

電子・情報工学講座

17

情報工学基礎論

田村進一 著

培風館

3.4	記憶のある情報源	18
3.5	一般化されたマルコフ情報源	19
3.6	状態の分類	21
3.7	マルコフ情報源の定常分布	22

4. 情報源符号化

25

4.1	離散的情報源の符号化	25
4.2	瞬時符号であるための必要十分条件	27
4.3	符号の木	28
4.4	クラフトの不等式	29
4.5	平均符号長の限界	31
4.6	ハフマン符号化	34
4.7	ブロック符号化	38
4.8	情報源符号化定理	39
4.9	非等長情報源記号系列に対するハフマン符号化	42

5. 離散的通信路

45

5.1	通信路モデル	45
5.2	記憶のない定常通信路	46
5.3	相互情報量	48
5.4	通信路容量	50
5.5	記憶のない一様通信路の通信路容量	51
5.6	情報伝送速度	53
5.7	通信路符号化定理	54

6. 通信路符号化法

58

6.1	組織符号	58
6.2	単一パリティ検査符号	59
6.3	水平垂直パリティ検査符号	60
6.4	線形符号	61
6.5	ハミング距離	61
6.6	線形符号の加法性	62
6.7	最小距離・最小重みと誤り訂正能力	62

6.8	パリティ検査方程式	64
6.9	単一誤り訂正ハミング符号	66
6.10	バースト誤り	67
6.11	巡回符号	68
6.12	巡回ハミング符号	75

7. 標本化定理

78

7.1	簡単な例	78
7.2	フーリエ級数	80
7.3	フーリエ変換	83
7.4	標本化定理の導出	84
7.5	デルタ関数	86
7.6	インパルス応答	88
7.7	自己相関関数とパワースペクトル	90
7.8	標本値列からの復元	90

8. アナログ通信路

96

8.1	確率密度関数	96
8.2	アナログ情報源のエントロピー	98
8.3	最大エントロピー定理	100
8.4	通信路容量	101
8.5	量子化雑音	103
8.6	ベクトル量子化	104

9. 統計的パターン認識

107

9.1	多次元ガウス分布	107
9.2	統計的決定法	109

10. 命題論理

115

10.1	論理式	115
10.2	論理式の構成法	117
10.3	論理式の解釈	117
10.4	意味の木	123

10.5 命題論理式の標準形	124
11. 述語論理	127
11.1 論理式の構成法	127
11.2 恒真性と充足性	129
11.3 節 形	130
11.4 導 出 節	132
11.5 単 一 化	134
11.6 導出原理	136
12. ニューラルネットワーク計算	141
12.1 ニューラルネットワークの形態	141
12.2 階層型ネットワークと誤差逆伝搬学習アルゴリズム	142
12.3 相互結合型ネットワーク	146
付 録	155
A 1 デイビス・パットナムのアルゴリズム	155
A 2 命題論理の公理系	157
A 3 エルブラン空間	159
文 献	163
演習問題解答	165
付 表	183
索 引	187

1

序 論

本章では本書の内容の概略について述べる。

1.7 節までは本書の8章までの内容であり、主として情報理論と信号理論の話である。情報理論は、情報伝達および蓄積の効率化、信頼性向上の基礎理論となるものである。これらは情報の内容や意味については立ち入らず、主として統計的性質に着目して、情報を発生・伝送・変換するときの定量的な諸性質を議論するものである。1.8 節以降は9章以降の内容であり、知的情報処理の基礎理論を述べている。これらはパターン認識、知識表現と推論、およびニューラルネットワークに関するものであり、情報をその意味内容にまで立ち入って議論する。これらは人工知能や知能情報の基礎理論となるものである。

1.1 情報理論とは

情報理論は情報伝達の効率化、高信頼化のための符号化の理論である。現代の情報理論の中心はシャノン理論である。シャノン理論は情報の量的表示および符号化の限界に関する理論である。

日常生活において、ある日の株価を知るとはAさんにとっては非常に値打ちのあることであろうし、Bさんにとってはまったく意味がないものであるかも知れない。このように、情報の価値は受け手により大きく変わってくるといえるであろう。このような情報の価値まで考慮した定量化は、重要ではあっても難しい。それに対して、シャノンは対象の統計的な性質に着目し、その定量化を行った。たとえば通信線を通して英文が送られてくる状況を考えてみよう。ここでは、英文の内容には立ち入らずにその文字や単語の出現頻度のみに

1.10 ニューラルネットワーク

ニューラルネットワークは生体の神経回路を模した回路で各種の情報処理を行うとするものである。論理学の体系が堅いのに対して、ニューラルネットワークの研究は人間のような柔らかな論理に従って柔軟に処理・行動を行う機械を作ることをめざしている。

ニューラルネットワークは多数の神経素子と、それらの間を結ぶ重み付き配線からなっている。それらの重みを変えることにより、目的の処理に適した回路を構成することができる。すなわち、人間のように学習することが可能であることがニューラルネットワークの大きな特徴である。これにより、プログラミングなしにさまざまな情報処理を行うことが可能になるであろうとの期待がかけられている。

ニューラルネットワークの代表的形態である階層型ネットワークでは、入力パターンから出力パターンへの任意の変換が可能であることが知られている。したがってこれは、画像の輪郭を抽出したり、ボケを取り除いたりするフィルタリングやパターン変換器に使うことができ、またその特殊な場合であるパターン分類やパターン認識に使うことができる。このネットワークは、多くのことを覚えすぎると新しいことを覚えられないなど人間の脳に似た性質をもつ。

もう一つのニューラルネットワークの代表的形態である相互結合型ネットワークはその基本的機能としてパターンの連想機能をもつ。たとえば覚え込んだ人の鼻写真を見て、顔全体を出力する機能である。これを利用すると、巡回セールスマン問題など、計算量的に問題の多い組合せ問題も人間のように近似解を速く求めることができる。

2 情報量

本章では、まず事象間の演算、およびその上で定義される確率の基本的な性質について述べる。これらの準備のもとに、事象系のもつ情報量、およびあいまいさを表わすエントロピーを定義する。

2.1 事象と確率

世の中には「今日は晴れている」や「硬貨を投げて裏が出る」のように、さまざまな事象(event; 出来事)が起こっている。

一般に人が感じる情報とは、自分や周囲の人に利害をもたらしたり、関心のある事柄が重要な情報となる。たとえば、まったく知らない人が宝くじに当たったという情報よりも、自分の家族が宝くじに当たったという情報のほうがはるかに大きい情報をもたらす。また、興味のない分野のニュースよりも関心の大きな分野のニュースのほうがより大きな情報をもたらす。しかしながら、このような受け手の状態や主観に依存する測り方では客観的な議論はできない。

シャノン(Shannon)に始まる“情報理論”では、確率的に定義することのできる事象の集合の上で客観的な情報の量の測り方を定義し、それに基づく情報伝達(通信)の理論を与えている。

任意の二つの事象を E, F とする。これらの間に次のような集合論的演算を定義することができる。

- (1) “ E と F の少なくとも一方が起こる”という事象を E と F の和事象といい、 $E \cup F$ で表わす。

(2) “ E と F がともに起こる” という事象を E と F の積事象といい、 $E \cap F$ で表わす。

(3) “ E が起こらない” という事象を E の余事象といい、 $\bar{E}$ で表わす。

基本的な事象(根元事象)から上のような積和、および余事象演算を何度か行つて、合成事象を導くことができる。さらに、つぎのような集合論的包含関係を定義できる。

(4) “ E が起これば必ず F も起こる” とき、それを $E \subset F$ で表わす。

【例 2.1】 A, B, C を事象とすると、 $\overline{(A \cup B)} \cap C$ も事象である。

根元事象を一個の要素からなる集合とみたとき、上の合成演算によって得られる合成事象は集合となる。したがって、事象には確率が定義できる。確率 $P(E)$ は事象 E の起こりやすさを示す値であり、つぎの(i)~(iii)の性質をもつ。なお、このような性質(コルモゴロフ(Kolmogorov)の公理)をもつ量は、より一般には測度(measure)と呼ばれる。

E, F を全事象 U に含まれる任意の事象、 ϕ を空事象とする。

- (i) $0 \leq P(E) \leq 1, \quad P(\phi) = 0, \quad P(U) = 1$
- (ii) $E \subset F$ ならば、 $P(E) \leq P(F)$ 。
- (iii) $E \cap F = \phi$ ならば、 $P(E \cup F) = P(E) + P(F)$ 。

【例 2.2】 サイコロを投げたとき、出た目の数が4またはそれ以下になる確率を求める。

この場合の全事象は $U = \{1, 2, 3, 4, 5, 6\}$ である。4またはそれ以下になる事象は $E = \{1\} \cup \{2\} \cup \{3\} \cup \{4\} = \{1, 2, 3, 4\}$ であるから、 $P(E) = 2/3$ である。

2.2 条件付き確率

いま、100 人の人(U)が自動車運転免許の試験を受けたとする。このうち40人(Y)が合格し、60人(N)が不合格であった。合格者のうちの32人、および不合格者のうちの40人は自動車学校の卒業生(G)であった。自動車学校の卒業生の合格率はいくらであろうか。

集合 E の要素の数を $n(E)$ で表わすと、図 2.1 のベン図 (Venn diagram) より、答は

$$P(Y|G) = \frac{n(Y \cap G)}{n(G)} = \frac{P(Y \cap G)}{P(G)} = \frac{32}{72} = \frac{4}{9}$$

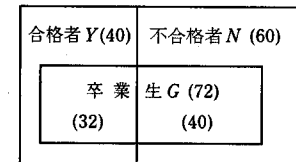


図 2.1 ベン図。()内は人数

となる。

なお一般に、

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (2.1)$$

は、事象 B が起こったとき、事象 A が起こる確率であり、条件付き確率と呼ばれる。また、 $P(A \cap B)$ は A と B の積事象、すなわち A と B が同時に起こる確率であり、 A と B の結合確率と呼ばれる。同様に

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \quad (2.2)$$

である。式(2.1)の分子に式(2.2)の分子を代入すれば、一般に

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2.3)$$

が成り立つ。これは、ベイズの定理(Bayes' theorem)と呼ばれる。事象 A が起こる確率は事前には $P(A)$ であるが、事象 B が起こったことを知った後は、その確率は $P(A|B)$ に変化する。そのような意味で $P(A)$ は事前確率、 $P(A|B)$ は事後確率と呼ばれることもある。それらの間の変換を与えるのがベイズの定理である。

また、 U が互いに素な n 個の事象 $A_1, A_2, \dots, A_n$ からなるとき、すなわち、 $U = A_1 \cup A_2 \cup \dots \cup A_n = \sum_{i=1}^n A_i$ のとき、

$$P(B) = \sum_{i=1}^n P(A_i \cap B) = \sum_{i=1}^n P(B|A_i)P(A_i)$$

であるから、ベイズの定理は

$$P(A_j|B) = \frac{P(B|A_j)P(A_j)}{\sum_{i=1}^n P(B|A_i)P(A_i)} \quad (2.4)$$

のように表現できる。なお、この式の右辺の分母は $\sum_{j=1}^n P(A_j|B) = 1$ とするための正規化係数とみなすことができる。

【例 2.3】 三つの壺がある。これらを U_1, U_2, U_3 とする。 U_1 には黒玉 1 個と白玉 1 個、 U_2 には黒玉 3 個と白玉 2 個、 U_3 には黒玉 2 個と白玉 1 個がある。いま、運まかせに一つの壺を選び、その中から玉を 1 個取り出す。その玉が黒であったとき、それが U_1 から取り出された確率を求めよ。

解 壺 U_i から玉が取り出されるという事象を U_i 、取り出された玉が黒、白であるという事象をそれぞれ黒、白で表わすと、

$$P(U_1|\text{黒}) = \frac{P(\text{黒}|U_1)P(U_1)}{\sum_{i=1}^3 P(\text{黒}|U_i)P(U_i)} = \frac{\frac{1}{2} \times \frac{1}{3}}{\left(\frac{1}{2} + \frac{3}{5} + \frac{2}{3}\right) \times \frac{1}{3}} = \frac{15}{53}$$

