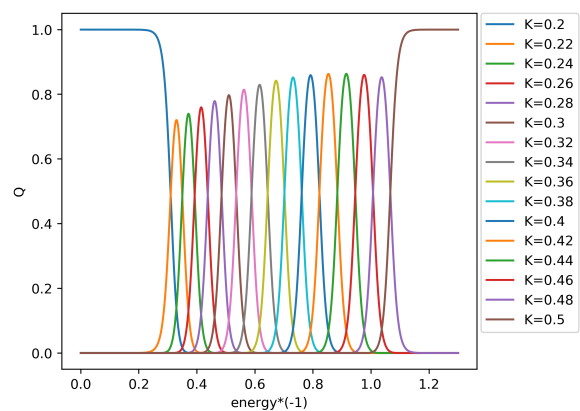


図 5.10: $N_a=4$ における確率

第6章 機械は何を学んでいるか？

6.1 ハミルトニアン認識率 (HRR)

配位はカノニカル分布に従っているので、機械が正答率の理論的上限值に近い値を出すためにはハミルトニアンを学ばなければならない。理論上限値と畳み込みニューラルネットワークの正答率を比較すると、概ね近い値を出していることから、畳み込みニューラルネットワークはハミルトニアンを認識していると考えられる。

その結果は入力データのヒストグラム図4.2、図4.4、図4.6、機械が出力したヒストグラム図4.3、図4.5、図4.7を比べてみると理論上限値の機械ならば、入力データのエネルギー分布の交点で分離され、16本の柱が立っているような分布になるはずである。例えば図4.3は、多少は分布の交点をもつが、図4.2と比べて分離されている。これは学習後の機械がハミルトニアン関数を持っており、ハミルトニアンを認識していることをしめしている。 N_a を増やして正答率を大きくしても機械が出力したヒストグラムが変わらない理由はハミルトニアンを認識して推定をしているからである。

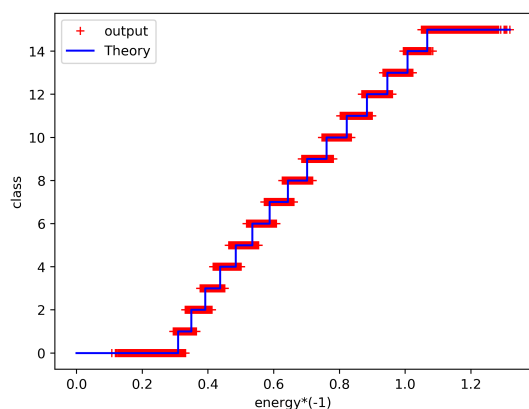


図 6.1: エネルギーと推定温度 (n=8)

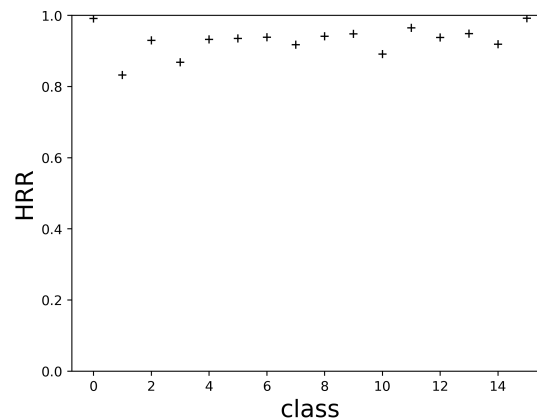


図 6.2: n=8 における HRR

理論上限値を実現するには、出力クラスは確率分布の交点で出力クラスが変わるようなハミルトニアン関数の関数となるはずである。その結果は図6.1にする。線は理論的なハミルトニアン出力関数の結果であり、プロットは畳み込みニューラルネットワークの出力関数の結果を表す。畳み込みニューラルネットワークが完全にハミルトニアンを認識しているならば、プロットは線上に乗っているはずである。畳み込みニューラルネットワークの出力は理論的な結果から少しずれていることがわかる。

次に機械はどれだけハミルトニアンを認識できているか考える。クラス n に出力されている配位数 N_{conf} 、線上に乗っているプロットの数 N_{max} とする。クラス n におけるハミルトニアン認識率は

$$R_n = \frac{N_{\text{max}}}{N_{\text{conf}}} \quad (6.1.1)$$

でありその結果を図 6.2 に示す。全体のハミルトニアン認識率は図 6.1 の青色の理論値にどれだけ赤色のプロットがのっているかの割合とする。 $n = 8$ のハミルトニアン認識率は 93.67% であった。

6.2 共分散不等式

まずエネルギー以外の物理量の情報では、正答率はどうなるか考える。自由エネルギーを

$$e^{W(K,h)} = \sum_{\{\sigma\}} e^{-KE(\sigma)-h\theta(\sigma)} \quad (6.2.1)$$

とする。また自由エネルギーは半正定値である。

$$\begin{pmatrix} \frac{\partial^2 W}{\partial K^2} & \frac{\partial^2 W}{\partial K \partial h} \\ \frac{\partial^2 W}{\partial K \partial h} & \frac{\partial^2 W}{\partial h^2} \end{pmatrix} \succeq 0 \quad (6.2.2)$$

(6.2.2) は任意のベクトル $(x \ y)$ を用いて

$$\frac{\partial^2 W}{\partial K^2} x^2 + 2xy \frac{\partial^2 W}{\partial K \partial h} + \frac{\partial^2 W}{\partial h^2} y^2 \geq 0 \quad (6.2.3)$$

と書き直せる。(6.2.3) が成り立つためには、判別式 (共分散不等式)

$$\left(\frac{\partial^2 W}{\partial K \partial h} \right)^2 - \frac{\partial^2 W}{\partial K^2} \frac{\partial^2 W}{\partial h^2} \leq 0 \quad (6.2.4)$$

を満たす必要があり、式変形を行うと

$$\left| \left(\frac{\partial^2 W}{\partial K \partial h} \right) \right| \leq \sqrt{\frac{\partial^2 W}{\partial K^2}} \sqrt{\frac{\partial^2 W}{\partial h^2}} \quad (6.2.5)$$

$$\frac{\frac{\partial \langle \theta \rangle}{\partial K}}{\sqrt{\frac{\partial^2 W}{\partial h^2}}} \leq \sqrt{\frac{\partial^2 W}{\partial K^2}} = \frac{\frac{\partial \langle E \rangle}{\partial K}}{\sqrt{\frac{\partial^2 W}{\partial K^2}}} \quad (6.2.6)$$

となる。正答率は、各温度における分布同士の距離と分布の分散の大きさによって決まる。よってこの不等式は温度推定する際、エネルギー以外の物理量の情報では、正答率は下がることを意味する。

第7章 相転移の特異性

機械はハミルトニアンを認識し温度を推定している。エネルギーのゆらぎすなわち比熱は相転移温度で特異性を持ち、この比熱の特異性を機械がどこかに情報を残しているはずである。畳み込みニューラルネットワークの分類器つまりソフトマックス関数に注目する。畳み込みの操作によって圧縮された配位 $z = z(\{\sigma\})$ とすると、重み ω ・バイアス b として出力 y は

$$y_j = \omega_j z + b_j \quad (7.0.1)$$

であり、ソフトマックス関数を用いて分類される。

$$q(K_j; x) = \frac{\exp(y_j)}{\sum_k \exp(y_k)} \quad (7.0.2)$$

混合サンプルの場合を考えると機械学習における誤差関数は、ある配位 $\{\sigma\}$ において、対数尤度関数は

$$C_{i,\{\sigma\}} = -\log q(\{\sigma\}; K_i) \quad (7.0.3)$$

と表す。また式 (7.0.3) をコスト関数と言う。配位 $\{\sigma\}$ を固定して i に関してのコスト関数の平均は確率 (5.5.2) を用いて

$$\begin{aligned} C_{\{\sigma\}} &= -\sum_i Q(\{\sigma\}; K_i) \log q(\{\sigma\}; K_i) \\ &= \sum_i Q(\{\sigma\}; K_i) \log \frac{Q(\{\sigma\}; K_i)}{q(\{\sigma\}; K_i)} - \sum_i Q(\{\sigma\}; K_i) \log Q(\{\sigma\}; K_i) \end{aligned} \quad (7.0.4)$$

である。また全サンプルの平均は式 (5.5.1) を用いて

$$C = \sum_{\{\sigma\}} P(\{\sigma\}) C_{\{\sigma\}} \quad (7.0.5)$$

とあらわせる。コスト関数 C を最小値するには $Q(\{\sigma\}; K_i)$ を固定して $q_{\max}(\{\sigma\}; K_i)$ を調整することになる。

式 (7.0.5) の最小値は

$$C_{\min} = -\sum_{\{\sigma\}} P(\{\sigma\}) \sum_i Q(\{\sigma\}; K_i) \log Q(\{\sigma\}; K_i) \quad (7.0.6)$$

となる。 C_{\min} は正規分布近似および Exact RBM から求めることができ計算することができる。正規分布近似における最小値は

$$C_{\min} = 1.2697 \quad (7.0.7)$$

であり Exact RBM おける最小値は

$$C_{\min} = 1.2673 \tag{7.0.8}$$

である。機械学習の結果 ($Na = 1$) は 1.274 なので、理論的な誤差関数の最小値に非常に近いことがわかる。つまり機械はソフトマックス関数を $Q(K_i; \{\sigma\})$ に近くようにパラメータを定めていることになる。

