

自動微分

柏木 雅英

kashi@waseda.jp

<http://verifiedby.me/>

早稲田大学 基幹理工学部 応用数理学科

自動微分 (1)

自動微分

関数 $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ (の計算手順) と、ある $a \in \mathbb{R}^n$ が与えられたとき、 $f(a)$ を計算すると同時に $f'(a)$ の計算を行うアルゴリズム。

- そのような値は、いわゆる数値微分 (例えば

$$\frac{\partial f_j}{\partial x_i}(a) \simeq \frac{f_j(a_1, \dots, a_i + \varepsilon, \dots, a_n) - f_j(a_1, \dots, a_i, \dots, a_n)}{\varepsilon}$$

とする) でも計算できるが、自動微分は数値微分と違って離散化誤差を含まない (誤差があるとすれば丸め誤差のみ)。また、数値微分は ε が大きすぎれば離散化誤差が増大し、小さすぎれば桁落ちによる誤差が増大するという困難を抱えているが、そのような問題は発生しない。

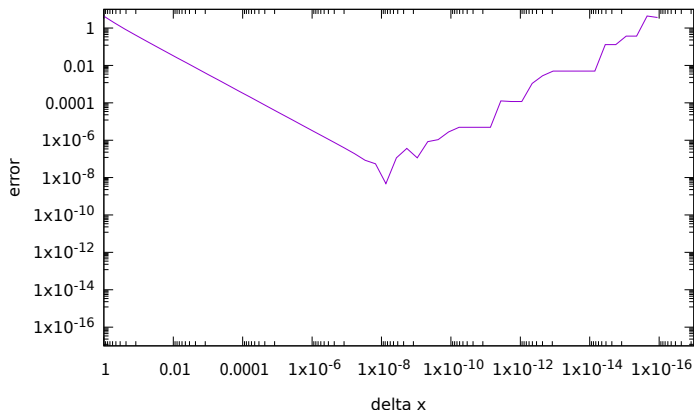
- Bottom Up 型 (forward mode) と Top Down 型 (backward mode) がある。
- 近年、機械学習のアルゴリズムを支える技術として自動微分が再注目されている。

数値微分の例

$f(x) = (x^2 + 1)(x - 3) + 2x$ について、 $f'(2)$ の値を、

$$\frac{f(2 + \Delta x) - f(2)}{\Delta x}$$

で (IEEE754 倍精度で) 計算したときの誤差。



自動微分 (2)

$f: \mathbb{R} \rightarrow \mathbb{R}$ (**1 変数関数**) と $a \in \mathbb{R}$ に対して、 $f(a)$ を計算すると同時に $f'(a)$ を Bottom-Up 型で計算することを考える。

方針

関数 $f(x)$ の計算中に現れる全ての中間変数 $f_1, f_2, \dots$ について、 $(f_i(a), \frac{df_i}{dx}(a))$ の値のペアを保持する。

アルゴリズム

$x = (a, 1)$ と初期化。以下、加減乗除と数学関数 h について、

- $(v_1, d_1) \pm (v_2, d_2) \rightarrow (v_1 \pm v_2, d_1 \pm d_2)$
- $(v_1, d_1) \times (v_2, d_2) \rightarrow (v_1 \times v_2, v_2 d_1 + v_1 d_2)$
- $(v_1, d_1) / (v_2, d_2) \rightarrow (v_1 / v_2, (1/v_2) d_1 + (-v_1 / (v_2)^2) d_2)$
- $h(v_1, d_1) \rightarrow (h(v_1), h'(v_1) d_1)$

の演算規則で計算。ただし式中の定数 c は、 $(c, 0)$ とみなす。最終的に得られた値を (v_f, d_f) とすると、 $v_f = f(a)$, $d_f = f'(a)$ となっている。

自動微分 (3)

例

$f(x) = (x^2 + 1)(x - 3) + 2x$ について、 $f(2)$ と $f'(2)$ の値を計算してみる。

- $x : (2, 1)$
- $x^2 : (2, 1) \times (2, 1) = (4, 4)$
- $x^2 + 1 : (4, 4) + 1 = (5, 4)$
- $x - 3 : (2, 1) - 3 = (-1, 1)$
- $(x^2 + 1)(x - 3) : (5, 4) \times (-1, 1) = (-5, 1)$
- $2x : 2 \times (2, 1) = (4, 2)$
- $(x^2 + 1)(x - 3) + 2x : (-5, 1) + (4, 2) = (-1, 3)$