2.3 完全事象系

いま標本空間 U が、図 2.2 のように、確率 P_k で生起する互いに排他的な n 個の事象 E_k に分割できるとする。このとき、このような事象の集合は**完全事象系**と呼ばれ、

$$\left| \frac{E}{P} \right| = \left| \begin{array}{cccc} E_1 & E_2 & \cdots & E_n \\ P_1 & P_2 & \cdots & P_n \end{array} \right| \quad (2.5)$$

のように表現される。ただし、 $\sum_{i=1}^n E_i = U$ (全事象)、 $\sum_{i=1}^n P_i = 1$ である。たとえば、サイコロを振ったとき i の目がでる事象を E_i で表わすと、

$$\left| \frac{E}{P} \right| = \left| \begin{array}{cccc} E_1 & E_2 & \cdots & E_6 \\ \frac{1}{6} & \frac{1}{6} & \cdots & \frac{1}{6} \end{array} \right|$$

のように表現される。

いま、ある事象の起こったことを観測した場合、何らかの情報を得たと考えることができる。確率 P で起こる事象を観測した場合、それによって得られる情報の量(これを自己情報量という)を

$$I = -a \log_k P \quad (2.6)$$

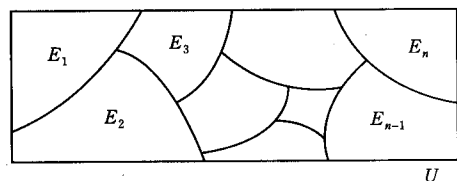


図 2.2 完全事象系

2.3 完全事象系

で定義するとつごうがよい。対数を用いることにより、二つの独立な事象を観測した場合、結合事象の生起確率はそれぞれの生起確率の積で表わされるから、自己情報量はそれぞれの自己情報量の和になる。さらに、定数は一般に $a=1, k=2$ にとられる。後で述べるように、定数をこのように決めることにより、デジタル情報や記憶の単位であるビット(bit)との整合性がとれるのである。このとき、自己情報量の単位も**ビット**となる。すなわち、

$$\begin{aligned} I &= -\log_2 P \quad [\text{ビット}] \\ &= \log_2 \frac{1}{P} \quad [\text{ビット}] \end{aligned} \quad (2.7)$$

したがって、非常にめずらしい事象を観測したとき、 P は非常に小さいから I は大きくなる。たとえば、1000 分の 1 の確率で当たるくじに当たったという情報は $I = \log_2 1000 \cong 10$ [ビット] の自己情報量をもっている。逆にそのくじにはずれたという情報はめずらしくない情報であり、自己情報量はほとんど 0 である。

さらに、完全事象系 E の事象が一つ起こったことを知ったときに平均的に得られる情報量を、それから得られる自己情報量の期待値で定義する。これを**平均情報量**、または単に**情報量**という。すなわち、得られる情報量 $\bar{I}$ は

$$\bar{I} = \sum_{i=1}^n (-P_i \log P_i) \quad [\text{ビット}] \quad (2.8)$$

で与えられる。

一方、完全事象系 E は最初 $H(E)$ なるあいまいさをもっていたとする。事象の観測後はこのあいまいさはなくなり、0 となる。その差が観測により得られる情報量であるとする。このあいまいさ $H(E)$ を完全事象系 E の**エントロピー**と呼ぶ。すなわち、エントロピーは情報量に等しく、

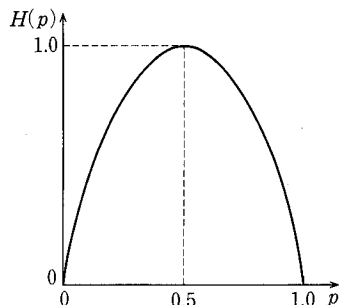
$$H(E) = \sum_{i=1}^n (-P_i \log P_i) \quad [\text{ビット}] \quad (2.9)$$

で定義される。特に断わらない限り、対数の底は 2 である。ときには自然対数が使われることもあるが、そのときにはエントロピーの単位は**ナット(nat)**となる。

【例 2.4】 いま、1, 0 の 2 種類の記号をそれぞれ確率 p および $(1-p)$ で発生する記憶のない情報源を考える。この情報源のエントロピーは

$$H(p) = -p \log p - (1-p) \log (1-p) \quad (2.10)$$

となる。これを p の関数として図示すると図 2.3 のようになる。この図より、一方

図 2.3 2元情報源のエントロピー $H(p)$

の生起確率が0または1なら、エントロピーは0であることがわかる。すなわち、その情報源からは常に0(または1)が発生し、あいまいさは0である。また、 $p=0.5$ 、すなわち0と1の両者が等確率で生じるときエントロピーは最大値1をとる。一般に、 n 個の事象があるとき、それらの生起確率がすべて等しい、すなわち $P_1=P_2=\dots=P_n=1/n$ のとき、エントロピーは最大値 $H=\log n$ をとる。エントロピー関数の値は巻末の付表1に示す。

上の例で述べたように、1, 0の2種類の記号のどちらかをとり二者択一の状態は最大1ビットの情報量をもつ。これはちょうど2元系列の1記号もしくは2進数の1桁や、それらのための記憶単位のビットに対応している。しかしながら、エントロピーは確率に関係した量であり、後者と区別するためエントロピーの単位をシャノンということもあるが、一般には普及していない。

なお、4.5, 4.8節では、エントロピーは情報源を一意的に復号可能なように符号化するときの2元符号の符号長の下限として自然な形で誘導される。このことから、エントロピーが情報量・あいまいさの測度として実質的な意味をもっていることがわかる。また詳しくは述べないが、このエントロピーは熱力学における分子などの無秩序さを表わすエントロピーと同じ概念である。

演習問題 2.

- 2.1 ある大学の自宅通学の学生の割合は2/3であり、その3/4がクラブ活動をしている。また、下宿している学生の割合は1/3であり、その1/2がクラブ活動をしている。クラブ活動をしている学生のうち、自宅通学の学生の割合はいくらか。
- 2.2 1/6の確率で起こる事象を観測したとき、平均的に得られる情報量はいくらか。ただし、 $\log_2 3=1.58$ 、 $\log_2 5=2.32$ である。

3

情報源

情報発生源には各種のものがあるが、本章ではそれらを取扱いやすい形にモデル化する。特に、出力がいつでも同じ統計的性質をもつ定常情報源や、一定以上過去には影響されないマルコフ情報源が重要であり、それらの性質について述べる。

3.1 情報源の統計的表現

情報源記号の集合 $A=\{a_1, a_2, \dots, a_M\}$ のうちからどれか一つずつ確率的に記号を順次発生・出力する離散的 M 元情報源を考える。時刻 i ($i=0, 1, 2, \dots$) において、情報源記号は A のうちのどれかをとり、それを確率変数 X_i で表わす。 A は情報源アルファベットと呼ばれる。時刻 $i=0, 1, 2, \dots, n-1$ において、出力がそれぞれ $x_0, x_1, x_2, \dots, x_{n-1}$ である結合確率を

$$P_{X_0, X_1, X_2, \dots, X_{n-1}}(x_0, x_1, x_2, \dots, x_{n-1}) \quad (3.1)$$

で表わす。確率変数名を省略し、

$$P(x_0, x_1, x_2, \dots, x_{n-1})$$

と略記することもある。

3.2 記憶のない定常情報源

各時刻における情報源記号の発生が他の時刻の発生記号とは独立であるとき、この情報源を記憶のない情報源という。定常ということは、出力記号の発生確率が、いつそれを観測するかによらないで同じであるということである。

演習問題 8.

- 8.1 いま、確率変数 X および Y は、それぞれ独立なガウス分布 $\mathcal{N}(0, \sigma_x^2)$ 、および $\mathcal{N}(0, \sigma_y^2)$ に従うものとする。このとき、それらの和である確率変数 $Z = X + Y$ はガウス分布 $\mathcal{N}(0, \sigma_x^2 + \sigma_y^2)$ に従うことを示せ。
- 8.2 (1) 単位帯域幅当たりの相加雑音電力 N_0 [ワット/Hz] をもつ帯域幅 W [Hz] の通信路を通して、信号電力 S [ワット] の信号を送る。このとき、通信路容量 C [ビット/秒] はいくらとなるか。
- (2) 通信路容量 C は W の単調増加関数である。すなわち、広い帯域を使用したほうが通信容量は大きくなる。 $W \rightarrow \infty$ の極限において C はいくらとなるか。
- [ヒント: $h \rightarrow \infty$ のとき、 $(1 + (1/h))^h \rightarrow e$]
- 8.3 連続的通信路において通信路容量 C 、情報伝送速度 R とするとき、 $C > R$ ならば誤り率を任意に小さくできる(そのような符号が存在する)。この符号の電力を S 、単位帯域幅当たりの雑音電力を N_0 、符号の語数を M 、その伝送時間を T とするとき、ビット当たりのエネルギー比には

$$\frac{ST_b}{N_0} > \log_e 2$$

の関係があることを示せ。ただし

$$T_b = \frac{T}{\log_2 M} = 1 \text{ ビット当たりの伝送時間}$$

とする。

$$\left[\begin{array}{l} \text{ヒント: } C = W \log_2 \left(1 + \frac{S}{N} \right), \quad N = WN_0, \quad W = \text{帯域幅} \\ \log_2 \left(1 + \frac{1}{n} \right)^n < \log_2 e \quad (n > 0) \end{array} \right]$$

- 8.4 変数 X の範囲が $-x_0 \leq X \leq x_0$ に制限されているとき、エントロピー $H(X)$ を最大にする X の確率密度関数 $p(x)$ を求めよ。

9

統計的パターン認識

帯域制限を受けた信号波形は、標準化や直交関数展開により、ベクトル表現できることは7章ですでに述べた。また、文字や画像の認識においては、画像から各種の特徴が計算される。したがって、信号や図形・画像の認識を含めて、一般にパターン認識問題においては特徴ベクトルをもとに、パターンの識別・クラス分けが行われる。これらのパターンはしばしば変形を受けたり、雑音で乱されたりする。パターンの変形も特徴ベクトル上では雑音加わったとみなせることが多い。そこで本章では、ガウス雑音加わった標本値が与えられたときに、統計的な見地からパターンの認識・分類の問題を考える。

9.1 多次元ガウス分布

正規分布は多くの連続的確率事象によくあてはまる分布である。複数の標本値が独立でそれらが同じガウス分布に従うときの確率密度関数については8.1節ですでに述べた。これらの場合も含め、一般の n 次元ガウス分布の確率密度関数は次式で与えられる。

$$p(x_1, x_2, \dots, x_n) = \frac{1}{\sqrt{|2\pi\Sigma|}} \exp \left\{ -\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right\} \quad (9.1)$$

ここに、

$$x = (x_1, x_2, \dots, x_n),$$

$$\mu = (\mu_1, \mu_2, \dots, \mu_n); \quad \text{平均値,}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix}; \quad \text{共分散行列}(n \times n \text{ 対称行列}),$$

$$\sigma_{ij} = \overline{(x_i - \mu_i)(x_j - \mu_j)}^\dagger,$$

T : 転置,

定常情報源では各標本値の確率密度関数は時刻によらず同一であるから,

$$\begin{aligned} \mu_i &= \mu \quad (i=1, 2, \dots, n) \\ \sigma_{i-l, j-l} &= \sigma_{ij} \quad (l=1, 2, \dots) \end{aligned} \quad (9.2)$$

となる。いちばん簡単な $n=2$ で、かつ平均値が $\mathbf{0}$ の場合についてこれを述べる。この2次元定常ガウス情報源では

$$\mu_1 = \mu_2 = 0$$

$$\sigma_{11} = \sigma_{22} = \sigma^2$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 \end{bmatrix}$$