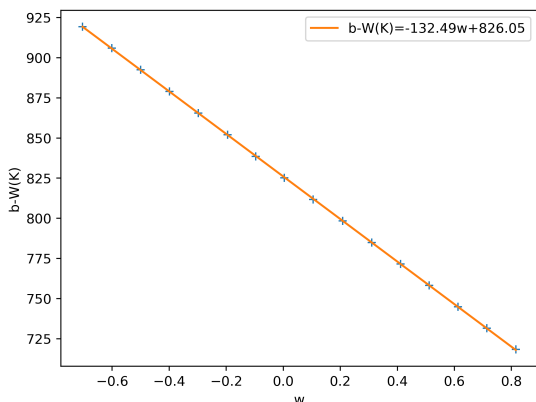


図 7.1: バイアス・自由エネルギーと重み

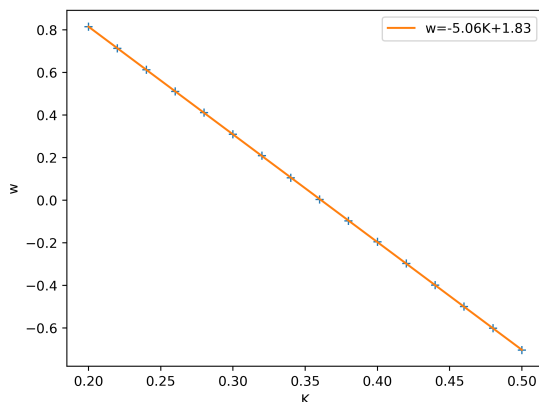


図 7.2: 重みと温度の関係

機械が学習によってハミルトニアンを認識していることと、誤差関数の最小値を求める議論から、ソフトマックス関数に入力される値 x は $x \propto E$ となると期待できる。よって

$$\begin{aligned} \exp(\omega_j x + b_j) R(E) &= \exp(-K_j E + W(K_j)) \\ \omega_j x + b_j + \log R(E) &= -K_j E + W(K_j) \end{aligned} \tag{7.0.9}$$

である。 $x = -a_1 E - a_0$ 、 $\log R(E) = c_0 + c_1 E$ とすると

$$\begin{aligned} \omega_j &= \frac{K_j + c_1}{a_1} \\ b_j &= W(K_j) + \omega_j a_0 - c_0 \end{aligned} \tag{7.0.10}$$

となる。 ω_j は K_j の 1 次式であり、 $b_j - W(K_j)$ は K_j の 1 次式である。よって b_j の温度の 2 階微分は不定性がなくなる。 b_j の温度の 2 階微分は比熱となる。

$$-\frac{a_1^2 (b_{j+1} - 2b_j + b_{j-1})}{(\omega_{j+1} - \omega_j)(\omega_j - \omega_{j-1})} = -\frac{\partial^2 W(K)}{\partial K^2} = C(K_j) \tag{7.0.11}$$

図 7.1 は、 ω_j と K_j をプロットしたものである。また図 7.2 は $b_j - W(K_j)$ と K_j をプロットしたものである。これらから ω_j は K_j の 1 次式であり、 $b_j - W(K_j)$ は K_j の 1 次式であることがわかる。

図 7.6 は、式 (7.0.11) をプロットしたものである。ヘルムホルツの自由エネルギーを

$$\begin{aligned} F(K_j) &= -\frac{1}{K_j} \log Z(K_j) = \frac{1}{K_j} W(K_j) \\ &= E - \frac{S(K_j)}{K_j} \end{aligned} \tag{7.0.12}$$

とすると、

$$\begin{aligned}
 y_j &= \omega_j x + b_j \\
 &= -\frac{K_j + c_1}{a_1}(a_1 E + a_0) + W(K_j) + \frac{K_j + c_1}{a_1}a_0 - c_0 \\
 &= -K_j E + F(K_j) + \mathcal{O}(E^1) + \text{const} \\
 &= -S(K_j) + \mathcal{O}(E^1) + \text{const}
 \end{aligned} \tag{7.0.13}$$

となる。図 7.7 は、式 (7.0.13) をプロットしたものであり、図 7.7 の包絡線はエントロピーを示している。

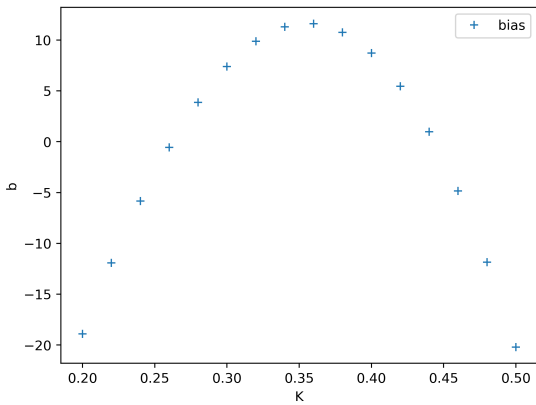


図 7.3: バイアス

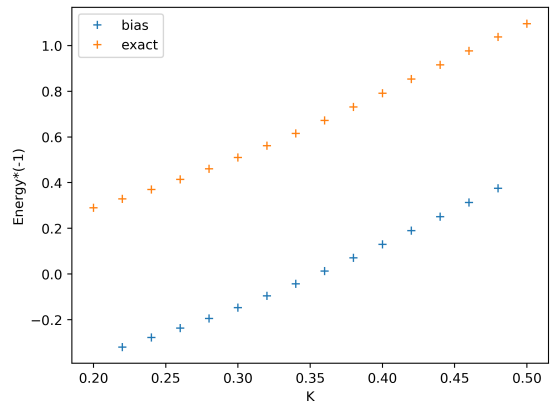


図 7.4: バイアスの1階差分

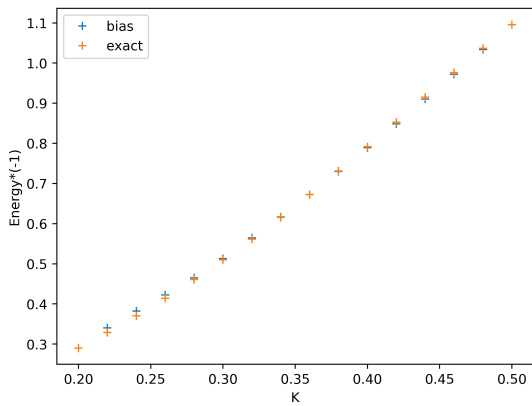


図 7.5: バイアスの1階差分を平行移動

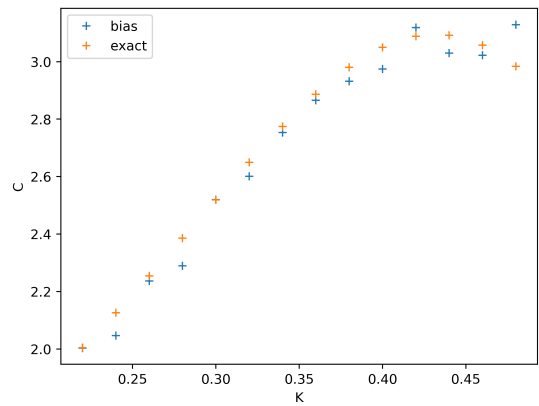


図 7.6: バイアスの2階差分

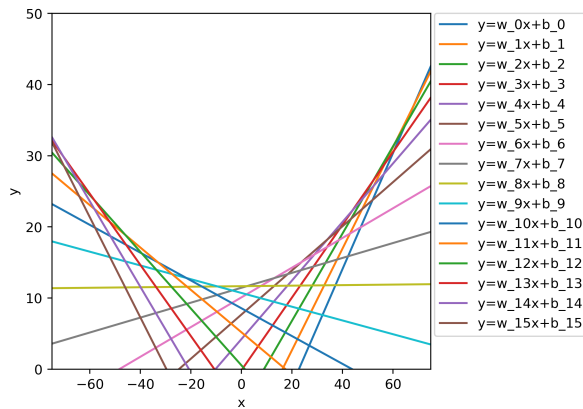


図 7.7: y と x の関係

第8章 まとめ

本論文では、統計系の配位をもとに深層学習を用いて温度推定を行い、相転移温度の推定を行った。機械はソフトマックス関数を混合サンプルにおけるカノニカル分布になるようにパラメータを最適化している。最適化された機械は、統計系の配位から温度を推定する過程において機械内部にエネルギー関数を構成し、ソフトマックス関数のパラメータに物理的な情報である温度と自由エネルギーの情報を残している。つまりソフトマックス関数の重み ω_j は温度 K_j の1次式であり、バイアスから自由エネルギーを引いた量 $b_j - W(K_j)$ は温度 K_j の1次式である。それらの情報から比熱の情報が得られ、相転移温度が推定できる。以上の温度推定の解析はモデルや空間次元に依存せず成り立つことがわかる。

配位集合を生成する際に用いた Exact RBM の方法は、RBM のパラメータをくりこみ群を用いて決めることができる。くりこみ群と逆の過程から自己相関がない配位集合を生成できるため計算時間が短縮される。この方法は厳密くりこみ群が可能な1次元（最近接・長距離）イジングモデルには適用できるが、他のモデルでも同様な計算ができるか今後の課題である。

深層学習における相転移温度推定について議論することで、くりこみ群との関係性を見出したいのだが、その関係性は未だ不明な状況である。

第9章 Appendix

9.1 peierls の議論

イジングモデルの相転移の有無は次元によることを説明する [18]。最近接相互作用イジングモデルのハミルトニアンは

$$H = -K \sum_{\langle i,j \rangle} \sigma_i \sigma_j - h \sum_j \sigma_j \quad (9.1.1)$$

