第3章 それぞれの特徴から ディープ・ラーニングにおける役割まで

# 常微分 / 偏微分 / 全微分の基礎

新井 正敏

ディープ・ラーニングにおいて，**微分**，**偏微分**，**全微分**は学習と最適化において重要な役割を果たします．特にバックプロパゲーションでは，ディープ・ニューラル・ネットワークの重みとバイアスを更新し，誤差を最小化するために使用されます．

バックプロパゲーションにおける微分の役割は関数の勾配を求めることであり，勾配により学習を進める方向を決める重要な指針となります．このため，本章でバックプロパゲーションを説明する前に微分について学びなおしておきたいと思います．

微分は注目する変数により，微分（常微分），偏微分，全微分の3種類があります．まずはそれぞれの特徴について説明し，3つの微分と**合成関数の微分**（チェインルール），**離散微分**，**ベクトル解析**（ベクトル場の変化）との関係についても触れます※．

主に常微分，偏微分，全微分の3種類の基本的な部分をしっかり学び直します．

※チェインルール，ベクトル解析に関しては第5章のバックプロパゲーションのからくりの説明の際に詳細説明します．離散微分については，第9章並びに第10章で示すCNNのフィルタの所で詳細を説明します．

## 3微分とそれぞれの特徴

高校数学で学んだ微分は常微分になります．大学初等で学ぶ偏微分や全微分も基本的な考え方は常微分です．ここでは，常微分，偏微分，全微分の特徴について**表1**に基づいて俯瞰します．

**表1　3微分のディープ・ラーニングにおける役割**

| 種　別 | 本特集の応用 |
|---|---|
| 常微分 | 1ニューロンの1つの重みの学習 |
| 偏微分 | 各層の複数の重み，バイアスの学習 |
| 全微分 | 学習の勾配ベクトル |

### ● 常微分 / 偏微分 / 全微分の概略

**▶常微分**

独立変数$x$のみに注目し，他の変数は考慮しません．常微分は単純な変化率を表すため，微分の最も基本的な形です．

- 一般式

$$\frac{df}{dx} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x) - f(x)}{\Delta x}$$

- 応用例

学習の際1つの重み変数$w$のみに依存する損失関数$E(w)$に対して，その変化率を表す際に使用されます．重みに対する更新量を計算するため，勾配降下法での更新式に用います．

**▶偏微分**

複数の変数を含む関数に対して，特定の変数についてのみ微分し，他の変数は定数として扱います．

- 一般式

$$\frac{\partial f}{\partial x} = \lim_{\Delta x \to 0} \frac{f(x+\Delta x, y, z, \ldots) - f(x, y, z, \ldots)}{\Delta x}$$

- 応用例

多層ニューラル・ネットワークにおける各層の重みやバイアスに関してのみの変化率を求め，それらを同時に最適化するために用いられます．

**▶全微分**

全ての変数の微小変化が関数に与える総合的な影響を求める際に使用されます．

- 一般式

$F(x,\ y,\ z)$の場合，全微分は次のように表されます

$$dF = \frac{\partial F}{\partial x}dx + \frac{\partial F}{\partial y}dy + \frac{\partial F}{\partial z}dz$$

- 応用例

主に微小変化が複数の変数にわたる場合に利用され，特に勾配ベクトルの概念を理解する際に重要です．

### ● 深掘りするための3要素

これら3種類の微分を基本としてバックプロパゲー