である。このとき

$$\begin{aligned} |2\pi\Sigma| &= |2\pi I\Sigma| \\ &= |2\pi I| |\Sigma| \\ &= (2\pi)^n (\sigma^4 - \rho^2\sigma^4) \\ &= (2\pi)^2 \sigma^4 (1 - \rho^2), \end{aligned} \quad (9.3)$$

$$\Sigma^{-1} = \frac{1}{\sigma^4(1-\rho^2)} \begin{bmatrix} \sigma^2 & -\rho\sigma^2 \\ -\rho\sigma^2 & \sigma^2 \end{bmatrix}, \quad (9.4)$$

$$\begin{aligned} \mathbf{x}\Sigma^{-1}\mathbf{x}^T &= \frac{1}{\sigma^4(1-\rho^2)} [x_1 \ x_2] \begin{bmatrix} \sigma^2 & -\rho\sigma^2 \\ -\rho\sigma^2 & \sigma^2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \frac{1}{\sigma^2(1-\rho^2)} (x_1^2 - 2\rho x_1 x_2 + x_2^2). \end{aligned} \quad (9.5)$$

したがって,

$$p(x_1, x_2) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left[-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2\sigma^2(1-\rho^2)}\right]. \quad (9.6)$$

なお,

$$\rho = \frac{\sigma_{12}}{\sigma^2} = \frac{\overline{x_1 x_2}}{\overline{x_1^2}}; \quad \text{相関係数}$$

[†] 本章では表記の簡単化のため、確率変数と単なる変数は同じ記号を用いる。

一般的には

$$\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}} = \frac{\overline{(x_1 - \mu_1)(x_2 - \mu_2)}}{\sqrt{\overline{(x_1 - \mu_1)^2} \overline{(x_2 - \mu_2)^2}}}$$

である。

2次元定常ガウス分布を $\rho=0$ および 0.9 の場合につき確率密度関数の等高線で表示すると図9.1のようになる。ガウス分布の場合、 $\rho=0$ は x_1 と x_2 が独立であることを示している。また、 $\rho=0.9$ は x_1 と x_2 が高い相関をもっており、たとえば、 x_1 が大きな値のときには x_2 も大きな値をとることが多いことを示している。

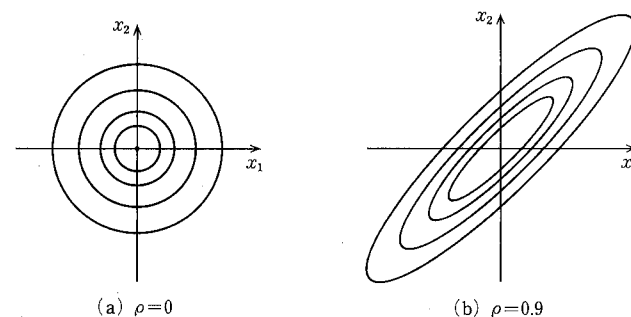


図 9.1 2次元ガウス分布の等高線

9.2 統計的決定法

いま、パターン認識の代表例として、図9.2のような通信系モデルを考える。受信信号は

$$\mathbf{Y} = \mathbf{S}_\theta + \mathbf{N} \quad (9.7)$$

で与えられる。受信機では $\mathbf{Y}$ を観測して $\mathbf{S}_0$ が送られたか、 $\mathbf{S}_1$ が送られたかの識別を行う (θ の推定値 $\hat{\theta}$ を得る)。なお $\mathbf{S}_0$ および $\mathbf{S}_1$ は既知であり、それらの生起確率はそれぞれ $1/2$ であることがわかっているとする。また通信路雑音 $\mathbf{N}$ は平均 $\mathbf{0}$ 、分散 Σ の既知のガウス分布 $\mathcal{N}(\mathbf{0}, \Sigma)$ に従う。これらのベクトルはいずれも n 次元行ベクトル、また行列は $n \times n$ 行列としておく。

$\mathbf{Y}$ を観測したとき、 θ に関する事後確率(条件付き確率) $P(\theta|\mathbf{Y})$ を最大にする決定を行うことは一つの妥当なやり方であろう。以下 $P(\cdot)$ は確率、 $p(\cdot)$ は確率密度関数を表わすものとする。いま、パターンには $\theta=0$ および $\theta=1$

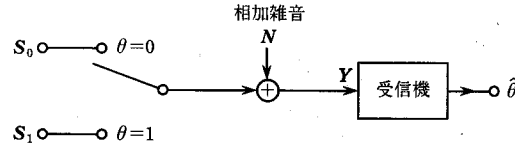


図 9.2 通信系モデル

の二つのクラスしかないので、この決定法に従えば、比

$$\begin{aligned}\lambda(Y) &= \frac{P(\theta=1|Y)}{P(\theta=0|Y)} \\ &= \frac{\frac{P(\theta=1)p(Y|\theta=1)}{p(Y)}}{\frac{P(\theta=0)p(Y|\theta=0)}{p(Y)}} \\ &= \frac{p(Y|\theta=1)}{p(Y|\theta=0)}\end{aligned}\quad (9.8)$$

を求め、

$$\begin{cases} \lambda \geq 1 \text{ ならば } \hat{\theta}=1, \\ \lambda < 1 \text{ ならば } \hat{\theta}=0 \end{cases}\quad (9.9)$$

と決定することになる。式(9.8)の

$$\lambda(Y) = \frac{p(Y|\theta=1)}{p(Y|\theta=0)} \quad (\text{尤度比}) \quad (9.8)'$$

は各クラスを仮定したときの Y の出現確率(密度)の比であり、尤度比と呼ばれる。したがって、式(9.9)に従って識別を行う決定方式は、尤度比を用いて決定を行うことから**尤度比検定**と呼ばれるものになっている。

ここで尤度比 $\lambda(Y)$ を求めてみよう。雑音 N の分布はガウス確率密度関数

$$p(N) = \frac{1}{\sqrt{2\pi\Sigma}} \exp\left\{-\frac{1}{2} N\Sigma^{-1}N^T\right\} \quad (9.10)$$

で与えられる。ここに $|\cdot|$ は行列式、 Σ^{-1} は Σ の逆行列を表わす。 $Y=S_\theta+N$ より、 Y の条件付き確率密度関数は

$$p(Y|\theta) = |2\pi\Sigma|^{-1/2} \exp\left\{-\frac{1}{2} (Y-S_\theta)\Sigma^{-1}(Y-S_\theta)^T\right\}$$

で与えられる。このとき、尤度比は

$$\begin{aligned}\lambda(Y) &= \frac{p(Y|\theta=1)}{p(Y|\theta=0)} \\ &= \exp\left\{-\frac{1}{2} [(Y-S_1)\Sigma^{-1}(Y-S_1)^T - (Y-S_0)\Sigma^{-1}(Y-S_0)^T]\right\}\end{aligned}\quad (9.11)$$

で計算できる。

ここで簡単な場合について式(9.8)の決定規則を考えてみよう。ON-OFF 通信系、すなわち $S_1=S$ および $S_0=0$ を仮定する。また、通信路雑音は無相関(uncorrelated)、すなわち、 $\Sigma=\sigma^2 I$ とする。このとき、式(9.11)は

$$\lambda(Y) = \exp\left[-\frac{1}{2\sigma^2}\{(Y-S)(Y-S)^T - YY^T\}\right] \quad (9.12)$$

となる。したがって、式(9.8)において λ を 1 と比較することは、上の $\exp$ の中身が正か負かを調べることである。すなわち、 $(Y-S)(Y-S)^T$ は Y と S の間の距離の二乗であり、 YY^T は Y と 0 の間の距離の二乗であるから、結局いま観測された Y が S と 0 のどちらにより近いかに観測値の識別を行っていることがわかる。この決定規則を改めて書くと次のようになる。

$$\begin{cases} YY^T \geq (Y-S)(Y-S)^T \text{ ならば } \hat{\theta}=1, \\ YY^T < (Y-S)(Y-S)^T \text{ ならば } \hat{\theta}=0, \end{cases}\quad (9.13)$$

また、別の見方も可能である。式(9.12)の $\exp$ の中は

$$\begin{aligned}-\frac{1}{2}[(Y-S)(Y-S)^T - YY^T] &= -\frac{1}{2}(YY^T - YS^T - SY^T + SS^T - YY^T) \\ &= YS^T - \frac{1}{2} SS^T\end{aligned}\quad (9.14)$$

である。ここで、内積 YS^T および SY^T はスカラー量であり、これらは

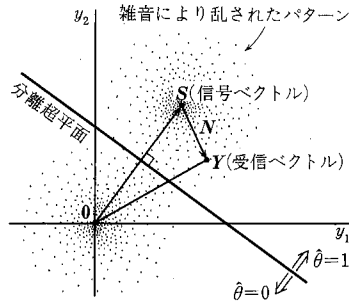
$$YS^T = SY^T = \sum_{i=1}^n y_i s_i \quad (9.15)$$

となって等しいことを用いた。したがって、式(9.14)の正負を調べて識別を行う上の決定規則は

$$\begin{cases} YS^T \geq \frac{1}{2} SS^T \text{ ならば } \hat{\theta}=1, \\ YS^T < \frac{1}{2} SS^T \text{ ならば } \hat{\theta}=0 \end{cases}\quad (9.16)$$

となるので、 Y と S の内積により識別を行っていることがわかる。このような方式は、信号の形をしたフィルタを用いていることから、**整合フィルタ**(matched filter)または**整合機械**(template machine)と呼ばれる。雑音が無相関でない場合には、式(9.16)の左辺は $Y\Sigma^{-1}S^T$ のように雑音の分散で正規化(白色化)した形となる(演習問題9.2)。このような一般形をした方式も整合機械と呼ばれる。

これらを各ベクトルが2次元の場合について図示すれば図9.3のようになる。すなわち、 S と 0 を結ぶ直線を垂直二等分する分離超平面で空間を二等

図 9.3 分離超平面 ($n=2$)

分し、この分離超平面より S に近い側に Y がくれば $\hat{\theta}=1$ と決定し、 0 に近い側に Y がくれば $\hat{\theta}=0$ と決定する。図 9.3 は $n=2$ の場合で直線により受信ベクトル平面が分割されているが、さらにもっと多次元の場合も含めて、この直線を超平面と呼んでいる。なお、一般に Y の線形関数を用いて信号クラスの識別を行なう方式を線形識別方式という。整合機械はこの線形識別方式の代表的なものである。

ここで、線形識別方式の誤り率 P_e を求めてみよう。いま線形識別規則を

$$\begin{cases} YA^T \geq \delta \text{ ならば } \hat{\theta}=1 \\ YA^T < \delta \text{ ならば } \hat{\theta}=0 \end{cases} \quad (9.17)$$

とする。式(9.17)の左辺を z とおく。すなわち

$$z = YA^T \quad (9.18)$$

とおくと、 z はガウス分布をする Y の各要素の荷重和である。したがって、 z もガウス分布に従う。その平均および分散はそれぞれつぎようになる：

$$\begin{aligned} \bar{z} &= \overline{YA^T} \\ &= \bar{Y}A^T \\ &= \theta SA^T, \\ \overline{(z - \bar{z})^2} &= \overline{(NA^T)^2} \\ &= \overline{(AN^T)(AN^T)} \\ &= \overline{AN^TNA^T} \\ &= \sigma^2 AA^T \\ &\triangleq \sigma^2. \end{aligned} \quad (9.19) \quad (9.20)$$

以上より、線形識別関数 $z = YA^T$ は平均 θSA^T 、分散 σ^2 の1次元ガウス分布に従うことがわかる。この様子を図 9.4 に示す。いま、いき値は δ である

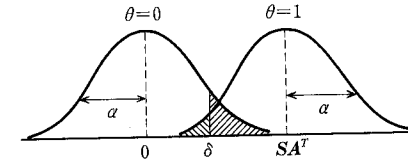


図 9.4 線形識別関数の分布

から図の斜線の部分が誤りとなる。標準ガウス分布関数を

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt \quad (9.21)$$

とすると、これらの誤り率はそれぞれ

$$\begin{aligned} P(\hat{\theta}=1 | \theta=0) &= 1 - \Phi\left(\frac{\delta}{\sigma}\right) \\ P(\hat{\theta}=0 | \theta=1) &= \Phi\left(\frac{\delta - SA^T}{\sigma}\right) \end{aligned} \quad (9.22)$$

となる。また、平均誤り率は

$$P_e = P(\theta=0)P(\hat{\theta}=1 | \theta=0) + P(\theta=1)P(\hat{\theta}=0 | \theta=1) \quad (9.23)$$