である。また以下の議論は周期境界条件および $h = 0$ を課すことにする。

まず1次元の場合を考える。低温極限すなわち $K = \infty$ の際、スピンは全て上向きか下向きになる。自由エネルギーは、エントロピーの効果を取り入れなくてよく、

$$F_0 = -N \quad (9.1.2)$$

となる。温度をあげるとスピンの揺らぎが生じるので、スピンを1個反転させることを考慮する必要があり、エネルギーは2損することになる。また反転するスピンの位置は任意であるので、そのような位置は N 通りある。よってエントロピーは $S = \log_N C_1$ である。よって温度を上げた場合の自由エネルギーは

$$F = -(N - 2) - \frac{1}{K} \log N \quad (9.1.3)$$

である。温度を上げた場合の変化量は

$$\Delta F = 2 - \frac{1}{K} \log N \quad (9.1.4)$$

である。よって有限温度で N を十分大きくとると、自由エネルギーが低下する。これは自由エネルギーは特異性を持たないことを意味しており、全て同じ向きに揃った状態は、少しの熱揺らぎによって破壊される。つまり温度を上げると全て同じ向きに揃った状態は不安定になる。よって1次元の場合は、長距離秩序が存在しないことすなわち、スピンの長距離まで揃いにくいことを意味する。

次に2次元の場合も同様に計算してみる。スピンの島が反転しているスピンの島を考えると、スピンの島によるエネルギーの損失は島の周囲の長さ L とすると $2L$ となる。周囲の長さ L の島が作られる状態数は $3^L N$ 通りある。ただし L 歩進んで元に戻らないといけないという条件は無視している。温度を上げた場合の自由エネルギーの変化量は

$$\begin{aligned} \Delta F &= 2L - \frac{1}{K} \log 3^L N \\ &\approx L \left(2 - \frac{1}{K} \log 3 \right) \end{aligned} \quad (9.1.5)$$

となる。島の大きさを十分大きくして $\log N$ を無視した。ここで

$$\frac{1}{K_c} = \frac{2}{\log 3} \quad (9.1.6)$$

とすると $K = K_c$ において自由エネルギーに特異性を持つことがわかる。 $K \geq K_c$ の場合は、自由エネルギーは増加し全て同じ向きに揃った状態は安定である。 $K < K_c$ の場合は、自由エネルギーは減少し不安定である。2次元の場合は長距離秩序があり、相転移が起こることを示唆している。

peierls の議論は、空間の次元は相転移の有無を決定する重要な要素であることを示唆している。 d 次元において、島の周囲の長さ L のドメインウォールの大きさは L^{d-1} となり、エネルギーの増加分は、ドメインウォールの大きさに比例する。1次元の場合自由エネルギーの増加分は、ドメインウォールが点なので、エネルギーに比べてエントロピーの増加分が大きく、スピンの揃った状態を壊しやすくなる。

最近接相互作用のイジングモデルは、1次元では有限温度で相転移を持たないことは厳密解からわかる。しかし長距離相互作用がある場合は1次元でも相転移が起こるとされている。1次元長距離イジングモデルのハミルトニアンは

$$H = -K \sum_n \sum_i \left(\frac{1}{n^p} \right) \sigma_i \sigma_{i+n} - h \sum_j \sigma_j \quad (9.1.7)$$

であり、減衰因子である $1/n^p$ の p が $1 < p \leq 2$ の場合に相転移が起こることが示唆されている [19][20]。平均場近似を用いて評価してみる。平均場近似を用いるとハミルトニアンは

$$H \approx - \sum_i \sigma_i (2K\zeta(p) \langle \sigma \rangle + h) \quad (9.1.8)$$

となる。ゼータ関数を以下のように定義した。

$$\zeta(p) = \sum_{n=1}^{\infty} \frac{1}{n^p} \quad (9.1.9)$$

磁化を計算すると

$$\langle \sigma \rangle = \frac{\prod_i \sum_{\{\sigma\}} \sigma_i \exp(\sigma_i (2K\zeta(p) \langle \sigma \rangle + h))}{\prod_i \sum_{\{\sigma\}} \exp(\sigma_i (2K\zeta(p) \langle \sigma \rangle + h))} = \tanh(2K\zeta(p) \langle \sigma \rangle + h) \quad (9.1.10)$$

となる。臨界点付近では磁化は十分小さいとすれば、式 (9.1.10) の右辺を磁化の1次で展開すれば

$$K_c = \frac{1}{2\zeta(p)} \quad (9.1.11)$$

臨界結合定数（または温度）が求まる。自発磁化を知りたいので $h = 0$ とした。臨界結合定数 K_c が式 (9.1.11) の右辺より大きい場合に自発磁化が起こることから、相転移が起こる下限

$$K_c > \frac{1}{2\zeta(p)} \quad (9.1.12)$$

が求まる。この結果は Griffiths[19] と Dyson[20] によって示されている下限 (9.1.12) と一致はする。しかし最近接相互作用また有限レンジでも1次元で相転移がおこってしまう。

9.2 Block Decimation くりこみ群

9.2.1 1次元イジングモデル

1次元最近接イジングモデルのハミルトニアンは

$$H = -K \sum_i \sigma_i \sigma_{i+1} - h \sum_j \sigma_j \quad (9.2.1)$$

と書きあらわせる。転送行列を用いて

$$T = \begin{pmatrix} e^{K+h} & e^{-K} \\ e^{-K} & e^{K-h} \end{pmatrix} \quad (9.2.2)$$

分配関数は、

$$Z = \sum_{\{\sigma\}} \exp \left(K \sum_i \sigma_i \sigma_{i+1} + h \sum_j \sigma_j \right) = \sum_{\{\sigma\}} \prod_i T_{\sigma_i \sigma_{i+1}} \quad (9.2.3)$$

と書くことができる。転送行列の2つの積は

$$\sum_b T_{ab} T_{bc} = T'_{ac} \quad (9.2.4)$$

となり、行列の積は**b**の添字の部分を積分することに相当するので、decimation くりこみ群に対応している。この関係式からくりこみ群方程式を導出することもできる。スピンの数 N 個の分配関数は、周期境界条件を課して以下のような計算できる。

$$Z_N = \text{Tr} T^{(N)} = \lambda_1^N \left(1 + \left(\frac{\lambda_2}{\lambda_1} \right)^N \right) \quad (9.2.5)$$

$$\begin{aligned} \lambda_1 &= e^K \cosh(h) + \sqrt{e^{2K} \cosh^2(h) - 2 \sinh(2K)} \\ \lambda_2 &= e^K \cosh(h) - \sqrt{e^{2K} \cosh^2(h) - 2 \sinh(2K)} \end{aligned} \quad (9.2.6)$$

熱力学極限を取る。自由エネルギー密度は

$$f = \frac{1}{N} \log(Z_\infty) = \log(\lambda_1) = \log \left(e^K \cosh(h) + \sqrt{e^{2K} \cosh^2(h) - 2 \sinh(2K)} \right) \quad (9.2.7)$$

求まる。

次に Green 関数を計算して相関長を求める。Green 関数は、1点関数と2点関数を用いて以下の式で定義される。

$$G(j) = \langle \sigma_0 \sigma_j \rangle - \langle \sigma_0 \rangle \langle \sigma_0 \rangle \quad (9.2.8)$$

ただし平行移動不変性を用いた。まず1点関数を計算する。

$$\langle \sigma_0 \rangle = \frac{1}{Z_N} \sum_{\{\sigma\}} [\sigma_0 T_{\sigma_0 \sigma_1} T_{\sigma_1 \sigma_2} \cdots T_{\sigma_{N-1} \sigma_N}] \quad (9.2.9)$$

スピン反転演算子は

$$\sigma = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad (9.2.10)$$

と書くことができる。転送行列の対角化行列 S を用いて

$$S^{-1}\sigma S = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (9.2.11)$$

と表すことにすると、

$$\langle \sigma_0 \rangle = \frac{1}{Z_N} \text{Tr} \left(S^{-1} \sigma S \begin{pmatrix} \lambda_1^N & 0 \\ 0 & \lambda_2^N \end{pmatrix} \right) = \frac{a\lambda_1^N + d\lambda_2^N}{\lambda_1^N + \lambda_2^N} \xrightarrow{N \rightarrow \infty} a \quad (9.2.12)$$

と計算できる。2点関数も同様に計算できる。

$$\langle \sigma_0 \sigma_j \rangle = \frac{1}{Z_N} \text{Tr} \left[\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \lambda_1^j & 0 \\ 0 & \lambda_2^j \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \lambda_1^{N-j} & 0 \\ 0 & \lambda_2^{N-j} \end{pmatrix} \right] \xrightarrow{N \rightarrow \infty} a^2 + bc \left(\frac{\lambda_2}{\lambda_1} \right)^j \quad (9.2.13)$$

よって Green 関数は

$$G(j) = \langle \sigma_0 \sigma_j \rangle - \langle \sigma_0 \rangle \langle \sigma_0 \rangle = bc \left(\frac{\lambda_2}{\lambda_1} \right)^j = bc \exp \left(-\frac{j}{\xi} \right) \quad (9.2.14)$$

と計算でき、相関長の定義から ($h = 0$)

$$\xi = \frac{1}{\log \left(\frac{\lambda_1}{\lambda_2} \right)} = \frac{1}{\log (\tanh K)} \quad (9.2.15)$$

と求まる。

9.2.2 1次元長距離イジングモデル

1次元長距離イジングモデルの場合も、同様に計算してみる。1次元長距離イジングモデルのハミルトニアンは以下の通りである。

$$H = -K \sum_n \sum_i \left(\frac{1}{n^p} \right) \sigma_i \sigma_{i+n} - h \sum_j \sigma_j \quad (9.2.16)$$

分配関数は、

$$\begin{aligned} Z_N &= \sum_{\{\sigma\}} \exp \left(K \sum_n \sum_i \left(\frac{1}{n^p} \right) \sigma_i \sigma_{i+n} + h \sum_j \sigma_j \right) \\ &= \sum_{\{\sigma\}} \prod_i T_{(\sigma_i \sigma_{i+1} \dots \sigma_{i+n})} (\sigma_{i+n+1} \sigma_{i+n+2} \dots \sigma_{i+2n}) \end{aligned} \quad (9.2.17)$$

