第8章 モーメンタムの概念を使って安定的な学習を進めるメカニズムを解き明かす

ベクトル解析の学び直しと転移学習のからくり

新井 正敏

第5章では，バックプロパゲーションによる学習のからくりを見てきました．中でも，確率的勾配降下法(SGD)の概略と課題についても触れました．また，第7章では，SGDにモーメンタムの概念を取り入れたSGDMとAdamを実際に使ってみて学習の経過から特性が把握できたと思います．

本章では，まずSGDの式をベクトル微分演算子$\vec{\nabla}$(ナブラ)を使って表します．∇は，一般的に矢印(ベクトル記号)を付けません．本特集ではベクトルであることを明示するためにあえて矢印を使っています．MATLABを使って直観的にベクトル解析の概念を学び直します．その上で，モーメンタムの概念を$\vec{\nabla}$を使って説明し，SGDMやAdamのからくりを解き明かします．

関数の勾配とベクトル解析

損失関数に対する重みの勾配と更新は，第5章の式(5)のように行われました．

$$w_{t+1} = w_t - \eta \boxed{\frac{\partial E}{\partial w}}\bigg|_{w=w_t}$$
$$w_{t+1} = w_t - \eta \boxed{\vec{\nabla} E(w_t)} \quad \cdots\cdots(1)$$

第5章の式(5)の勾配は，式(1)のようにベクトル微分演算子$\vec{\nabla}$を使って表されます．なお，$\vec{\nabla}$は式(2)のように定義されたベクトルです．

$$\vec{\nabla} \equiv \left(\frac{\partial}{\partial u_1}, \frac{\partial}{\partial u_2}, \cdots, \frac{\partial}{\partial u_n}\right) \quad \cdots\cdots(2)$$

ここで，関数の勾配と$\vec{\nabla}$の関係について，しっかりと学び直します．

● 2変数関数の勾配と$\vec{\nabla}$の関係

1変数の関数$f(u)$の勾配(微分)は，次式のように表されました．本特集では$\frac{d}{du}$をスカラ微分演算子(一般的には微分演算子)と呼ぶことにします．

$$\mathrm{grad}\ f(u) = \frac{df(u)}{du} = \frac{d}{du}f(u) \quad \cdots\cdots(3)$$

一方，n個の多変数関数$f(u_1, u_2, \cdots, u_n)$の勾配を式(3)から考えてみます．

多変数関数の勾配は各変数に関数する偏微分のベクトルになり，式(4)のようになります．

$$\mathrm{grad}\ f(u_1, u_2, \cdots, u_n) = \left(\frac{\partial}{\partial u_1}, \frac{\partial}{\partial u_2}, \cdots, \frac{\partial}{\partial u_n}\right) f(u_1, u_2, \cdots, u_n) \quad \cdots\cdots(4)$$

ここで，式(2)の$\vec{\nabla}$を適用すると，多変数の勾配は式(5)のようになります．

$$\vec{\nabla} f = \left(\frac{\partial f}{\partial u_1}, \frac{\partial f}{\partial u_2}, \cdots, \frac{\partial f}{\partial u_n}\right) \quad \cdots\cdots(5)$$

$\vec{\nabla} f$は関数fの勾配ベクトルであり，各成分は関数fを変数u_1，u_2，…，u_nで偏微分したものになります．

2変数$f(u_1, u_2)$の勾配は式(6)になります．

$$\vec{\nabla} f(u_1, u_2) = \left(\frac{\partial f}{\partial u_1}, \frac{\partial f}{\partial u_2}\right) \quad \cdots\cdots(6)$$

式(6)の勾配ベクトルは，関数の各方向に対する変化率を表しています．すなわち，関数$f(u_1, u_2)$がどの方向u_1，u_2にどれだけ急勾配で増加または減少しているかを示しています．

本特集では，主に損失関数について扱っています．損失関数に対する$\vec{\nabla}$(重みベクトル)に関する微分で，w_t(t時点での重み)の値を示すと，式(7)となります．

$$\vec{\nabla} E(w_t) = \left(\frac{\partial E(w_t)}{\partial u_1}, \frac{\partial E(w_t)}{\partial u_2}\right) \quad \cdots\cdots(7)$$

このため，第5章の式(5)の勾配は，式(1)と同等であることが分かります．

2変数関数の勾配と$\vec{\nabla}$についてまとめます．

- $\vec{\nabla}$はベクトル微分演算子で，勾配や発散，回転を計算するために使用する