で与えられる。

いま、線形識別の重みが式(9.16)で与えられる整合フィルタの場合の誤り率をもう少し具体的に求めてみる。このとき、

$$A = S, \quad \delta = \frac{1}{2} SS^T \quad (9.24)$$

であるから、式(9.22)において

$$\begin{aligned} \frac{\delta}{\sigma} &= \frac{\frac{1}{2} SS^T}{\sqrt{\sigma^2 SS^T}} \\ &= \frac{1}{2} \sqrt{R} \end{aligned} \quad (9.25)$$

となる。また

$$\begin{aligned} \frac{\delta - SA^T}{\sigma} &= \frac{\frac{1}{2} SS^T - SS^T}{\sqrt{\sigma^2 SS^T}} \\ &= -\frac{1}{2} \sqrt{R}. \end{aligned} \quad (9.26)$$

ここに

$$R = \frac{SS^T}{\sigma^2}; \quad \text{S/N 比}$$

これらを用いて式(9.22)は

$$\begin{aligned} P(\hat{\theta}=1|\theta=0) &= 1 - \Phi\left(\frac{1}{2}\sqrt{R}\right) \\ &= \Phi\left(-\frac{1}{2}\sqrt{R}\right), \end{aligned} \quad (9.27)$$

$$P(\hat{\theta}=0|\theta=1) = \Phi\left(-\frac{1}{2}\sqrt{R}\right) \quad (9.28)$$

となる。これら両者は等しいから平均誤り率も

$$P_e = \Phi\left(-\frac{1}{2}\sqrt{R}\right) \quad (9.29)$$

となり、これらは信号のエネルギーと雑音電力の比 R のみで決まることがわかる。なお、標準ガウス分布関数 $\Phi(\cdot)$ の値は付表3に示されている。

これらの誤り率を図示すると図9.5のようになる。

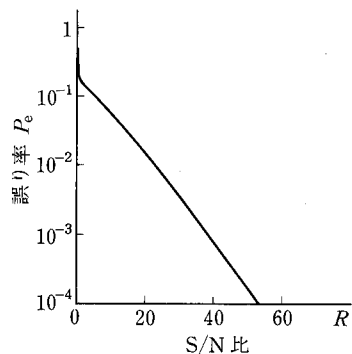


図 9.5 S/N 比と整合機械の誤り率

演習問題 9.

9.1 x_1, x_2 の結合確率密度関数を

$$p(x_1, x_2) = \frac{1}{2\pi\sigma^2\sqrt{1-\rho^2}} \exp\left\{-\frac{x_1^2 - 2\rho x_1 x_2 + x_2^2}{2\sigma^2(1-\rho^2)}\right\}$$

とする。このとき、 $x_1 x_2$ および x_1^2 の期待値をそれぞれ求めよ。

9.2 (1) 雑音の共分散行列が一般の Σ のとき、式(9.16)の決定規則はどのようなになるか。

(2) また、

$$S = [1 \ 2], \quad \Sigma = \begin{bmatrix} \frac{1}{3} & 0 \\ 0 & 1 \end{bmatrix}$$

のとき、分離超平面はどのようなになるか、図示せよ。

(3) このときの誤り率を求めよ。

10

命題論理

人間の知的活動の主要なものは、既知の事実から必要な推論を行い、与えられた問題の解答を得たり、判断を行うことである。

計算機にこのようなことを行わせるためには、推論過程の定式化と理論的な考察が必要となる。本章ではその基本となる命題論理について述べる。

命題論理では真偽を与えることのできる命題(基本論理式)をもとに、論理記号を用いてそれらの複合命題(論理式)をつくることができる。これらの命題相互間の真偽の関係を論じるのが命題論理である。これらの命題はその構成法から人工的な言語とみなすことができる。

10.1 論理式

命題論理が対象とする命題は、たとえば次のような文である。

P (今日は風が強い)

Q (今日は雨だ)

R (1たす1は3である)

これらの命題は論理式の基本となるので基本論理式または基本命題と呼ぶ。基本論理式はこれ以上分解することのできない論理式である。基本論理式は現在対象とする世界において真(T)または偽(F)のいずれかの値をとる。それを真理値と呼ぶ。

表 10.1 に示す論理記号を用いてこれらの基本論理式より論理式をつくることができる。なお、次章で述べる述語論理と区別する必要がある場合にはこれ

(1)'~(3)' から(4)'を証明するため、(1)'~(3)'の節形に(4)'を否定した節を加えた節集合
 $\{\sim\text{human}(x) \vee \text{mortal}(x), \sim\text{god}(y) \vee \sim\text{mortal}(y), \text{human}(\text{socrates}), \text{god}(z)\}$
 をつくる。これから空節が導出される導出木の例を念のため図 11.4 に二つ示すが、いずれも解 $z=\text{socrates}$ を与えている。

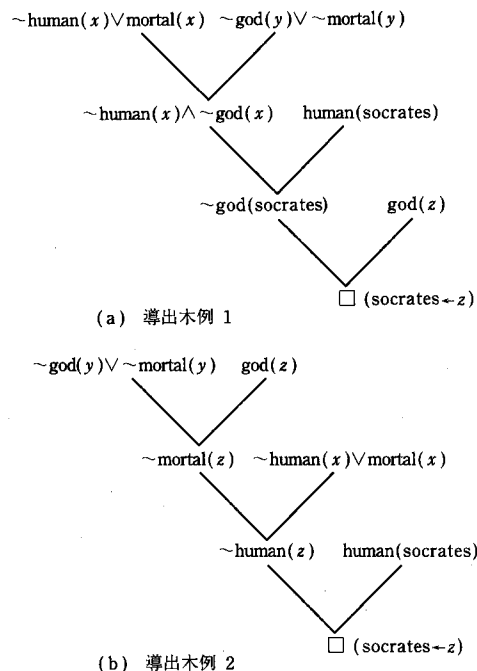


図 11.4 導出原理による質問応答。(a), (b)は導出木は異なるが、いずれにしろ同じ結果を与える。

演習問題 11.

11.1 (1) $\{Q(x, f(x, y)), Q(u, v)\}$

および,

(2) $\{Q(x, f(y, z)), Q(u, v), Q(a, f(h(b), z))\}$

をそれぞれ最も一般的な単一化置換により単一化せよ。

11.2 つぎの節集合(1), (2)の導出木を書け。

(1) $\{P(x) \vee P(f(y)) \vee Q(x), \sim P(x) \vee R(f(x)), \sim R(f(z)) \vee \sim Q(z)\}$

(2) $\{P(x), \sim P(x) \vee Q(x), \sim Q(x) \vee R(f(x)), \sim Q(f(g(y))) \vee S(y), \sim R(f(a)) \vee \sim S(z)\}$

12

ニューラルネットワーク計算

ニューラルネットワークは多数の神経素子と、それらの間を結ぶ重み付き配線からなっている。これらの重みは可変であり、学習によって処理対象に適した重みを得ることも可能である。すなわち、人間のように学習が可能であることが、ニューラルネットワークの大きな特徴である。また、いろいろな処理を同時に行う並列処理が可能であることも大きな特徴である。

ネットワーク表現は前章のような記号表現に比し、より①低レベル、②均一的である。したがって、ニューラルネットワークはパターン認識、画像処理など、あいまいな情報の処理に向いているといえよう。

本章では、ニューラルネットワークの基本原理について述べる。以下、12.1節ではニューラルネットワークの形態について、また12.2節および12.3節ではその代表例である階層型、および相互結合型ネットワークについてそれぞれ述べる。

12.1 ニューラルネットワークの形態

ニューラルネットワークは図 12.1(a)に示すような神経素子を多数接続したものである。個々の素子(神経素子)は入力加重和を求め、それをシグモイド関数(S字型単調増加関数; sigmoid function)に通したものを出力する。シグモイド関数の代表例がロジスティック(logistic)関数

$$f(x) = \frac{1}{1 + e^{-x}} \quad (12.1)$$

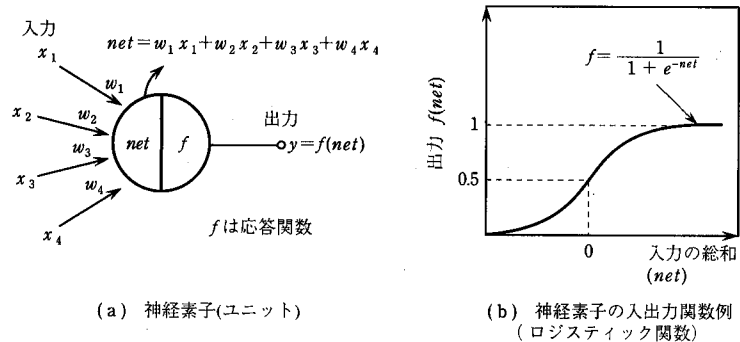


図 12.1 ニューラルネットワークで使われる素子と入出力関係。

である(図 12.1(b))。実際の神経細胞はパルスの頻度で信号を伝えるが、人工のニューラルネットワークでは、パルス頻度をアナログ値(神経の活性値)で置き換えて表現する。ニューラルネットワークの代表的な形態には図 12.2 に示すような階層型ネットワークと相互結合型ネットワークがある。

12.2 階層型ネットワークと誤差逆伝搬学習アルゴリズム

階層型ネットワークにおいては、各神経素子は入力層、中間層、出力層に分かれている。入力層から入ったパターンは、ネットワークで変換を受け、出力層から出力される。同一層内での結合はないとする。階層型ネットワークは基本的にはパターン変換を行うネットワークであり、その特別な場合がパターン認識・識別・分類である。

階層型ネットワークの学習は誤差逆伝搬学習アルゴリズムにより行われることが多い。この学習アルゴリズムにおいては、教師の教えと出力の誤差の責任量が、中間層を通して各結線の重みへフィードバックされる。これにより、3層、4層など多層のネットワークを教育することができる。この学習機能を利用して、手書き文字認識、発音記号の学習、音声認識など多数の応用システムが開発されている。

誤差逆伝搬学習アルゴリズム 誤差逆伝搬学習(error back-propagation learning)アルゴリズムは、教師付き学習アルゴリズムであり、階層型ネット

ワーク(別名 BPN; Back-Propagation Network)における出力素子の出力値ができるだけ教師の教えに近づくよう、重みの調整(学習)が行われる。

図 12.2(a)に示す階層型モデルについて説明する。最上層は出力層であり、出力層を構成する各素子 l の出力 o_l のとるべき値 t_l が教師から与えられる。入力を入力層の各素子に与えられる。中間層も含め、各素子は一つ下の層の素子からの入力の重み付き線形和(net)に対応する出力 $f(net)$ を出す。これらを式で表現すると次のようになる。

出力層の素子 l の出力誤差を次式で定義する：

$$E = \frac{1}{2} (t_l - o_l)^2 \quad (12.2)$$

素子 l の出力 o_l は、

$$o_l = f_l(net_l) \quad (12.3)$$

で与えられる。ここに、

$$net_l = \sum_k w_{kl} o_k \quad (12.4)$$

この o_k は一つ下の層の素子 k の出力である。なお、添字 l, k などは対応する層(l 層および k 層など)内の素子についてとるものとする。以下、同じである。また同様に、

$$o_k = f_k(net_k) \quad (12.5)$$

$$net_k = \sum_j w_{jk} o_j \quad (12.6)$$

⋮

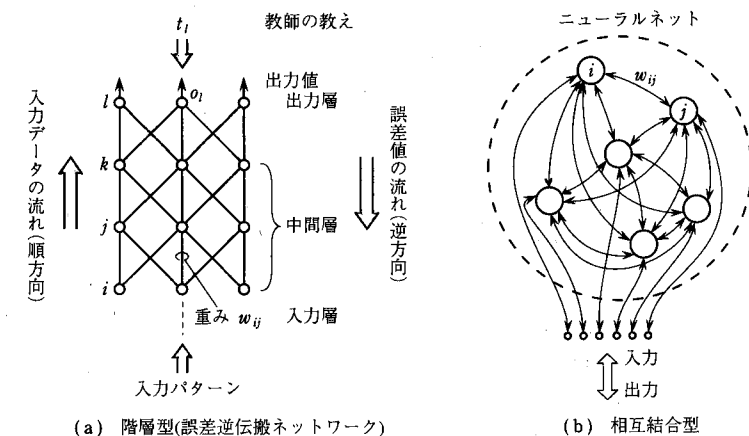


図 12.2 ニューラルネットワークの基本形態

ここで、出力誤差 E を素子 i から j への重み w_{ij} で偏微分する。これは w_{ij} に関する E の感度を表わしており、これが大きいことは、 w_{ij} を変化させることによる E の減少が大きいことを示している。したがって、重み w_{ij} の調整量は $-\partial E / \partial w_{ij}$ に比例した大きさとすればよい。すなわち、