と書くことができる。ただし相互作用の長さ n は有限とする。転送行列の2つの積は

$$\begin{aligned} \sum_{\sigma_{i+n+1}\sigma_{i+n+2}\dots\sigma_{i+2n}} T_{(\sigma_i\sigma_{i+1}\dots\sigma_{i+n})(\sigma_{i+n+1}\sigma_{i+n+2}\dots\sigma_{i+2n})} T_{(\sigma_{i+n+1}\sigma_{i+n+2}\dots\sigma_{i+2n})(\sigma_{i+2n+1}\sigma_{i+2n+2}\dots\sigma_{i+3n})} \\ = T'_{(\sigma_i\sigma_{i+1}\dots\sigma_{i+n})(\sigma_{i+2n+1}\sigma_{i+2n+2}\dots\sigma_{i+3n})} \end{aligned} \quad (9.2.18)$$

となり、これはブロック $\{\sigma_{i+n+1}\sigma_{i+n+2}\dots\sigma_{i+2n}\}$ を decimation している。この方法は、Block decimation くりこみ群 (BDRG) と呼ばれる [21]。BDRG を用いて物理量を評価する。転送行列を掛け合わせることで、すなわちくりこむことで分配関数が計算できる。

$$Z(K, h) = \text{Tr} T^{(k)}(K, h) \quad (9.2.19)$$

k はくりこみ回数である。

自由エネルギー密度は

$$f(K, h) = \frac{1}{2^k n} \log(Z(K, h)) \quad (9.2.20)$$

である。自由エネルギー密度を $h = 0$ の近傍でテイラー展開することによって

$$\begin{aligned} f(K, h) &= f(K, 0) + \frac{1}{2} \left. \frac{\partial^2 f(K, h)}{\partial h^2} \right|_{h=0} h^2 + \frac{1}{4!} \left. \frac{\partial^4 f(K, h)}{\partial h^4} \right|_{h=0} h^4 + \dots \\ &= f(K, 0) + \frac{1}{2} \chi h^2 + \frac{1}{4!} \chi_{nl} h^4 + \dots \end{aligned} \quad (9.2.21)$$

磁化率と非線形磁化率が求まる。有限レンジでは、対称性は破れないことから h の奇数乗は0である。 K に関して微分することで、エネルギー密度および比熱を求めることができる。くりこみ回数を $k = 17$ とした結果を図 9.1, 図 9.2, 図 9.3, 図 9.4 に示す。

次に熱力学極限を考える。ただし長距離相互作用がある場合は、転送行列が対称行列ではないので対角化はできないが、ジョルダン標準形には持っていくことができる。スピン数 N 個の分配関数は、周期境界条件を課して以下のような計算できる。

$$Z_N = \text{Tr} T^{(N/n)} = \lambda_1^{N/n} \left(1 + \left(\frac{\lambda_2}{\lambda_1} \right)^{N/n} + \dots \right) \quad (9.2.22)$$

熱力学極限を取ると自由エネルギー密度は

$$f = \frac{1}{N} \log(Z_\infty) = \frac{1}{n} \log(\lambda_1) \quad (9.2.23)$$

求まる。 K に関して微分することで、熱力学極限をとったエネルギー密度および比熱を求めることができる。その結果を図 9.5, 図 9.6 に示す。計算結果は、くりこみ回数を $k = 17$ とした結果と同じになる。よってくりこみ回数を $k = 17$ は体積無限大とみなすことができる。

Green 関数は

$$G(j) = \langle \sigma_0 \sigma_j \rangle - \langle \sigma_0 \rangle \langle \sigma_j \rangle \quad (9.2.24)$$

である。1次元長距離イジングモデルの場合は、転送行列に n 個のスピン要素がある。よってスピン反転演算子 σ は相互作用の長さの数 n だけある。

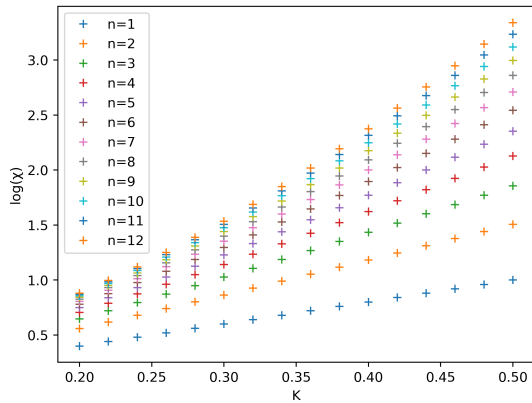


図 9.1: $p = 1.8$ における磁化率

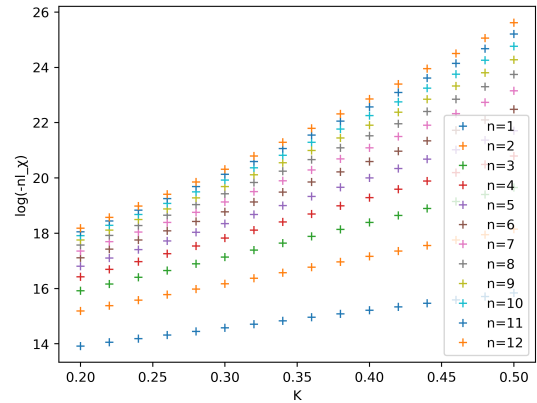


図 9.2: $p = 1.8$ における非線形磁化率

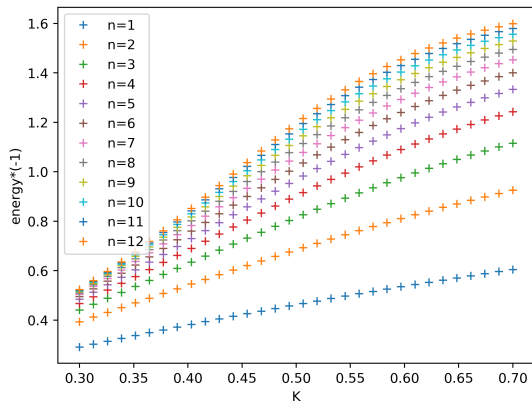


図 9.3: $p = 1.8$ におけるエネルギー密度

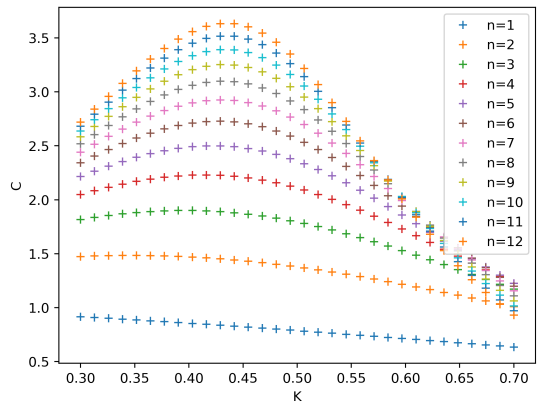


図 9.4: $p = 1.8$ における比熱

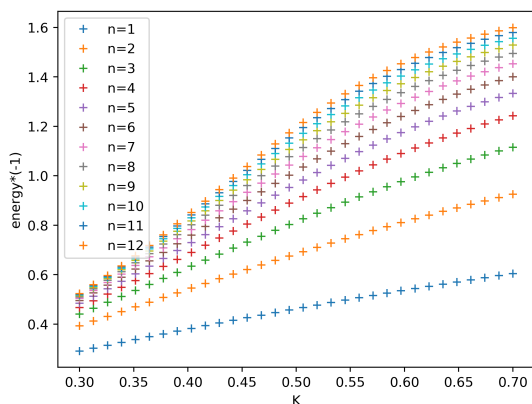


図 9.5: 熱力学極限、 $p = 1.8$ 、エネルギー密度

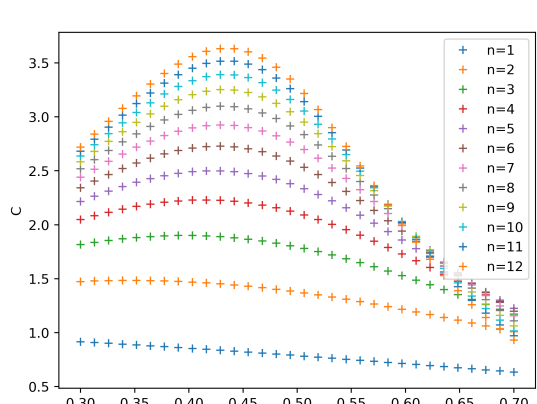


図 9.6: 熱力学極限、 $p = 1.8$ 、比熱

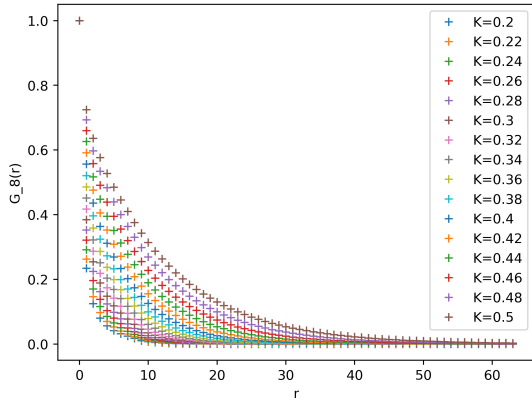


図 9.7: $p = 1.8, n = 8$ Green 関数

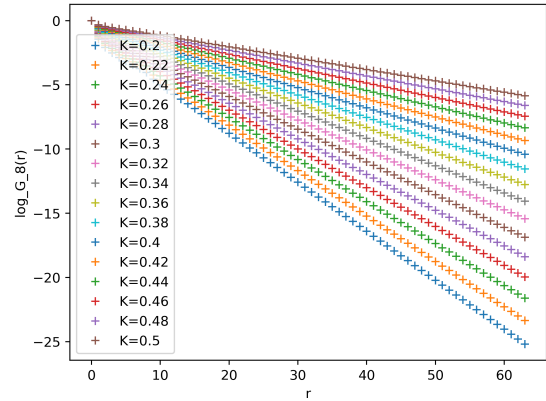


図 9.8: $p = 1.8, n = 8$ 対数 Green 関数

2点関数は転送行列を用いて

$$\langle \sigma_0 \sigma_j \rangle = \frac{1}{Z} \text{Tr} [\sigma_0 T^{(j)} \sigma_j T^{(N-j)}] \quad (9.2.25)$$