実際、 $f(2) = -1$ で、 $f'(x) = 3x^2 - 6x + 3$ なので $f'(2) = 3$ であることが確かめられる。

自動微分 (4)

$f: \mathbb{R}^n \rightarrow \mathbb{R}$ (n 変数関数) と $a \in \mathbb{R}^n$ に対して、 $f(a)$ を計算すると同時に $\frac{\partial f}{\partial x_1}(a), \dots, \frac{\partial f}{\partial x_n}(a)$ を計算することを考える (Bottom Up 型)。

方針

関数 $f(x)$ の計算中に現れる全ての中間変数 $f_1, f_2, \dots$ について、 $(f_i(a), \frac{\partial f_i}{\partial x_1}(a), \dots, \frac{\partial f_i}{\partial x_n}(a))$ の $n+1$ 個の値を保持する。

アルゴリズム

以下、保持する値を、 $v \in \mathbb{R}$ 、 $d \in \mathbb{R}^n$ のペアとして (v, d) のように書く。 $x_i = (a_i, e_i)$ ($1 \leq i \leq n$) と初期化 (e_i は単位ベクトル)。以下、加減乗除と数学関数 h について、

- $(v_1, d_1) \pm (v_2, d_2) \rightarrow (v_1 \pm v_2, d_1 \pm d_2)$
- $(v_1, d_1) \times (v_2, d_2) \rightarrow (v_1 \times v_2, v_2 d_1 + v_1 d_2)$
- $(v_1, d_1) / (v_2, d_2) \rightarrow (v_1 / v_2, (1/v_2) d_1 + (-v_1 / (v_2)^2) d_2)$
- $h(v_1, d_1) \rightarrow (h(v_1), h'(v_1) d_1)$

の演算規則で計算。ただし式中の定数 c は、 $(c, 0)$ とみなす。最終的に得られた値を (v_f, d_f) とすると、 $v_f = f(a)$ 、 $d_f = f'(a)$ となっている。

自動微分 (5)

例

$f(x, y) = (x - y)(x^2 + y) + xy$ について、 $a = (2, 1)$ として、 $f(a)$ 、 $\frac{\partial f}{\partial x}(a)$ 、 $\frac{\partial f}{\partial y}(a)$ を計算してみる。

- $x : (2, 1, 0)$
- $y : (1, 0, 1)$
- $x - y : (2, 1, 0) - (1, 0, 1) = (1, 1, -1)$
- $x^2 : (2, 1, 0) \times (2, 1, 0) = (4, 4, 0)$
- $x^2 + y : (4, 4, 0) + (1, 0, 1) = (5, 4, 1)$
- $(x - y)(x^2 + y) : (1, 1, -1) \times (5, 4, 1) = (5, 9, -4)$
- $xy : (2, 1, 0) \times (1, 0, 1) = (2, 1, 2)$
- $(x - y)(x^2 + y) + xy : (5, 9, -4) + (2, 1, 2) = (7, 10, -2)$

実際、 $f(2, 1) = 7$ で、 $\frac{\partial f}{\partial x} = 3x^2 - 2yx + 2y$ なので $\frac{\partial f}{\partial x}(2, 1) = 10$ 、
 $\frac{\partial f}{\partial y} = -2y + 2x - x^2$ なので $\frac{\partial f}{\partial y}(2, 1) = -2$ であることが確かめられる。

自動微分 (6)

(注意)

- 二項演算の演算規則は、その演算を $g(x, y)$ として

$$g((v_1, d_1), (v_2, d_2)) \rightarrow (g(v_1, v_2), \frac{\partial g}{\partial x}(v_1, v_2)d_1 + \frac{\partial g}{\partial y}(v_1, v_2)d_2)$$

のように作られている。

- n 変数関数の自動微分は、「ある i を固定し、 x_i のみを変数、残りの変数を定数とみなすことによって f を自動微分で計算」を $1 \leq i \leq n$ に対して繰り返すことによって実行できる。この場合、1 変数用のプログラムがそのまま使えるが、 $f(a)$ という同じ値の計算を n 回繰り返すことになり計算量が増えてしまう。
- Top Down 型 (backward mode) の自動微分というアルゴリズムもあるが、本稿では説明を省略する。一般に、 $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ の自動微分の計算量は、自動微分を行わない場合の f の計算量に比べて、
 - Bottom Up なら、 n 倍程度で、 m によらない。
 - Top Down なら、 m 倍程度で、 n によらない。ということが知られている。