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}} \quad (12.7)$$

$$\begin{aligned} &= \frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}} \\ &= \left[\sum_k \frac{\partial E}{\partial net_k} \frac{\partial net_k}{\partial o_j} \right] \frac{\partial o_j}{\partial net_j} \frac{\partial net_j}{\partial w_{ij}}. \end{aligned} \quad (12.8)$$

いま、

$$\delta_k \triangleq -\frac{\partial E}{\partial net_k}$$

とおく。

$$\frac{\partial net_k}{\partial o_j} = w_{jk}, \quad \frac{\partial net_j}{\partial w_{ij}} = o_i, \quad \frac{\partial o_j}{\partial net_j} = f'_j(net_j) \quad (12.9)$$

であるから、中間層では、

$$\begin{aligned} -\frac{\partial E}{\partial w_{ij}} &= \delta_j o_i \\ &= \left[\sum_k \delta_k w_{jk} \right] f'_j(net_j) o_i \end{aligned} \quad (12.10)$$

なる関係が得られる。ここに、 δ_j は誤差伝搬量(誤差に対する責任の大きさ)であり、次の逆方向伝搬関係式で出力層側から順に求められる。

$$\delta_j = \left[\sum_k \delta_k w_{jk} \right] f'_j(net_j) \quad (12.11)$$

なお、出力層においては δ_i は、

$$\begin{aligned} \delta_i &= -\frac{\partial E}{\partial net_i} \\ &= -\frac{\partial E}{\partial o_i} \frac{\partial o_i}{\partial net_i} = (t_i - o_i) f'_i(net_i) \end{aligned} \quad (12.12)$$

で与えられる。

出力における δ_i は式(12.12)から出発して、順次逆向きに式(12.11)を適用していくと、各素子における誤差伝搬量 δ_j が求められる。これらをもとに、式(12.10)から結合重み w_{ij} の E に関する責任量である $\partial E / \partial w_{ij}$ が求められる。これから、たとえば時刻 n における w_{ij} の増加量を、適当な正係数 η を用いて、

$$\Delta w_{ij}^n = -\eta \frac{\partial E^n}{\partial w_{ij}} \quad (12.13)$$

とする。ここに E^n は時刻 n における出力誤差量である。この方法は、 E のいちばん小さくなる方向に向ってパラメータを調整することから最急降下法(steepest descent method)と呼ばれる。あるいは、雑音に惑わされずに安定に降下するため過去の重み変化に対する慣性項 α ($0 \leq \alpha < 1$) を導入して、

$$\Delta w_{ij}^n = -\eta \frac{\partial E^n}{\partial w_{ij}} + \alpha \Delta w_{ij}^{n-1} \quad (12.14)$$

とする(モーメント法)。

【例 12.1】 図 12.3 のようなネットワークがある。○印は単純ないき値素子であり、そのいき値が○内に記されている。入力 x_1, x_2 がいき値以下のとき、出力は 0、それ以外のときには出力は 1 となる。線上的数値は結合の重みである。 x_1, x_2 は入力、 z_1, z_2 は中間層の出力、 y はネットワークの出力である。

(1) x_1, x_2 が $\{0, 1\}$ の値のみをとるとき、真理値表は表 12.1 のようになる。このとき、出力 y は x_1, x_2 の排他的論理和(XOR)になっている。XOR のような非線形識別は単独の神経素子では困難であっても、多層にすると可能となる。なお、問題の簡単化のためにいき値素子を使用したがる、いき値素子の入出力特性は微分可能ではない。そのため誤差逆伝搬学習は不可能である。しかしながら、パターン変換・識別は可能である。

(2) x_1, x_2 が連続値をとるとき、出力 y は図 12.4 のようになる。ここでは線形識別関数を二つ並べた非線形識別関数でパターンの識別が行われていることがわかる。

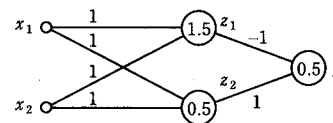


図 12.3 例 12.1 のネットワーク

表 12.1 図 12.3 のネットワークの真理値表

$x_1 \backslash x_2$	0	1
0	$z_1 = 0$ $z_2 = 0$ $\therefore y = 0$	$z_1 = 0$ $z_2 = 1$ $\therefore y = 1$
1	$z_1 = 0$ $z_2 = 1$ $\therefore y = 1$	$z_1 = 1$ $z_2 = 1$ $\therefore y = 0$

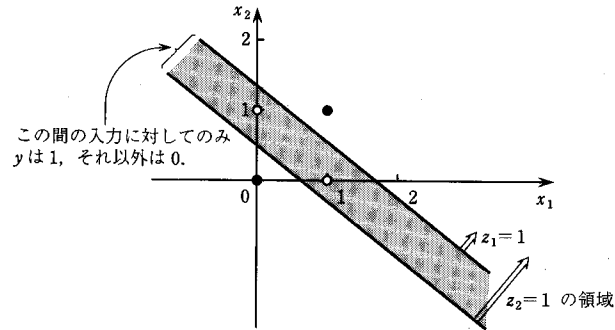


図 12.4 連続値入力に対する図 12.3 のネットワークの出力。

12.3 相互結合型ネットワーク

相互結合型ネットワーク (Hopfield ネットワーク) は、素子間に入出力の区別がなく、それらの間に相互の結合を許すものである。基本的な動作として、複数のパターンを結合の重みとしてネットワークに格納したのち、あるパターンの一部を入力すると、それに最も近いパターンがしばらくするとネットワーク上に素子の活性状態として想起・連想される。たとえば、多数の顔画像を格納し、ある人の鼻の部分だけ入力すると、その人の顔全体が出力される。また、人間の意識・無意識活動のように、多くのあいまいな要因が絡んだ相互作用のある系における問題解決やシミュレーションには相互結合型ネットワークが適している。

相互結合型ネットワークによる処理において使われる評価基準はネットワーク全体のエネルギー最小化である。問題に応じて適切な重みを選択する必要があるが、これにより、巡回セールスマン問題のような組合せ問題も解ける。個々の素子の活性状態を局所的状態に応じて変化させると全体のエネルギーが最小値に向い、解が得られる。好ましくない極小値に引っかかるのを避けるため、金属の焼きなましとのアナロジーから、確率的に素子の活性度を変化させる模擬焼きなまし (simulated annealing) の手法なども提案されているが本書ではふれない。

12.3.1 エネルギー最小化と連想

図 12.2(b) のようなニューラルネットワークにおいて、 w_{ij} を素子 i と素子 j の間の重みであるとする。素子 i の活性状態を u_i ($=0$ または 1) で表わす。すなわち、 $u_i=0$ のときには素子 i は非活性状態であることを表わし、 $u_i=1$ のときには活性状態であることを表わす。さらに、重み w_{ij} は i, j に関して対称、すなわち

$$w_{ij} = w_{ji} \quad (12.15)$$

であるとする。また、

$$w_{ii} = 0 \quad (12.16)$$

すなわち、自分自身への結合はないとする。

ここで、ネットワークのエネルギーを

$$E = -\frac{1}{2} \sum_i \sum_j w_{ij} u_i u_j - \sum_i (s_i - \theta_i) u_i \quad (12.17)$$

で定義する。ここで、 s_i および θ_i は、それぞれ素子 i の外部入力およびいき値である。このとき簡単にいえば、大きな重み w_{ij} で結ばれた素子は同時に興奮し ($u_i=1$ かつ $u_j=1$)、小さな重み w_{ij} で結ばれた素子どうしは同時に興奮することがない ($u_i=0$ または $u_j=0$) とき、第 1 項は小さく (エネルギー状態が低く) なる。また第 2 項は、外部入力 s_i が大きな素子はできるだけ活性化されたほうがネットワーク全体のエネルギーは小さくなることを表わしている。

いま、素子 i の値 (活性値) が 0 から 1 に変わったとする。これによる E の変化は

$$\Delta_i E = - \sum_j w_{ij} u_j - (s_i - \theta_i) \quad (12.18)$$

である。また、素子 i の値が 1 から 0 に変わった場合には、これによる E の変化は

$$\Delta_i E = \sum_j w_{ij} u_j + (s_i - \theta_i) \quad (12.19)$$

である。素子 i が以下の規則にしたがって状態変化をするとする。

$$\begin{cases} \sum_j w_{ij} u_j + s_i \geq \theta_i \text{ ならば, } u_i = 1 \\ \sum_j w_{ij} u_j + s_i < \theta_i \text{ ならば, } u_i = 0 \end{cases} \quad (12.20)$$

すなわち、他の素子からの入力と外部入力 s_i の総和がいき値より大きいとき、素子は 1 となり、小さいとき 0 となる。このとき、 $\Delta_i E$ は負または零であり、

正になることはない。したがって、上のような状態変化規則に従うとき、系全体のエネルギーは順次減少していくことになる。ただし、ネットワークのエネルギーはどこまでも小さくなれるわけではなく、最小値が存在する。なぜなら各 u_i は有界で 0 と 1 の間の値(実際上は 0 と 1 のみ考えておけば十分)しかとれないからである。最小値を与える u_i の組 $\{u_i\}$ は 1 個とは限らない。また、最小値以外に極小値があるとそこに引っかかって抜け出れないことがある。

このように、ある初期状態から出発したネットワークは状態変化規則に従って状態変化を繰り返すうちに、最小値または極小値に落ち込むことになる。そこで、エネルギーが最小または極小になるこれらの状態をある意味での望ましい状態としておくと、状態遷移によりそれらの状態が誘導されて出てくことになる。すなわち、望ましい状態にある程度近いパターンを入力すると、望ましい状態が連想出力されるのである。最小値を含め、エネルギーの極小値は一般に複数個存在するから、連想される状態もそれに対応して一般には複数個存在する。

図 12.5、図 12.6 には三つのパターン C, F, 4 を覚え込ませたネットワークに、ある初期状態を入力するとそれに近いパターンが連想出力されてくる様子を示す。連想したいパターンをネットワークエネルギー最小または極小状態に対応させる方法については次の項に示す。

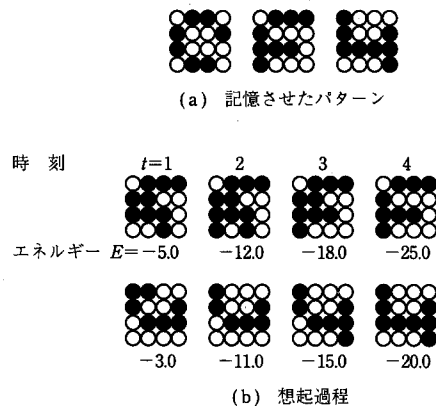


図 12.5 自己想起型文字連想過程におけるネットワークの状態変化とエネルギーの減少

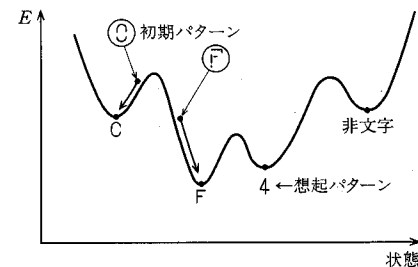


図 12.6 連想過程とネットワークのエネルギー変化

[例 12.2] (連想ネットワーク)

図 12.7 のようなネットワークがある。各素子の入出力関数 f は図 12.8 に示すように、区間 $[0.0, 1.0]$ 内の入力はそのまま出力し、負入力に対しては最小値 0.0、また 1.0 を越える入力に対しては最大値 1.0 をそれぞれ出力する。