1点関数は

$$\langle \sigma_j \rangle = \frac{1}{Z} \text{Tr} [\sigma_j T^{(N)}] \quad (9.2.26)$$

となる。 N はくりこむ回数である。ここで不純物行列を

$$T_c = T \sigma_j T \quad (9.2.27)$$

とする。

スピン反転演算子は

$$\sigma = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

であり、最近接相互作用の場合は、 $0, j$ 番目にこの演算子を挟む。長距離相互作用の場合は、相互作用の長さ n とし、 j 番目のスピン演算子 $\sigma^{(j)}$ を

$$\sigma^{(j)} = \hat{1} \otimes \hat{1} \otimes \dots \otimes \sigma \otimes \dots \otimes \hat{1} \quad (9.2.28)$$

と書くことができる。 $\hat{1}$ は単位行列である。上付き添字は何番目にスピン演算子を挿入するか記述している。

くりこむ際に不純物行列を代入することで、Green 関数を計算することができる。その結果を図 9.7 と図 9.8 にしるす。

転送行列の対角化行列 S を用いて、 j 番目のスピン反転演算子を $\sigma^{(j)}$ とすると

$$[S^{-1} \sigma^{(j)} S]_{kl} = A_{kl}^{(j)} \quad (9.2.29)$$

と表すことにする。2点関数は

$$\begin{aligned} \langle \sigma_0 \sigma_j \rangle &= \frac{1}{Z} \left\{ A_{ab}^{(0)} \lambda_b^{j/n} \delta_{bc} A_{cd}^{(j)} \lambda_d^{(N-j)/n} \delta_{da} \right\} \\ &= \frac{1}{Z} \left\{ A_{ab}^{(0)} A_{ba}^{(j)} \lambda_b^{j/n} \lambda_a^{(N-j)/n} \right\} \\ &\xrightarrow{N \rightarrow \infty} A_{1b}^{(0)} A_{b1}^{(j)} \left(\frac{\lambda_b}{\lambda_1} \right)^{j/n} \end{aligned} \quad (9.2.30)$$

となる。添え字 b は1 から n まで走る。また転送行列の最大固有値は λ_1 とする。

1点関数も同様に

$$\begin{aligned} \langle \sigma_j \rangle &= \frac{1}{Z} \left\{ A_{ab}^{(j)} \lambda_b^{N/n} \delta_{ba} \right\} \\ &\xrightarrow{N \rightarrow \infty} A_{11}^{(j)} \end{aligned} \quad (9.2.31)$$

求めることができ Green 関数は

$$G(j) = \sum_{b=2}^n A_{1b}^{(0)} A_{b1}^{(j)} \left(\frac{\lambda_b}{\lambda_1} \right)^{j/n} \quad (9.2.32)$$

と求めることができる。

以上の議論から物理量を評価できるが、相互作用の長さ $n \rightarrow \infty$ として計算し相転移の情報を引き出すことは困難である。ただし、有限レンジスケーリング仮説を用いることで、有限レンジ n の情報から無限レンジ $n \rightarrow \infty$ の情報を得て相転移温度を評価することができる。

9.2.3 有限レンジスケーリング仮説

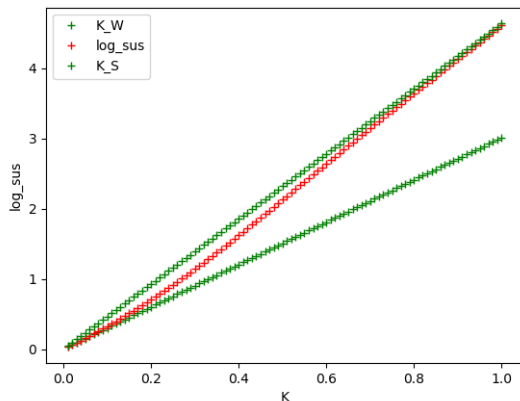


図 9.9: $p = 1.8, n = 4$ における境界

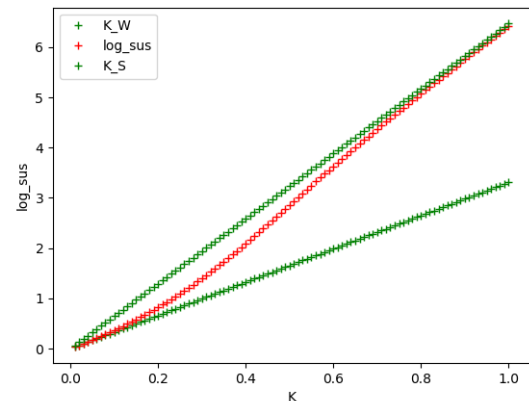


図 9.10: $p = 1.8, n = 8$ における境界

1次元長距離イジングモデルでは、体積以外にも相互作用の長さすなわちレンジを無限大にしなければならない。有限レンジ n の情報から無限レンジ $n \rightarrow \infty$ の情報を得る

方法を有限レンジスケーリング仮説 (FRS) と言う [22]。次の不等式が成り立つことを仮定する。この不等式は有限レンジの場合、数値的には成立している。その結果を図 9.9、図 9.10 に示す。

$$2K_{eff}^{[W]} \leq \log \chi \leq 2K_{eff}^{[S]} \quad (9.2.33)$$

$$K_{eff}^{[W]} \equiv \sum_n \frac{K}{n^p}, \quad K_{eff}^{[S]} \equiv \sum_n n \frac{K}{n^p} \quad (9.2.34)$$

この不等式はゼータ関数を用いて記述できる。

$$2K\zeta(p) \leq \log \chi \leq 2K\zeta(p-1) \quad (9.2.35)$$

次に磁化率のレンジの差を以下の量で定義する。

$$\Delta(n, p, K) \equiv \frac{1}{2K} (\log \chi(n) - \log \chi(n-1)) \quad (9.2.36)$$

式 (9.2.36) は、境界を満たす。

$$\Delta^{[W]} \leq \Delta(n, p, K) \leq \Delta^{[S]} \quad (9.2.37)$$

$$\Delta^{[W]} = \left(\frac{1}{n}\right)^p, \quad \Delta^{[S]} = \left(\frac{1}{n}\right)^{p-1}$$

ここで $\beta(n, p, K)$ を以下の量で定義する。

$$\Delta(n, p, K) \equiv \left(\frac{1}{n}\right)^{\beta(n, p, K)}$$

式 (9.2.38) から

$$p-1 \leq \beta(\infty, p, K) \leq p \quad (9.2.38)$$

が導ける。よって磁化率の対数を取ったものは、ゼータ関数で書くことができる。

$$\lim_{n \rightarrow \infty} \log \chi(n) \propto \zeta(\beta(\infty, p, K)) \quad (9.2.39)$$

磁化率は実関数であるので、 $\beta(z)$ は $z > 1$ では有限の値を持つが、 $z \rightarrow 1_+$ では発散する。つまり

$$\beta(\infty, p, K_c) = 1 \quad (9.2.40)$$

を満たす K_c が臨界点である。 $\beta(\infty, p, K)$ は

$$\beta(\infty, p, K) = \lim_{n \rightarrow \infty} -\frac{\log \Delta(n, p, K)}{\log(n)} \quad (9.2.41)$$

である。

臨界点である K_c を数値的に求める。 $\beta(n, p, K)$ を n の微小変化量で定義する。

$$\beta(n, p, K) \equiv -\frac{\log \Delta(n, p, K) - \log \Delta(n-1, p, K)}{\log(n) - \log(n-1)} \quad (9.2.42)$$

これらの結果を図 9.11 に示す。図 9.11 はくりこみ回数 $k = 17$ の場合であり、 $K = 0.41$ において全ての n が $\beta(n, p, K = 0.41) = 1$ となる。よってこの $p = 1.8$ の場合 $K = 0.41$ が相転移温度である。

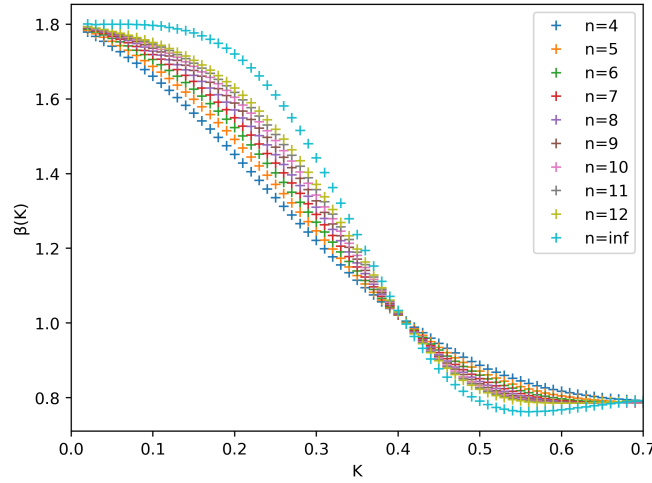


図 9.11: $p = 1.8$ における $\beta(n, p, K)$

9.3 勾配法について

9.3.1 誤差逆伝播法

勾配降下法を用いてパラメータを更新していく際、誤差関数 $E(\omega)$ の微分係数

$$\nabla_{\omega} E(\omega) = \frac{\partial E(\omega)}{\partial \omega} \quad (9.3.1)$$

