

# 写像とは

---

集合の各元を他の集合(または同じ集合)の元にそれぞれ対応させること。すなわち、集合から集合への変換のこと。写像  $f$  が集合  $A$  から集合  $B$  へ変換する役割を担っている場合

$$f : A \rightarrow B$$

などと表現する。また、集合  $A$  から集合  $B$  への写像を考えたとき、気木ウンのに要素の順番というのは順不同となる。すなわち、集合  $A$  から集合  $B$  へ対応する要素が過不足無く含まれていれば問題ない。集合と写像は非常に抽象的な概念で、ほとんど具体的な制限が掛かっていない論理体系である。従って、近代的な数学の基礎となっている。

## 写像の具体例：確率

### 確率変数

「確率変数」という名前でありながら、これも写像の例だととらえることができる。

コインを投げた場合、集合  $A = \text{表, 裏}$  の要素のいずれかが発生する。この際に、数値でわかりやすくしておくために、集合  $B = 0, 1$  というものを準備しておいて、写像を考える。

$$X : A \rightarrow B$$

つまり確率変数は、具体的な現実の事象を、予め準備していた数値の集合へ写像してくれるものだと考えることができる。

### 確率密度関数

確率密度関数  $p$  も列記とした写像です。「関数」は全て写像だと考えてもよい(分野が異なるだけで定義は同じ)。確率密度関数として最も馴染み深いものは

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

このガウス分布かもしれません。  $x$  の値を入力すると、値域  $[0, 1]$  の値を返してくる。この帰ってきた値を、  $x$  が発生する確率だと捉えるわけです。もちろん  $x$  というのは**確率変数**  $X$  の要素の1つであり、そしてこのケースでは  $X = [-\infty, \infty]$  という集合を扱っていることになります。(  $\mu, \sigma$  はガウス分布のパラメータです)

# 写像ニューラルネットワーク

---

写像ニューラルネットワークが扱う問題は、  $n$  次元ユークリッド空間の有界部分集合  $A$  から  $m$  次元ユークリッド空間の有界部分集合  $f[A]$  への有界写像または関数  $f : A \subset R^n \rightarrow R^m$  を、その写像動作の実際の事例  $(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k), \dots$  に基づく訓練手段により、近似的に実現することである。

## 種類

基本的に、2つのタイプの写像ネットワークがある。すなわち、**特徴による**(feature-based)ネットワークまたは単に**特徴**ネットワークと、**プロトタイプ**ネットワークまたは単に**プロトタイプ**ネットワークである。

### 特徴ネットワーク

特徴ネットワークは、ある関数の入出力関係を実現する。それは、一般的かつ変更可能な関数形で表現される。この関数形は近似される写像にフィットするように、(典型的には、ネットワーク処理要素内で、1つ以上の学習則を適用することにより重みを適応的に設定することを介して)変更される。ネットワークの例としては、

- バックプロパゲーション
- データ処理の群的手法(Group Method of Data Handling, GMDH)

がある。

## 関数近似精度の測定

---

現在, 感心を持っている写像ネットワークはすべて, 結合データを変えるのではなく, 適応的な係数(すなわち重み)を変えることにより事例データに自己適応すると仮定する.

ベクトル $\mathbf{x}$ (ある固定された確率密度関数  $\sigma(\mathbf{x})$  に従ってランダムに選ばれと仮定する) が写像ニューラルネットワークに入れられたとき, そのネットワークの出力ベクトルを $\mathbf{Y}(\mathbf{x}, \mathbf{w})$ と表す. ここで,  $\mathbf{w}$ はネットワークの重みベクトルである. ネットワークの近似精度を測るには, 多数のテスト試行について, ネットワークの実際の出力  $\mathbf{Y}(\mathbf{x}, \mathbf{w})$ を, "正しい" 出力  $f(\mathbf{x})$ と何らかの方法で比較しなければならない.

## テスト集合

### 定義

1. 写像ネットワークの近似精度テストするためだけに使われる事例の集合(すなわち, **訓練で用いられていない事例の集合**)を, **テスト集合(test set)** と呼ぶ.
2. テスト集合は, ランダムに選ばれた事例  $(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_k, \mathbf{y}_k), \dots$  である.

## テスト試行

### 定義

1. テスト集合を入力とするネットワークの出力と写像が与える正しい関数値とを比較し, ネットワークをテストする.
2. テストは1事例ごとに行われる.

## 平均2乗和誤差

$k$  番目のテスト試行に使われた事例を  $(\mathbf{x}_k, \mathbf{y}_k)$  とする (すなわち  $\mathbf{y}_k = f(\mathbf{x}_k)$ ). 以前と同じく,  $\mathbf{x}_k$  は, ある固定された確率密度関数  $\rho$  に従い, 領域集合  $A$  からランダムに取り出されたと仮定する. 次に,

$$\mathbf{F}_k(\mathbf{x}_k, \mathbf{w}) = |f(\mathbf{x}_k) - \mathbf{Y}(\mathbf{x}_k, \mathbf{w})|^2$$

とする.