(1) 最初(時刻 $t=1$)に a(肉食)だけを 0.1 の活性値で刺激したとき、ネットワークの活性状態はどのように変化していくであろうか。刺激は $t=2$ 以降では取り去るものとする。

図 12.7 より、ネットワークの重み行列は

$$W = \begin{bmatrix} a & b & c & d & e \\ 0 & -1 & 1 & 1 & -1 \\ -1 & 0 & 1 & -1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & -1 & 1 & 0 & -1 \\ -1 & 1 & 1 & -1 & 0 \end{bmatrix} \begin{matrix} a \\ b \\ c \\ d \\ e \end{matrix}$$

である。

時刻 $t=1$ における初期活性値ベクトルは

$$u_1 = (0.1, 0, 0, 0, 0)$$

である。時刻 $t=2$, およびそれ以降の活性値ベクトルは

$$u_2 = f(u_1 W)$$

$$= (0, 0, 0.1, 0.1, 0)$$

$$u_3 = f(u_2 W)$$

$$= (0.2, 0, 0.1, 0.1, 0)$$

$$u_4 = (0.2, 0, 0.3, 0.3, 0)$$

$$u_5 = (0.6, 0, 0.5, 0.5, 0)$$

$$u_6 = (1.0, 0, 1.0, 1.0, 0)$$

$$u_7 = (1.0, 0, 1.0, 1.0, 0)$$

⋮

となる。したがって、a(肉食), c(動物), d(ライオン)の三つが最終的に活性化されることがわかる。これは a(肉食)からこれらが連想されたことを表わしている。

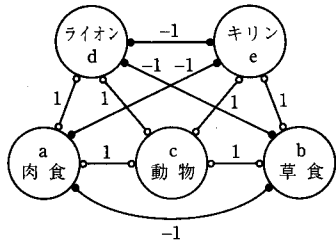
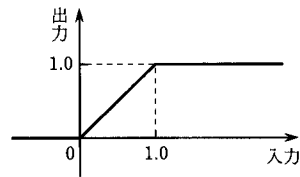


図 12.7 動物連想ネットワーク

図 12.8 入出力関数 f

(2) 時刻 t におけるネットワークのエネルギーを

$$\begin{aligned} E(t) &= -\frac{1}{2} \sum_{i,j} w_{ij} u_i u_j \\ &= -\frac{1}{2} \mathbf{u}_t^T \mathbf{W} \mathbf{u}_t \end{aligned} \quad (12.21)$$

で定義する。 u_i は素子 i の活性値, w_{ij} は素子 i, j 間の重みである。このとき, 各時刻におけるエネルギーは

$$\begin{aligned} E(1) &= -\frac{1}{2} \mathbf{u}_1^T \mathbf{W} \mathbf{u}_1 = 0, \\ E(2) &= -\frac{1}{2} \mathbf{u}_2^T \mathbf{W} \mathbf{u}_2 = -0.01, \\ E(3) &= -0.05, \\ E(4) &= -0.21, \\ E(5) &= -0.85, \\ E(6) &= E(7) = \dots = -3.0. \end{aligned}$$

時刻 $t=6$ 以降では, $E(t)$ は最小値に落ち込んでいる。

この例で

12.3.2 パターンの記憶

ここでは K 個の n 次元パターンベクトル $\mathbf{u}^s = (u_1^s, u_2^s, \dots, u_n^s)$ ($s=1, 2, \dots, K$) をネットワークに記憶させることを考える。そこで, 重み w_{ij} を

$$\begin{aligned} w_{ij} &= \sum_s (2u_i^s - 1)(2u_j^s - 1) \\ w_{ii} &= 0 \end{aligned} \quad (12.22)$$

としよう。ここで, $s_i=0, \theta_i=0$ としておく。上の w_{ij} は, 同時に興奮 ($u_i^s = u_j^s = 1$) する, または同時に抑制される ($u_i^s = u_j^s = 0$) ような素子間の重みは大きくなることを示している。

いまネットワークの状態が記憶された状態 $s=L$ にあるとする。このとき, 素子 i への入力 I_i の総和は

$$I_i = \sum_j w_{ij} u_j^L = \sum_s (2u_i^s - 1) \left[\sum_j u_j^L (2u_j^s - 1) \right] \quad (12.23)$$

となる。 $[\]$ 内の値は s により変化する。 u_j^L は 0, 1 をとり, $(2u_j^s - 1)$ は ± 1 をとる。パターン L と s が独立であれば $[\]$ 内の値はほぼ 0 となる。実際, 記憶パターンにそのような条件をつけることは厳しいことではない。したがって, $[\]$ は $s=L$ の場合のみ正の値 C をもち, それ以外の場合には 0 とすると,

$$I_i = (2u_i^L - 1) C \quad (12.24)$$

となる。いま仮定により, ネットワークは状態 L にあるから, 素子 i の値は $u_i = u_i^L$ である。つぎの時刻における u_i は

$$\begin{cases} u_i^L = 1 \text{ ならば, } I_i > 0. \text{ したがって, 次の時刻における } u_i \text{ も } 1, \\ u_i^L = 0 \text{ ならば, } I_i < 0. \text{ したがって, 次の時刻における } u_i \text{ も } 0 \end{cases}$$

となる。すなわち, ネットワークの状態は変化しない。これはネットワークが平衡状態にあることを示している。したがって, 状態が記憶パターン $\mathbf{u}^L$ の近くにあれば, 引き込まれて状態は $\mathbf{u}^L$ となり, そこで落ちつく。すなわち, 初期状態に近い記憶パターンが連想されることになる。

【例 12.3】巡回セールスマン問題

巡回セールスマン問題とは, いくつかの都市を一度ずつ通り, もとに戻ってくる問題で, そのときたとえば費用や所要時間などのコストを最小にするにはどういう順番で回ればよいかを求める問題である。これは人にとっては近似解がすぐにみえてくるが, 従来法で計算機にやらせようとする, 都市の数 n とともに急激に計算時間が増大する (n の多項式オーダーの時間では解けない) やっかいな問題として知られている。

これに対して, 図 12.9(a) に示すように, 都市名と訪問の順番の組を一つの素子に割り当てる。このとき, ネットワークのエネルギー E が問題の拘束と巡回のコストの加重和になるようにする。すなわち, (1) 同一都市は一度しか訪れないから, 同一都市の異なる訪問順の素子間には抑制性重みをつける, また, (2) 同じ訪問順番のときには一つの都市しか訪問できないから, 同一訪問順番の異なる都市間には抑制性重みをつける, (3) それ以外の素子間には, 都市間移動のコストに応じた重みをつける。

これらの各拘束 (1) ~ (3) はいくつの素子間結合を与えるか求めてみよう。

(1) 神経素子配列 (図 12.9(a)) の各行においてはすべての組合せ nC_2 個の結合に相互抑制性重みがつけられる。行の数は n である。したがって, 総計

$$nC_2 \cdot n = \frac{1}{2} n^2 (n-1)$$

個の重みを与える。

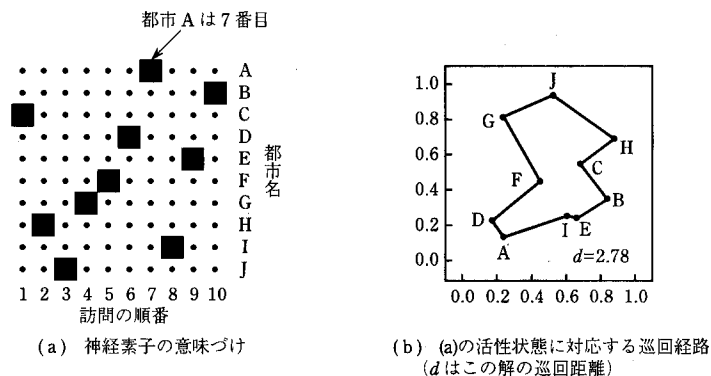


図 12.9 巡回セールスマン問題

(2) 同様にして

$${}_nC_2 \cdot n = \frac{1}{2} n^2 (n-1).$$

(3) 図 12.9(a)の各神経素子からはすぐ右隣の列の異なる都市の神経素子へ抑制性重みがつけられる。右端の列から左端の列へ同様に付けられる。すなわち、各神経素子からは $(n-1)$ 個の相互抑制性結合が出ていく。神経素子は n^2 個あるから、総計

$$n^2(n-1)$$

個の結合を与える。

なお、上の(1)~(3)の結合の場所は重ならないから、(1)~(3)合せて $2n^2(n-1)$ 個の結合がつくられる。これらの結合の重みは通常抑制性のみであり、このままではすべての神経素子の活性値が0になったときネットワークのエネルギーが最小になってしまう。そこで実際には、たとえば n^2 個の神経素子の活性値の総和と n との差の2乗に比例した量をエネルギーにつけて加えておく。

適当な初期状態から出発し、各素子は局所的な状況に従って状態を変化させていくと、ネットワークのエネルギーは順次低下していき、問題の拘束と巡回のコストを両方バランス良く満足させる最小値へ到達することが期待できる。そのときの各素子の活性値が解を与える(図 12.9(b))。しかしながら、しばしば最小値に到達する前に極小値に引っかかる。このときには、最良解ではなく近似解となる。

演習問題 12.

12.1 最急降下法

図 12.10 の 2 層ネットワークにおいて、出力素子の入出力特性を $f(net) = (net)^k$ とする。ここに $net = xW^T$, $x = (x_1, x_2, x_3)$ はネットワークの入力, $W = (w_1, w_2, w_3)$ は重みベクトルである。なお, $net > 0$ の範囲のみ考える。ネットワークの出力と教師の教え t の 2 乗誤差 E が小さくなる方向に修正する最急降下法に従って、重み W を修正せよ。

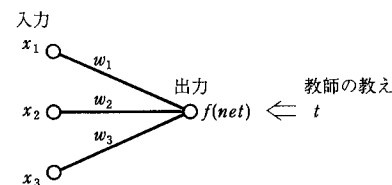


図 12.10 2層ネットワーク

12.2 例 12.2 において、c (動物) だけを 1.0 の活性値で刺激したとき、ネットワークの状態および $E(t)$ は振動する可能性があるが、この状態は不安定で、素子のばらつきなどにより容易に最小エネルギーの安定状態におちつくことを説明せよ。

12.3 A/D 変換器

相互結合型ネットワークにより A/D 変換器を作りたい。入力電圧を V_{in} , LSB (最下位ビット) に相当する最小量子化ステップ電圧を v_0 , n ビットの変換結果を $x_{n-1}, x_{n-2}, \dots, x_k, \dots, x_0$ ($x_k = 0$ または 1) とする。3 ビットの A/D 変換器を実現する相互結合型ネットワークは図 12.11 で与えられることを示せ。

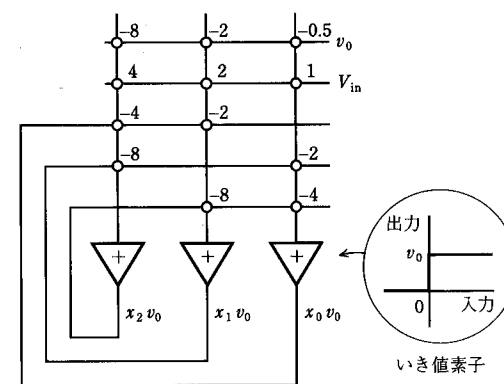


図 12.11 3 ビット A/D 変換ネットワーク (結線交点
○印の数値は結合重み)

$$-\log_e p(x) - 1 + \lambda = 0.$$

したがって

$$p(x) = e^{\lambda-1}$$

となる。この右辺は x によらない。すなわち $p(x)$ は一様分布

$$p(x) = \begin{cases} \frac{1}{2x_0} & (-x_0 \leq x \leq x_0) \\ 0 & (\text{その他}) \end{cases}$$

のとき、 $H(X)$ を最大にする。このとき、最大値 $H(X)$ は

$$\begin{aligned} H(X) &= \log_e(2x_0) & [\text{ナット}] \\ &= \log_2(2x_0) & [\text{ビット}] \end{aligned}$$