チェインルールを用いて

$$\frac{\partial E(\omega)}{\partial \omega_{ij}} = \sum_k \frac{\partial E(\omega)}{\partial \hat{y}_k} \frac{\partial \hat{y}_k}{\partial \omega_{ij}} \quad (9.3.2)$$

を計算すると \hat{y} は

$$\hat{y}(\mathbf{W}, \mathbf{x}) = f^{(L)}(\mathbf{W}^{(L)} f^{(L-1)}(\mathbf{W}^{(L-1)} f^{(L-2)}(\dots \mathbf{W}^{(2)} f^{(1)}(x))) \quad (9.3.3)$$

であるので、層が深くなるとこの方法では合成関数の微分が複雑となる。

ここでニューラルネットワークでは

$$u^{(l)} = W^{(l)} z^{(l-1)} \quad (9.3.4)$$

である事を用いて微分係数の計算を考える。

チェインルールを用いて

$$\frac{\partial E(\omega)}{\partial \omega_{ij}^{(l)}} = \sum_k \frac{\partial E(\omega)}{\partial u_k^{(l)}} \frac{\partial u_k^{(l)}}{\partial \omega_{ij}^{(l)}} = \sum_k \sum_m \frac{\partial E(\omega)}{\partial u_k^{(l)}} \frac{\partial \omega_{km}^{(l)}}{\partial \omega_{ij}^{(l)}} z_m^{(l-1)} \equiv \delta_i^{(l)} z_j^{(l-1)} \quad (9.3.5)$$

となる。さらに $\delta^{(l)}$ は

$$\delta_i^{(l)} = \sum_k \frac{\partial E(\omega)}{\partial u_k^{(l+1)}} \frac{\partial u_k^{(l+1)}}{\partial u_i^{(l)}} = \sum_k \delta_k^{(l+1)} \omega_{ki}^{(l+1)} \frac{\partial z_i^{(l)}}{\partial u_i^{(l)}} \quad (9.3.6)$$

漸化式となる。漸化式 (9.3.6) を見ると l 層から $l+1$ 層と方程式が解かれていく。方程式の解く方向が、出力側から入力側に向かって解かれている。この方法を誤差逆伝播法 [23][24] という。

誤差逆伝播法を用いて漸化式を解くために、初期値 $\delta^{(L)}$ を計算すると出力は

$$\hat{y}_i = z_i^{(L)} = f^{(L)}\left(u_i^{(L)}\right) \quad (9.3.7)$$

なので

$$\delta_i^{(L)} = \sum_m \frac{\partial E(\omega)}{\partial \hat{y}_m} \frac{\partial \hat{y}_m}{\partial u_i^{(L)}} \quad (9.3.8)$$

である。誤差関数をクロスエントロピーとすると、出力はソフトマックス関数なので

$$\begin{aligned} \delta_i^{(L)} &= \sum_k \sum_m \frac{\partial (t_k \log(\hat{y}_k))}{\partial \hat{y}_m} \frac{\partial}{\partial u_i^{(L)}} \left(\frac{\exp(u_m^{(L)})}{\sum_l \exp(u_l^{(L)})} \right) \\ &= \sum_k \frac{t_k}{\hat{y}_k} (\delta_{ik} \hat{y}_k - \hat{y}_i \hat{y}_k) \\ &= t_i - \hat{y}_i \end{aligned} \quad (9.3.9)$$

となる。 $\sum_k t_k = 1$ を用いた。 $\delta^{(L)}$ は最終出力と訓練データの出力の差である。もし $\delta^{(L)}$ が大きい場合、勾配は大きくなるように寄与する。

数値微分は十分小さな値 ϵ として

$$\frac{\partial E(\omega)}{\partial \omega_i} \approx \frac{E(\omega_1, \dots, \omega_i + \epsilon, \dots) - E(\omega_1, \dots, \omega_i, \dots)}{\epsilon} \quad (9.3.10)$$

と書くことができる。誤差関数の微分は数値微分でも計算できるが、計算の回数は(重みの数+1)×(サンプル数)である。誤差逆伝播法の場合、一回の計算で出力値 $z^{(l)}$ が求まり1回の誤差伝播で $\delta^{(l)}$ が求まる。計算の回数は2×(サンプル数)であるので数値微分で求めるより計算量が少ないことがわかる。

9.3.2 勾配消失問題・勾配爆発問題

誤差逆伝播法は効率よく勾配を計算することができる。しかし深層のため勾配消失・爆発問題が生じる。漸化式 (9.3.6) は

$$\begin{aligned} \delta_i^{(l)} &= \sum_k \delta_k^{(l+1)} \omega_{ki}^{(l+1)} \frac{\partial z_i^{(l)}}{\partial u_i^{(l)}} \\ &= \sum_q \sum_k \delta_q^{(l+2)} \omega_{qk}^{(l+2)} \frac{\partial z_k^{(l+1)}}{\partial u_k^{(l+1)}} \omega_{ki}^{(l+1)} \frac{\partial z_i^{(l)}}{\partial u_i^{(l)}} \\ &= \sum_a \dots \sum_q \sum_k \delta_a^{(L)} \omega_{ab}^{(L)} \frac{\partial z_b^{(L-1)}}{\partial u_b^{(L-1)}} \dots \omega_{qk}^{(l+2)} \frac{\partial z_k^{(l+1)}}{\partial u_k^{(l+1)}} \omega_{ki}^{(l+1)} \frac{\partial z_i^{(l)}}{\partial u_i^{(l)}} \\ &= \sum_a \dots \sum_q \sum_k \delta_a^{(L)} \omega_{ab}^{(L)} f'(u_b^{(L-1)}) \dots \omega_{qk}^{(l+2)} f'(u_k^{(l+1)}) \omega_{ki}^{(l+1)} f'(u_i^{(l)}) \end{aligned} \quad (9.3.11)$$

とかける。重み ω の寄与は一旦無視すると出力層 L から中間層 l まで逆伝播したとすると

$$\delta^{(l)} \propto (f'(u))^{(L-l)} \quad (9.3.12)$$

となる。活性化関数の微分 $f'(u)$ が小さい値であると冪的な振る舞いから、 $\delta^{(l)}$ は小さい値となり勾配が更新されない。この問題を勾配消失問題、その逆を勾配爆発問題という。

ReLU 関数

以上の問題を解決するため、正規化線形関数（ReLU 関数）を用いる。

$$f(x) = \max\{0, x\} = \begin{cases} x & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

この関数の微分は

$$f'(x) = \begin{cases} 1 & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

である。漸化式 (9.3.6) は

$$\delta_i^{(l)} = \sum_a \cdots \sum_q \sum_k \delta_a^{(L)} \omega_{ab}^{(L)} \cdots \omega_{qk}^{(l+2)} \omega_{ki}^{(l+1)} \quad (9.3.13)$$

となるので、活性化関数に関しては勾配消失・爆発問題を解決することができる。

He の初期化

以上の議論では重みの寄与を無視して行った。重みの値によっては $\delta^{(l)}$ が小さくなり（大きくなり）勾配消失・爆発問題が起きる。He の初期化 [14] [25] を用いる事で、式 (9.3.13) の重みの積は 1 にできる。

He の初期化は分散 $2/d_l$ 、平均 0 のガウス分布から重み $\omega^{(l)}$ をサンプルする。 d_l は l 層における重みの数である。

$$\omega^{(l)} \sim N\left(\omega^{(l)}; 0, \sqrt{\frac{2}{d_l}}\right) \quad (9.3.14)$$

分散 $2/d_l$ である必要性について説明する。 $\omega^{(l)}$ と $u^{(l)}$ は互いに独立で確率変数とする。平均および分散は

$$\begin{aligned} E[u_i^{(l)}] &= \sum_j E[\omega_{ij}^{(l)} z_j^{(l-1)}] = d_l E[\omega^{(l)}] E[z^{(l-1)}] = 0 \\ V[u_i^{(l)}] &= \sum_j V[\omega_{ij}^{(l)} z_j^{(l-1)}] = d_l V[\omega^{(l)}] V[z^{(l-1)}] \end{aligned} \quad (9.3.15)$$

である。また $V[z^{(l-1)}]$ は、活性化関数を ReLU 関数とすると

$$\begin{aligned} V[z^{(l-1)}] &= E[(z^{(l-1)})^2] - (E[z^{(l-1)}])^2 \\ &= \sum_{u^{(l-1)}} P(u^{(l-1)}) (\max(0, u^{(l-1)}))^2 \\ &= \sum_{u^{(l-1)} \geq 0} P(u^{(l-1)}) (u^{(l-1)})^2 \\ &= \frac{1}{2} \sum_{u^{(l-1)}} P(u^{(l-1)}) (u^{(l-1)})^2 = \frac{1}{2} V[u^{(l-1)}] \end{aligned} \quad (9.3.16)$$

となる。出力層 L おける $u^{(L)}$ の分散は

$$V[u^{(L)}] = V[u^{(1)}] \prod_{l=2}^L \frac{d_l}{2} V[\omega^{(l)}] \quad (9.3.17)$$

となる。 $u^{(L)}$ の分散は、層が深くなれば大きくなる。よって

$$\frac{d_l}{2} V[\omega^{(l)}] \approx 1 \quad (9.3.18)$$

と取れば分散が大きくなる。

9.3.3 勾配降下法

誤差関数を最小化することによって、最適なパラメータ ω^*

$$\omega^* = \arg \min_{\omega} E(\omega) \quad (9.3.19)$$

を見つけなければならない。(9.3.19) を数値的に解く手法として勾配降下法がある。

勾配降下法は、ステップ数 t における $\omega^{(t)}$ を

$$\begin{aligned} \omega^{(t+1)} &\leftarrow \omega^{(t)} + \Delta\omega^{(t)} \\ \Delta\omega^{(t)} &= -\eta \nabla E(\omega^{(t)}) \end{aligned} \quad (9.3.20)$$