$\mathbf{F}_k$ は $k$

番目のテスト試行において, 写像ネットワークにより作られた近似誤差の2乗である. この近似誤差の検討のために,  $\mathbf{w}$  はテストの間, 固定されていると仮定する. テスト試行を繰り返し, すべての  $\mathbf{F}_k$  が得られると, ネットワークの**平均2乗誤差**  $\mathbf{F}(\mathbf{x})$  は,  $\mathbf{F}_k(\mathbf{x}_k, \mathbf{w}) \equiv \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N \mathbf{F}_k(\mathbf{x}_k, \mathbf{w})$  と定義される. ここで極限は, ランダムに選ばれた  $\mathbf{x}_k$  のほとんど全ての集合に対し存在すると仮定する.  $\mathbf{F}$ は負でない量の平均なので  $\mathbf{F}(\mathbf{w}) \geq 0$  に注意しよう. あるいは, 確率積分を用いて, 次式のように  $\mathbf{F}(\mathbf{w})$  を定義できる.  $\mathbf{F}(\mathbf{x}) \equiv \int_A |f(\mathbf{x}) - \mathbf{Y}(\mathbf{x}, \mathbf{w})|^2 \rho(\mathbf{x}) d\mathbf{x}$  平均2乗誤差  $\mathbf{F}(\mathbf{w})$  はほとんどのニューラルネットワークに対し, うまく定義される. しかし, 平均2乗和誤差がうまく定義されないようなニューラルネットワークを定めることも可能である. ここで"うまく定義される"とは, 式(1)の極限が"ほとんど確実に"(すなわち確率1で)収束すること, あるいは, 関数  $|f(\mathbf{x}) - \mathbf{Y}(\mathbf{x}, \mathbf{w})|^2 \rho(\mathbf{x})$  が領域  $A$  で可積分であることの少なくとも一つが成り立つことを意味する. 以後  $\mathbf{F}(\mathbf{w})$  は存在するものと仮定する. さらに,  $\mathbf{F}(\mathbf{w})$  は, ほとんどいつも, 連続かつ微分可能な $\mathbf{w}$ の関数であると仮定する.

## 平均2乗和誤差の特徴

$\mathbf{F}(\mathbf{w})$  の注目すべき点は, 私たちが  $\rho$  に従い, テスト事例をランダムに選択する限り, どんな事例を使おうとも違いがないことである.

平均2乗誤差は, 興味ある唯一の誤差というわけではない. しかしながら, それは, がぜん, 最もポピュラーなものである. 平均2乗和誤差の特徴は, 誤差ベクトル

$$f(\mathbf{x}_k) - \mathbf{Y}(\mathbf{x}_k, \mathbf{w})$$

の大きさの2乗に応じて訓練の試行誤差を一様に重みづける点である.

この誤差測定方法は, 大きな誤差は小さな誤差よりはるかに大きな注目を受けることを保証し, これは通常望まれることである(ほとんどの場合, 大きな誤差は小さな誤差よりはるかに大きな損害を与える). また, 平均2乗誤差は, 個々の入力の発生頻度を考慮している. その結果, 平均2乗誤差は, 稀な入力の誤差よりも, よく出会う入力の誤差に, はるかに影響を受けやすい.

## 平均2乗誤差の欠点

ネットワーク性能の平均2乗誤差測定は、あらゆる場合に望ましいものというわけではない。例えば、写像およびネットワーク出力が画像の場合、平均2乗誤差は人間の観測者にとって意味のある誤差測定とはならないだろう。その具体例として、写像の出力画像とネットワークの出力画像間において、256グレースケール(明度)レベルのうち、30レベルだけが一定オフセットの形で異なり、あとは同じ画像の場合、その平均2乗誤差は900と計算される。しかし、観測した人は恐らく、これら画像間に何ら意味のある違いを感じることはないだろう。一方、写像の出力画像とネットワークの出力画像が、100画素あたり平均1画素が(0から255の間の)ランダムな値に設定されている点だけを除き、あとは同じ場合、平均2乗誤差は小さいが、2つの画像を観測した人には**非常**に違って見えるだろう。このように画像に対しては、平均2乗誤差評価基準は、意味のある誤差を測るのに適切ではない。

## 重み空間との関係

平均2乗誤差  $F(\mathbf{w})$  は評価されるニューラルネットワークの重みベクトル  $\{\mathbf{w}\}$  の関数である。違った重みを選べば違った平均2乗誤差が得られる。従って、 $F(\mathbf{w})$  はネットワークの重み空間"上"に位置する表面と考えることができる。ここで  $F$  は重み  $\mathbf{w}$  における表面の高さである。この表面はニューラルネットワークの**誤差表面**(error surface)と呼ばれる。

$F$  は非負関数だから、誤差表面が常に重み空間上の非負の高さにあることは明白である。明らかに目標は、 $F$  を最小化するような重みを見つけることである。ほとんどの場合、ニューラルネットワークは、希望の写像を**正確**に実現することはできないだろうから、 $F$  の最小値は0ではないだろう。すなわち、 $F$  の最小値は、典型的には、 $F_{min} > 0$  となろう。

後にわかるように、誤差表面の構造は、バックプロパゲーション・ニューラルネットワークにとって、非常に重要な問題である。他のニューラルネットワークにとっては、単に、ネットワークの性能を絶対値の形で理解するための一方法にすぎない。