である。

9章

9.1 x_1x_2 の期待値は

$$\begin{aligned} \overline{x_1x_2} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1x_2 p(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \left\{ \int_{-\infty}^{\infty} x_1 \frac{1}{\sqrt{2\pi\sigma^2(1-\rho^2)}} \exp\left[-\frac{(x_1-\rho x_2)^2 + x_2^2(1-\rho^2)}{2\sigma^2(1-\rho^2)}\right] dx_1 \right\} \\ &\quad \cdot \frac{1}{\sqrt{2\pi\sigma^2}} x_2 \cdot dx_2 \end{aligned}$$

$$= \int_{-\infty}^{\infty} \rho x_2^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x_2^2}{2\sigma^2}\right] dx_2$$

$$= \rho\sigma^2.$$

x_1^2 の期待値は

$$\begin{aligned} \overline{x_1^2} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1^2 p(x_1, x_2) dx_1 dx_2 \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1^2 \frac{1}{\sqrt{2\pi\sigma^2(1-\rho^2)}} \exp\left[-\frac{(x_1-\rho x_2)^2}{2\sigma^2(1-\rho^2)}\right] dx_1 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x_2^2}{2\sigma^2}\right] dx_2 \\ &= \int_{-\infty}^{\infty} [\sigma^2(1-\rho^2) + \rho^2 x_2^2] \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x_2^2}{2\sigma^2}\right] dx_2 \\ &= \sigma^2(1-\rho^2) + \rho^2\sigma^2 = \sigma^2. \end{aligned}$$

なお、ここではよく知られた公式

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = 1,$$

$$\int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = \mu,$$

$$\int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] dx = \sigma^2 + \mu^2$$

を用いた。

9.2 (1) 式(9.11)より

$$\begin{aligned} \lambda(Y) &= \exp\left[-\frac{1}{2}\{(Y-S)\Sigma^{-1}(Y-S)^T - Y\Sigma^{-1}Y^T\}\right] \\ &= \exp\left[-\frac{1}{2}(Y\Sigma^{-1}Y^T - 2Y\Sigma^{-1}S^T + S\Sigma^{-1}S^T - Y\Sigma^{-1}Y)\right] \\ &= \exp\left(Y\Sigma^{-1}S^T - \frac{1}{2}S\Sigma^{-1}S^T\right) \end{aligned}$$

これと 1 との大きさを比較することから、決定規則はつぎのようになる。

$$\begin{cases} Y\Sigma^{-1}S^T \geq \frac{1}{2}S\Sigma^{-1}S^T \text{ ならば, } \hat{\theta}=1, \\ Y\Sigma^{-1}S^T < \frac{1}{2}S\Sigma^{-1}S^T \text{ ならば, } \hat{\theta}=0. \end{cases}$$

$$(2) \Sigma^{-1}S^T = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \quad S\Sigma^{-1}S^T = [1 \quad 2] \begin{bmatrix} 3 \\ 2 \end{bmatrix} = 7$$

ここで、 $Y = [y_1 \ y_2]$ とおくと、識別関数は

$$[y_1 \ y_2] \begin{bmatrix} 3 \\ 2 \end{bmatrix} - \frac{7}{2} = 3y_1 + 2y_2 - \frac{7}{2}.$$

これを 0 に等しいとおいた分離超平面(いまの場合は直線)を図示すると、図 5 のようになる。

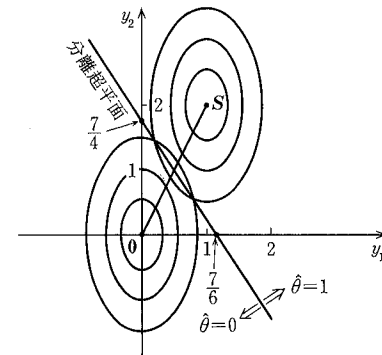


図 5

(3) 式(9.17)~(9.20)において、 $A^T = \Sigma^{-1}S^T$ 、 $\delta = \frac{1}{2}S\Sigma^{-1}S^T$ であるから、

$$\begin{aligned} \bar{z} &= \theta S\Sigma^{-1}S^T, \\ \frac{\bar{z} - \bar{z}}{(\bar{z} - \bar{z})^2} &= A\Sigma A^T = S\Sigma^{-1}S^T = \alpha^2. \end{aligned}$$

このとき、S/N 比は

$$R = S\Sigma^{-1}S^T = 7.$$

したがって、誤り率は式(9.29)および付表 3 を用いて、

$$\begin{aligned} P_e &= \Phi\left(-\frac{1}{2}\sqrt{R}\right) = \Phi(-1.323) \\ &= 1 - \Phi(1.323) = 1 - 0.907 = 0.093 \end{aligned}$$

11.2 (1) 図 11 (2) 図 12

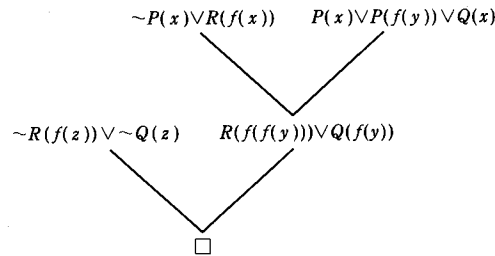


図 11 節集合 $\{P(x) \vee P(f(y)) \vee Q(x), \sim P(x) \vee R(f(x)), \sim R(f(z)) \vee \sim Q(z)\}$ の導出木。

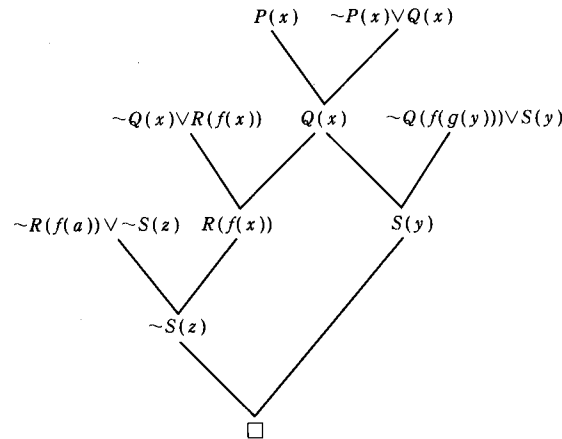


図 12 充足不可能な節集合 $\{P(x), \sim P(x) \vee Q(x), \sim Q(x) \vee R(f(x)), \sim Q(f(g(y))) \vee S(y), \sim R(f(a)) \vee \sim S(z)\}$ の導出木。

12 章

12.1 $E = [t - (xW^T)^k]^2$

最急降下法に従うと重みの修正量は、 η を正定数として

$$\Delta w_i = -\eta \frac{\partial E}{\partial w_i} = 2\eta [t - (xW^T)^k] k(xW^T)^{k-1} \cdot x_i$$

で与えられる。

12.2 (1) ネットワークの状態変化： 例 12.2 と同様に

$$\begin{aligned} s_1 &= (0, 0, 1, 0, 0) \\ s_2 &= (1, 1, 0, 1, 1) \\ s_3 &= (0, 0, 1, 0, 0) \\ s_4 &= (1, 1, 0, 1, 1) \\ s_5 &= (0, 0, 1, 0, 0) \\ s_6 &= (1, 1, 0, 1, 1) \\ &\vdots \end{aligned}$$

したがって、c(動物)とそれ以外が交互に興奮している。ただし後で示すように、このサイクルは不安定であり、素子の反応速度のばらつきなどにより容易に例 12.2 のような安定状態におちつく。

(2) ネットワークのエネルギー変化： この場合のエネルギーは

$$E(1)=0, E(2)=2, E(3)=0, E(4)=2, \dots$$

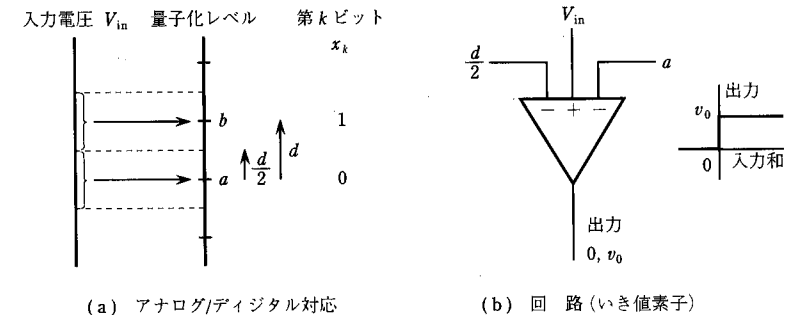
のように振動する。このことは、12.3.1 項で述べたエネルギーの単調減少性に反しているようにみえる。これは、12.3.1 項では非同期的に各素子を独立に動作させているのに対して、この解では同期的に動作させたことに原因がある。

この問題のネットワークは左右対称につくられており、重みも ± 1 の 2 種類しかない。ネットワークに非対称性、重みや素子の応答時間のばらつきなどがあれば、このサイクルから抜け出る。たとえば $t=1$ から $t=2$ へ遷移するとき、素子 e の興奮が他より少し早く起こったとする。そうすると a と d の活性化は抑えられ、b の活性化は加速されるから、e および b が興奮し、 $t=3$ 以降で b, c, e が興奮し、他は抑制された状態となり遅かれ早かれ安定化する。この安定状態のエネルギーは例 12.2(2) と同じく

$$E(t) = -3.0$$

である。

12.3 第 k ビット ($k=0, 1, \dots, n-1$) の変換について考える。他のビットはすべて確定しているとする。このとき、第 k ビットの値 x_k は図 13(a) にしたがって、つぎのように得られる。



(a) アナログ/デジタル対応

(b) 回路 (いき値素子)

図 13 第 k ビットの変換

$$x_k = \begin{cases} 0 & \left(a + \frac{1}{2}d \geq V_{in} \text{ のとき} \right) \\ 1 & \text{(その他)} \end{cases}$$

ここに、 a および b は第 k ビットをそれぞれ 0, 1 としたときの量子化レベル、 d は第 k ビットの量子化ステップ幅であり、つぎで与えられる：

$$a = (x_{n-1} \cdot 2^{n-1} + \dots + 0 \cdot 2^k + \dots + x_0 \cdot 2^0) v_0,$$

$$b = (x_{n-1} \cdot 2^{n-1} + \dots + 1 \cdot 2^k + \dots + x_0 \cdot 2^0) v_0,$$

$$d = 2^k v_0.$$

これを実現する回路は図 13 (b) で与えられる。この回路をそのまま接続すると図 14 のようになる。相互結合型ネットワークの重みは対称である。各素子はいき値素子であるから、入力を正の定数倍しても特性は変らない。そこで、第 k ビットの素子の入力重みを 2^k 倍すると、図 12.11 が得られる。

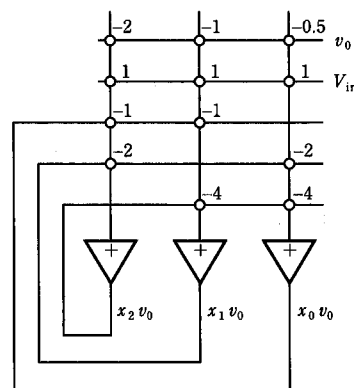


図 14 3ビット A/D 変換器の直接実現法

付 表 1

(a) 2 を底とする対数の表

n	$\log n$	n	$\log n$	n	$\log n$	n	$\log n$
1	0.000000	26	4.700439	51	5.672425	76	6.247927
2	1.000000	27	4.754887	52	5.700439	77	6.266786
3	1.584962	28	4.807355	53	5.727920	78	6.285402
4	2.000000	29	4.857981	54	5.754887	79	6.303780
5	2.321928	30	4.906890	55	5.781359	80	6.321928
6	2.584962	31	4.954196	56	5.807355	81	6.339850
7	2.807355	32	5.000000	57	5.832890	82	6.357552
8	3.000000	33	5.044394	58	5.857981	83	6.375039
9	3.169925	34	5.087463	59	5.882643	84	6.392317
10	3.321928	35	5.129283	60	5.906890	85	6.409391
11	3.459431	36	5.169925	61	5.930737	86	6.426264
12	3.584962	37	5.209453	62	5.954196	87	6.442943
13	3.700440	38	5.247927	63	5.977280	88	6.459431
14	3.807355	39	5.285402	64	6.000000	89	6.475733
15	3.906890	40	5.321928	65	6.022367	90	6.491853
16	4.000000	41	5.357552	66	6.044394	91	6.507794
17	4.087463	42	5.392317	67	6.066089	92	6.523562
18	4.169925	43	5.426264	68	6.087462	93	6.539158
19	4.247927	44	5.459431	69	6.108524	94	6.554583
20	4.321928	45	5.491853	70	6.129283	95	6.569855
21	4.392317	46	5.523562	71	6.149747	96	6.584962
22	4.459431	47	5.554589	72	6.169925	97	6.599912
23	4.523562	48	5.584962	73	6.189824	98	6.614709
24	4.584962	49	5.614710	74	6.209453	99	6.629356
25	4.643856	50	5.643856	75	6.228818	100	6.643856