となるように更新していく。 η は学習率または、ハイパーパラメータと呼ばれる。勾配 $\Delta\omega^{(t)}$ に沿って $\omega^{(t)}$ を ω^* に近づけていく。勾配が $\Delta\omega^{(t)} = 0$ となる ω が ω^* である。

確率的勾配降下法

訓練データ $D = \{(x_n, y_n)\}_{n=1, \dots, N}$ があり、誤差関数を

$$E(\omega) = \sum_{n=1}^N E_n(\omega) \quad (9.3.21)$$

と表すことにする。全てのサンプル数 N を用いて学習させる方法をバッチ学習という。

訓練データの1部を取り出したサンプルを $B^{(t)}$ とする。上付き添字の (t) は、 $\omega^{(t)}$ を更新していく際に用いる訓練データを表す。 $B^{(t)}$ をミニバッチと呼び、 $B^{(t)}$ を使って学習する事をミニバッチ学習という。ミニバッチは学習前にランダムに生成する。

あるステップ t における誤差関数を

$$E^{(t)}(\omega) = \frac{1}{|B^{(t)}|} \sum_{n \in B^{(t)}} E_n(\omega) \quad (9.3.22)$$

として学習を行う。ミニバッチによってランダムに選ばれた訓練データを用いて勾配降下法を行うので確率的勾配降下法 [14] と呼ぶ。

ミニバッチを用いることで、ステップごとに $E^{(t)}(\omega)$ が変化する。その変化によって、誤差関数の最小値から離れた極小値に引っかかる可能性が減る。また訓練データのサイズが大きくなると、似たようなサンプルが増える。ミニバッチを用いることで、似たようなデータを重複して使用する無駄が省かれる。

9.3.4 勾配降下法の問題点

局所的最適解の問題

深層学習の誤差関数は一般的に、非凸関数であるとされる。非凸関数には多数の極小値があり、学習の過程において極小値にひかかってしまい最小値にたどりつくことが困難となる。この問題を局所的最適解の問題という。深層学習は、最小値に辿り着けなくても極小値であればいいとされており、これは深層学習の未解決の問題である。ただし、簡単なモデルにおいて極小値は最小値となることが知られている [26][27]。

誤差関数の形における問題点

- 誤差関数が深い谷のような形をしていたとする。勾配が急であるのでパラメータの更新が大きくなり、谷の反対側に移動してしまう。次のパラメータの更新も同様なので、振動してしまい学習が進まなくなる。
- 誤差関数が高原のような形をしていたとする。これはプラトーと呼ばれる。勾配が平坦に近いので、パラメータの更新が遅くなり学習が進まなくなる。
- 誤差関数が絶壁のような形をしていたとする。絶壁にさしかかると勾配が急になるので、反射して前の状態に戻ってしまう。パラメータを更新していくと同じ絶壁にさしかかるので学習が進まなくなる。
- 以上ではパラメータは1次元方向のみ考えてきた。多次元空間を考えた場合鞍点が存在する。鞍点にさしかかるとある方向ではパラメータを更新したいが、他の方向では勾配が0となって学習が進まなくなる。

9.3.5 モーメンタム法

誤差関数が深い谷のような形をしていると、急な勾配の正負の入れ替えによって振動してしまい学習が進まなくなる。この問題を解決するために前の勾配の情報を取り入れることによって振動を抑える方法があり、その方法をモーメンタム法 [14] という。パラメータの更新は以下のように行う。

$$\begin{aligned}\omega^{(t+1)} &\leftarrow \omega^{(t)} + \Delta\omega^{(t)} \\ \Delta\omega^{(t)} &= \mu\Delta\omega^{(t-1)} - (1 - \mu)\eta\nabla E(\omega^{(t)})\end{aligned}\tag{9.3.23}$$

$\mu = 0$ の場合は勾配降下法である。前の情報 $\Delta\omega^{(t-1)}$ を取り込むことで、急な勾配の正負の入れ替えが生じにくくなる。また、勾配があまり変化しない場合

$$\begin{aligned}\Delta\omega^{(t)} &\approx \Delta\omega^{(t+1)} \\ \Delta\omega^{(t)} &= -\eta\nabla E(\omega^{(t)})\end{aligned}\tag{9.3.24}$$

となり、平坦なところでは通常の勾配降下法になる。

9.3.6 AdaGrad

勾配降下法ではハイパーパラメータ η が1つのみである。最適化したいパラメータ ω は多次元空間であり、誤差関数は各方向によって勾配が違う可能性が高い。例えば鞍点があると一方の方向では勾配が無くなるので学習が進まなくなる。よって各方向に異なるハイパーパラメータを導入する必要があるが、多くのハイパーパラメータの値をトライアンドエラーで決めることは困難である。

AdaGrad(adaptive subgradient descent)[14][28] は過去の勾配の情報を取り入れ、パラメータの更新は以下のように行う。

$$\Delta\omega_i^{(t)} = -\frac{\eta}{\sqrt{\sum_{s=1}^t (\nabla E(\omega^{(s)})_i)^2}} \nabla E(\omega^{(t)})_i\tag{9.3.25}$$

(9.3.25) の分母は、過去の情報によって各方向におけるハイパーパラメータを変化させる。例えば過去に大きな勾配をとってきた場合、ハイパーパラメータを減衰させる。しかし学習初期の勾配が大きくなるとハイパーパラメータが大きく減衰してしまい学習しなくなる問題が生じる。

9.3.7 Adam

Adam[14][29] はモーメンタム法と AdaGrad を組み合わせた方法である。AdaGrad の欠点を補うために、Adam は過去の勾配の情報を減衰因子 ρ によって冪的に減衰させる。

$$\begin{aligned}
 m_{i,t} &= \rho_1 m_{i,t-1} + (1 - \rho_1) \nabla E(\omega^{(t)})_i = (1 - \rho_1) \sum_{s=1}^t (\rho_1)^{t-s} \nabla E(\omega^{(s)})_i \\
 v_{i,t} &= \rho_2 v_{i,t-1} + (1 - \rho_2) (\nabla E(\omega^{(t)})_i)^2 = (1 - \rho_2) \sum_{s=1}^t (\rho_2)^{t-s} (\nabla E(\omega^{(s)})_i)^2 \\
 \Delta\omega_i^{(t)} &= -\frac{\eta}{\sqrt{v_{i,t} + \epsilon}} m_{i,t}
 \end{aligned} \tag{9.3.26}$$

ϵ は分母が0にならないように調節する。確率的勾配降下法を用いた場合、訓練データはランダムに抽出されるので、勾配もある確率分布に従う。初期値は $m_{i,0} = 0, v_{i,0} = 0$ とすると推定量 $m_{i,t}$ や $v_{i,t}$ が0に偏ってしまう。 $v_{i,t}$ の期待値を計算すると

$$\begin{aligned}
 E[v_{i,t}] &= (1 - \rho_2) \sum_{s=1}^t (\rho_2)^{t-s} E[(\nabla E(\omega^{(s)})_i)^2] \\
 &\approx E[(\nabla E(\omega^{(t)})_i)^2] (1 - \rho_2) \sum_{s=1}^t (\rho_2)^{t-s} \\
 &= (1 - (\rho_2)^t) E[(\nabla E(\omega^{(t)})_i)^2]
 \end{aligned} \tag{9.3.27}$$

となる。2行目の近似は、過去の情報はパラメータを上手く取ることにより抑えられたつまり $\omega^{(s)} \approx \omega^{(t)}$ とする。因子 $1 - (\rho_2)^t$ が推定量を0にする要因となる。よってこの因子で規格化を行う。

$$\begin{aligned}
 \hat{m}_{i,t} &= \frac{m_{i,t}}{1 - (\rho_1)^t} \\
 \hat{v}_{i,t} &= \frac{v_{i,t}}{1 - (\rho_2)^t} \\
 \Delta\omega_i^{(t)} &= -\frac{\eta}{\sqrt{\hat{v}_{i,t} + \epsilon}} \hat{m}_{i,t}
 \end{aligned} \tag{9.3.28}$$

式 (9.3.28) の勾配法を Adam とよばれる。推奨値があり

$$\eta = 10^{-3}, \rho_1 = 0.9, \rho_2 = 0.999, \epsilon = 10^{-8} \tag{9.3.29}$$

である。

9.4 Kadanoff 形式

ここでは Kadanoff 形式のくりこみ群について紹介する [30][31]。ハミルトニアンを $H_K(\sigma)$ とする。 K は結合定数、 σ は変数である。自由エネルギー $F(K)$ は

$$F(H) = -\ln \text{Tr}_\sigma e^{-H_K(\sigma)} \tag{9.4.1}$$

となる。次に新しいハミルトニアンをくりこみ変換を用いて定義する。新しい変数 μ を用いて $H'(\mu)$ を

$$e^{H'(\mu)} = \text{Tr}_\sigma e^{T(\mu,\sigma) - H_K(\sigma)} \tag{9.4.2}$$

と表すことができ、 $T(\mu, \sigma)$ はくりこみ変換関数である。くりこみ変換後の $H'(\mu)$ は一般的には複雑な形となる。くりこみ変換によって自由エネルギーは不変でなければならないので

$$\begin{aligned} F(H') &= -\ln \text{Tr}_\mu e^{H'(\mu)} = -\ln \text{Tr}_\mu \text{Tr}_\sigma e^{T(\mu, \sigma) - H_K(\sigma)} \\ &= F(H) \end{aligned} \quad (9.4.3)$$