(b) 2元アルファベットのエントロピーの表

$$H(p) = -p \log_2 p - (1-p) \log_2 (1-p)$$

p	$H(p)$	p	$H(p)$	p	$H(p)$	p	$H(p)$
0.005	0.045415	0.130	0.557438	0.255	0.819107	0.380	0.958042
0.010	0.080793	0.135	0.570993	0.260	0.826746	0.385	0.961497
0.015	0.112364	0.140	0.584239	0.265	0.834198	0.390	0.964800
0.020	0.141441	0.145	0.597185	0.270	0.841465	0.395	0.967951
0.025	0.168661	0.150	0.609840	0.275	0.848548	0.400	0.970951
0.030	0.194392	0.155	0.622213	0.280	0.855451	0.405	0.973800
0.035	0.218878	0.160	0.634310	0.285	0.862175	0.410	0.976550
0.040	0.242292	0.165	0.646138	0.290	0.868721	0.415	0.979051
0.045	0.264765	0.170	0.657705	0.295	0.875093	0.420	0.981454
0.050	0.286397	0.175	0.669016	0.300	0.881291	0.425	0.983708
0.055	0.307268	0.180	0.680077	0.305	0.887317	0.430	0.985815
0.060	0.327445	0.185	0.690894	0.310	0.893173	0.435	0.987775
0.065	0.346981	0.190	0.701471	0.315	0.898861	0.440	0.989588
0.070	0.365924	0.195	0.711815	0.320	0.904381	0.445	0.991254
0.075	0.384312	0.200	0.721928	0.325	0.909736	0.450	0.992774
0.080	0.402179	0.205	0.731816	0.330	0.914925	0.455	0.994149
0.085	0.419556	0.210	0.741483	0.335	0.919953	0.460	0.995378
0.090	0.436470	0.215	0.750932	0.340	0.924819	0.465	0.996462
0.095	0.452943	0.220	0.760167	0.345	0.929523	0.470	0.997402
0.100	0.468996	0.225	0.769193	0.350	0.934068	0.475	0.998196
0.105	0.484648	0.230	0.778011	0.355	0.938454	0.480	0.998846
0.110	0.499916	0.235	0.786626	0.360	0.942683	0.485	0.999351
0.115	0.514816	0.240	0.795040	0.365	0.946755	0.490	0.999711
0.120	0.529361	0.245	0.803257	0.370	0.950672	0.495	0.999928
0.125	0.543564	0.250	0.811278	0.375	0.954434	0.500	1.000000

索引

あ 行

アナログ情報 4, 5
 アナログ情報源 96
 誤り検出 72
 誤り検出能力 75
 誤り検出領域 63
 誤り訂正能力 64
 誤り訂正領域 63
 誤りの検出・訂正 2
 誤りパターン 72
 誤り率 112, 114
 r 元通信路 45
 r 元符号 30
 暗号 4
 いき値素子 145
 ECC 符号 2
 一意的復号可能 26, 28, 40
 1 次エントロピー 39, 41, 99
 位置不変性 95
 パワースペクトルの—— 95
 一様な通信路 48
 出力に関して—— 48
 2 重に—— 48, 51
 入力に関して—— 48
 意味の木 123

意味論 157
 因果律 88
 因子 136
 インパルス応答 88
 エイリアシング 5, 78, 93
 エキスパートシステム 7
 XOR 145
 枝 28
 A/D 変換 5, 78
 A/D 変換器 153
 (n, k) 符号 58
 n 次エントロピー 40, 100
 n 次拡大情報源 39, 40
 エネルギー 147
 MR 符号 2
 m 重マルコフ情報源 18
 LSB 153
 エルゴード情報源 16
 エルゴード性 16, 97
 エルゴード的 16
 エルブラン解釈 161
 エルブラン基底 160
 エルブラン空間 132, 159
 エントロピー 13, 31, 40, 42, 99, 100
 結合—— 50

重み 62
 折り返し歪 5, 95

か行

解釈 118

階層型ネットワーク 8, 142

ガウス分布 97, 107

学習 8, 143

拡大の手法 51

確率 9, 10

活性値 142

過渡状態部分 21

加法性 62
 線形符号の—— 62

加法的白色ガウス通信路 101

含意 120

含意記号
 ——の除去 130

慣性項 145

完全
 関数的に—— 117

完全事象系 12

完全性 159

完全性定理 159

偽(F) 115

記憶がない
 情報源に—— 3

記憶のある情報源 18

記憶のない情報源 15

記憶のない通信路 46

記号論理学 7

奇数パリティ検査 59

基礎例 160

基本命題 115

基本論理式 115, 127

既約多項式 74

既約マルコフ情報源 22

教師付き学習 142

共分散行列 108

距離の三公理 62

偶数パリティ検査 59

空節 132, 133

組合せ問題 8

クラフトの不等式 29, 30

形式的証明 158

結合エントロピー 50

結合確率密度関数 98

結合則 135

決定規則 111

決定不能 129

限界距離復号法 63

検査記号 58

検査ビット 60

原始多項式 75

限定作用素 128

限量記号 128

項 127

高階述語論理 129

恒偽性 124, 132

恒真 118

恒真性 123, 124, 129

構文論的方法 157

公理系
 命題論理の—— 157

効率 53, 58

効率符号の—— 53

誤差逆伝搬学習アルゴリズム 142

誤差伝搬量 144

個体定項 127

個体変項 127

根元事象 9

コンパクト 34, 36

コンパクト符号 34, 37

コンマ符号 27

さ行

最急降下法 145, 153

最小重み 62, 63

最小距離 63, 74

最小ハミング距離 63

最大エントロピー定理 100

最短符号 34

作用範囲 128

三角不等式 62

三段論法 7, 121, 126, 133, 157

サンプリング 78

CRC 68, 72

CRC 符号 2

時間平均 16, 97

シグモイド関数 141

事後エントロピー 49

事後確率 11, 109

自己情報量 12

自己相関関数 90

CCITT 68, 75

事象 9

事前エントロピー 49

事前確率 11

質問応答 139

質問応答システム 7

シャノンの補助定理 32

シャノン・ハートレーの定理 102

シャノン理論 1

周期 71

周期的状態部分 21

集合平均 16, 97

集合論的演算 9

充足可能 118

充足する 118

充足性 129

充足不可能 118, 129, 132, 136

周波数成分 83

自由変項 128

受信記号 45

述語 127

述語論理 7, 127

述語論理式 128, 129

出力アルファベット 45

出力記号 45

出力誤差 143

出力層 142

巡回セールスマン問題 151

巡回ハミング符号 75

巡回符号 68, 69, 71

瞬時符号 27, 28, 29, 31

条件付きエントロピー 49

条件付き確率 109, 10, 11, 97

消失 47

状態 18, 21

状態確率分布 23

状態図 18

状態分布 23

商多項式 70

冗長度 54
 符号の—— 54

情報源 2

情報源アルファベット 15

情報源記号系列
 平均的に長い—— 42

情報源符号化 2, 3, 27

情報源符号化定理 39, 40

情報速度 53

情報伝送速度 53, 56

情報ビット 59

情報量 5, 9, 13

情報理論 1, 9

証明可能 158

剰余多項式 70

真(T) 115
 常に—— 119

真偽 116

神経回路 8

シンドローム 65

振幅スペクトル密度 83

真理値 115, 129

水平垂直パリティ検査符号 60
 水平垂直パリティ検査方式 68
 推 論 7, 115
 推論規則 117, 152
 スコーレム関数 131
 スペクトル 83

 正規分布 97, 107
 正規マルコフ情報源 22, 23
 整合機械 111
 整合フィルタ 111
 生成行列 61
 生成多項式 75
 世 界 129
 積事象 10
 積標準形 131
 節 124, 130, 132
 節 点 28
 接頭語 27
 接頭語条件 27
 遷移確率行列 19
 遷移図 18
 線形識別関数 145
 線形識別方式 112
 線形符号 58, 61
 選 言 124
 選言標準形 124, 125
 全称記号 128
 —の除去 132
 前置形 131

 相関係数 108
 相互結合型ネットワーク 8, 146
 相互情報量 48, 49, 54
 送信記号 45
 測 度 10
 束縛変項 128
 組織符号 58
 存在記号 128
 —の除去 131

た 行
 帯域制限 96
 第一階述語論理 129
 代 入 132
 多項式 68
 畳み込み積分 89
 wff 117, 127
 単位インパルス関数 86
 単一誤り 72
 単一誤り訂正ハミング符号 66
 単一化 133, 134
 単一化アルゴリズム 135
 単一化置換
 最も一般的な—— 134
 単一パリティ検査符号 59
 探 索 137
 単純マルコフ情報源 18

 置 換 134
 —の合成 134
 知的符号化 4
 中間層 142
 超関数 87
 超平面 112
 直線量子化 103
 直交関数展開 86

 通信系モデル 2, 109
 通信路 45
 通信路行列 46, 48
 通信路線図 47
 通信路符号化 2, 3, 25
 通信路符号化定理 56
 通信路モデル 45
 通信路容量 102, 51, 56

 低域通過フィルタ 93
 定常状態 23
 定常情報源 15, 97
 定常通信路 46

デイビス・パットナムのアルゴリズム 155
 デイビス・パットナムの方法 124, 133
 定 理 158
 適応的符号化 4
 δ 関数 86
 典型的系列 54
 伝達関数 89

 統計的パターン認識 6
 導 出 133
 導出原理 7, 130, 132, 137
 導出節 133, 156
 同 値 120
 等長符号 26
 特徴ベクトル 107
 閉じた状態部分 21
 閉じている 62
 加法に関して—— 62
 ド・モルガンの法則 125

 な 行
 ナット 13

 2元対称消失通信路 52
 2元対称通信路 47, 52
 2次元エントロピー 41
 2次拡大情報源 38
 二重否定 125
 ニューラルネットワーク 8, 141
 —の形態 141
 入力アルファベット 45
 入力記号 45
 入力層 142

 根 28

 は 行
 葉 28
 バースト誤り 67, 73, 74
 バースト誤り訂正(検出)能力 67

パーセバルの定理 95
 倍多項式 69
 排他的論理和 59, 68, 145
 白色ガウス情報源 98
 パターン認識 6, 8, 107, 109, 142
 パターンの記憶 150
 破綻節点 161
 幅優先探索 138
 ハフマン符号 34, 42
 ハフマンブロック符号 39, 43
 ハミング重み 62
 ハミング距離 61
 ハミング符号 66
 パラメータ表現
 命題論理式の—— 123
 パリティ検査 73
 パリティ検査行列 64
 パリティ検査ビット 25
 パリティ検査方程式 59, 64
 パワースペクトル 90

 BSC 47
 引 数 127
 BCH 符号 77
 非周期的状態部分 22
 非瞬時符号 27
 非直線量子化 103
 ビット 13
 否定記号
 —の個別化 130
 非等長情報源記号系列 42
 非等長符号 26
 BPN 143
 標準ガウス分布関数 112
 標準形 124
 命題論理式の—— 124
 標本化 4, 78
 標本化関数 85
 標本化定理 4, 78, 85
 標本化パルス 90