よってくりこみ変換関数は以下の条件

$$\text{Tr}_\mu e^{T(\mu, \sigma)} = 1 \quad (9.4.4)$$

を満たさなければならない。この条件を Tr 条件と言うことにする。新しいハミルトニアンは形式的に

$$H' = R_T[H] \quad (9.4.5)$$

と表すことができる。 R は recursion function と呼ばれる。

また新しいハミルトニアンの結合定数は、新しい結合定数 K' を用いて表現できる。新しい結合定数と古い結合定数の関係はパラメータ p を用いて

$$K' = R_p[K] \quad (9.4.6)$$

と表すことができる。ここで変数を $\sigma \rightarrow \mu$ に変換した際、格子間隔が $a \rightarrow \lambda a$ に変化したとする。この場合の、くりこみ変換を以下のように定義する。

$$K' = R^\lambda[K] \quad (9.4.7)$$

R^{λ_1} は格子間隔が $a \rightarrow \lambda_1 a$ の変化を表現しているとし、 R^{λ_2} は格子間隔が $a \rightarrow \lambda_2 a$ の変化を表現しているとする、定義によって

$$K' = R^\lambda[K] = R^{\lambda_2}[R^{\lambda_1}[K]] \quad (9.4.8)$$

と書くことができる。格子間隔の変化は $\lambda = \lambda_2 \lambda_1$ である。

具体的に1次元最近接イジングモデルについて計算する。最近接相互作用のみのイジングモデルのハミルトニアンは

$$H(\sigma) = -K \sum_i \sigma_i \sigma_{i+1} \quad (9.4.9)$$

とかける。くりこみ変換関数を

$$e^{T(\mu, \sigma)} = \prod_{k:\text{even}} \delta(\mu_k - \sigma_k) \quad (9.4.10)$$

とする。簡単なため k は偶数の値をとるとする。新しいハミルトニアンは

$$\begin{aligned} e^{-H'(\mu)} &= \text{Tr}_\sigma \prod_{k:\text{even}} \delta(\mu_k - \sigma_k) \prod_i e^{K\sigma_i \sigma_{i+1}} \\ &= \text{Tr}_\sigma \prod_{i:\text{odd}} e^{K\sigma_i(\mu_{i-1} + \mu_{i+1})} \\ &= \prod_{i:\text{odd}} 2 \cosh(K(\mu_{i-1} + \mu_{i+1})) \end{aligned} \quad (9.4.11)$$

となり、これは Decimation くりこみ群である。

次にくりこみ変換関数を

$$T(\mu, \sigma) = p(\mu\sigma) + p(\mu\sigma') - u(\sigma\sigma') - C \quad (9.4.12)$$

とする。 p はフリーパラメータである。 C は定数、 u はパラメータであるが Tr 条件から決まる。

$$\begin{aligned} u(\sigma\sigma') + C &= \ln \text{Tr}_\mu e^{p(\mu\sigma) + p(\mu\sigma')} \\ &= \ln 2 \cosh(p(\sigma + \sigma')) \end{aligned} \quad (9.4.13)$$

簡単なため σ と σ' は偶数番目のイジング変数とする。新しいハミルトニアンは

$$\begin{aligned} e^{-H'(\mu)} &= \text{Tr}_\sigma \prod_{i=\text{odd}} e^{p(\mu_i\sigma_{i-1}) + p(\mu_i\sigma_{i+1}) - u(\sigma_{i-1}\sigma_{i+1}) - C} \prod_i e^{K\sigma_i\sigma_{i+1}} \\ &= \text{Tr}_\sigma \prod_{i=\text{odd}} e^{p(\mu_i\sigma_{i-1}) + p(\mu_i\sigma_{i+1})} [2 \cosh(p(\sigma_{i-1} + \sigma_{i+1}))]^{-1} [2 \cosh(K(\sigma_{i-1} + \sigma_{i+1}))] \end{aligned} \quad (9.4.14)$$

となり、フリーパラメータを $p = K$ とすると

$$e^{-H'(\mu)} = \prod_{i=\text{odd}} 2 \cosh(K(\mu_i + \mu_{i+2})) \quad (9.4.15)$$

Decimation くりこみ群の結果を再現する。

この方程式を解く。具体的に $\mu_i = \pm 1$ を代入していくと

$$\begin{aligned} K' \mu_i \mu_{i+1} + C &= \log(2 \cosh(K(\mu_i + \mu_{i+1}))) \\ K' + C &= \log(2 \cosh(2K)) \\ -K' + C &= \log(2) \end{aligned} \quad (9.4.16)$$

よってくりこみ変換後の結合定数 K は式変形を行い

$$\begin{aligned} K' &= \frac{1}{2} \log(\cosh(2K)) \\ \tanh(K') &= [\tanh(K)]^2 \end{aligned} \quad (9.4.17)$$

となる。ここで格子間隔 a を復活させると

$$\tanh(K(2a)) = [\tanh(K(a))]^2 \quad (9.4.18)$$

$\lambda - 1$ 個 decimation したものは式 (9.4.7) を用いて

$$\tanh(K(\lambda a)) = [\tanh(K(a))]^\lambda \quad (9.4.19)$$

と書くことができる。

参考文献

- [1] Y.Nagai, H.Shen, Y.Qi, J.Liu, and L.Fu, *Phy Rev B* **96**, 161102 (2017).
- [2] S.Iso, S.Shiba, and S.Yokoo, *Phys. Rev. E* **97**, 053304 (2018).
- [3] K.Aoki, and T.Kobayashi, *Mod. Phys. Lett. B* **30**,1650401(2016).
- [4] H.Fujita, Yuya.O. Nakagawa, S.Sugiura, and M.Oshikawa ,*Phys. Rev. B* **97**, 075114 (2018).
- [5] H.Saito, *J. Phys. Soc. Jpn.* **87**, 074002 (2018).
- [6] T.Mano, and T.Ohtsuki, *J. Phys. Soc.Jpn.* **86**, 113704 (2017).
- [7] Y.Fujimoto, K.Fukushima, and K.Murase, *Phys. Rev. D* **98**, 023019 (2018).
- [8] A.Tanaka, and A.Tomiya, arXiv:1712.03893.
- [9] Y. Mori, K.Kashiwa, and A. Ohnishi, *PTEP* **2018** 023B04 (2018) .
Y. Mori, K.Kashiwa, and A. Ohnishi, *Phys. Rev. D* **96** 111501(2017) .
K.Kashiwa, Y. Mori, and A. Ohnishi, arXiv:1805.08940.
- [10] K.Hashimoto, S.Sugishita, A.Tanaka,and A.Tomiya, *Phys. Rev. D* **98**, 046019 (2018).
- [11] P. Mehta, and D. J. Schwab, arXiv:1410.3831 (2014).
- [12] A.Tanaka, and A.Tomiya, *J. Phys. Soc. Jpn.* **86**, 063001 (2017).
- [13] S.Arai, M.Ohzeki, and K.Tanaka, *J. Phys. Soc. Jpn.* **87**, 033001 (2018) .
- [14] 瀧雅人, これならわかる深層学習入門
- [15] C. M. Bishop, *Pattern Recognition and Machine Learning*(Information Science and Statistics) (Springer, New York, 2006).
- [16] K.Fukushima, *Biological Cybernetics* **36**,193(1980).
- [17] Y. LeCun, B. Boser, J. S. Denker, and D. Henderson, *Neural computation* **1**,541(1989).
- [18] 高橋和孝, 西森秀稔, 相転移・臨海現象とくりこみ群

- [19] R.B.Griffiths , Commun. Math. Phys.**6**,121(1967).
- [20] F.J.Dyson, Commun. Math. Phys.**12**,91(1969).
- [21] 小林玉青, 博士論文 (2008)
- [22] K.Aoki, T.Kobayashi, and H.Tomita, Prog.Theor.Phys.**119**, 509(2008).
- [23] S. Amari, IEEE Trans. Electron. Comput. **EC-16**, 299 (1967).
- [24] D.E.Rumelhart, G.E.Hinton, and R.J.Williams ,Nature**323**,533(1986).
- [25] K.He, X.Zhang, S.Ren, and J.Sun, Proceedings of the WEEE International Conference on Computer Vision,(2015).
- [26] A.Choromanska, M.Henaff, M.Mathieu, G.B.Arous, and Y.LeCun, arXiv:1412.0233, (2014).
- [27] K. Kawaguchi, arXiv:1605.07110 (2016).
- [28] J.Duchi, E.Hazan, and Y.Singer. Journal of Machine Learning Research, **12**,2121(2011).
- [29] D.P.Kingma, and J.Ba, arXiv:1412.6980(2014).
- [30] L.P.Kadanoff, Annals Phys. **100**,359(1976).
- [31] L.P.Kadanoff, Rev.Mod.Phys.**49**,267(1977).

謝辞

研究室のスタッフの方々、学生、OB、OGの皆様には3年間大変お世話になりました。指導教員の青木健一教授には、理工学域長という大変忙しい職務の中、面倒をみてくださりとても感謝しております。研究で培われたものは、今後役に立つと思います。共同研究者の小林玉青准教授には、高等専門学校の職務に追われ研究に時間が割けない中、私と議論してくださりました。熊本さん、小内さんには、専門である対称性の破れ・弱解について、他にも研究に限らず様々なことを教えてくださりました。堀くんにはpythonのプログラミングを教えていただきました。また堀くんにはこの研究の基盤を作ってくれたおかげで今の研究成果があります。

河内さん、坂井くんには研究室について、また数値計算、特にモンテカルロ法およびLinux・クラスタコンピュータの使い方を丁寧に教えてくださりました。お二方がおられなければ研究はできなかったと思います。神宮くん、飯島さんにはゼミで議論をしてくれました。久保治輔客員教授、末松大二郎教授、青木真由美准教授、武田真滋准教授そして石渡弘治助教には、様々な話をしていただきました。西川涼子さんには様々な書類管理・申請等をしていただきました。

またA.SPACE社長の小波さんまた娘さん達には、議論する場および親睦の場を設けてもらいとても感謝しております。高確率で現れるネコにはとても癒されました。最後に今まで育ててくれた家族に深い感謝の意を表して謝辞といたします。