最急降下法

大枝 真一

1 はじめに

機械学習では，なんらかの数理モデルを仮定する．このときに，目的関数（誤差関数）を設定して，この関数の最小値（or 最大値）求める必要がある．この関数が解析的に陽に解けるときは問題ないが，一般的に機械学習で扱う目的関数は非常に複雑であるため，解析的に陽に解けないことがほとんどである．あるいは，人間が時間をかけて解析的に解いて厳密解を得るよりも，計算機を用いて近似値を短時間で求める方が有益なことが多い．このような繰り返しの計算で代表的な手法が最急降下法である．

2 最急降下法

ここでは，最急降下法の説明のため，式 (1) に示す目的関数の最小値を求めることを考える．

$$f(x) = (x-2)^2 + 1 \tag{1}$$

関数を図に示すと図 1 のようになる．

図 1 からわかるように，この関数は x=2 のときに最小値をとる．今は，最小値の値にはとくに関心がなく，x がいくらのとき最小値をとるかを知りたい．このような目的関数への入力 x はパラメータと呼ぶ．一般的な機械学習では事前に設定した目的関数を最小にするパラメータを，観測したデータから求めることになる．

2.1 解析的に解く

関数の最小値を求めるには，関数の凸性を調べた後，微分して 0 になる点を探す．

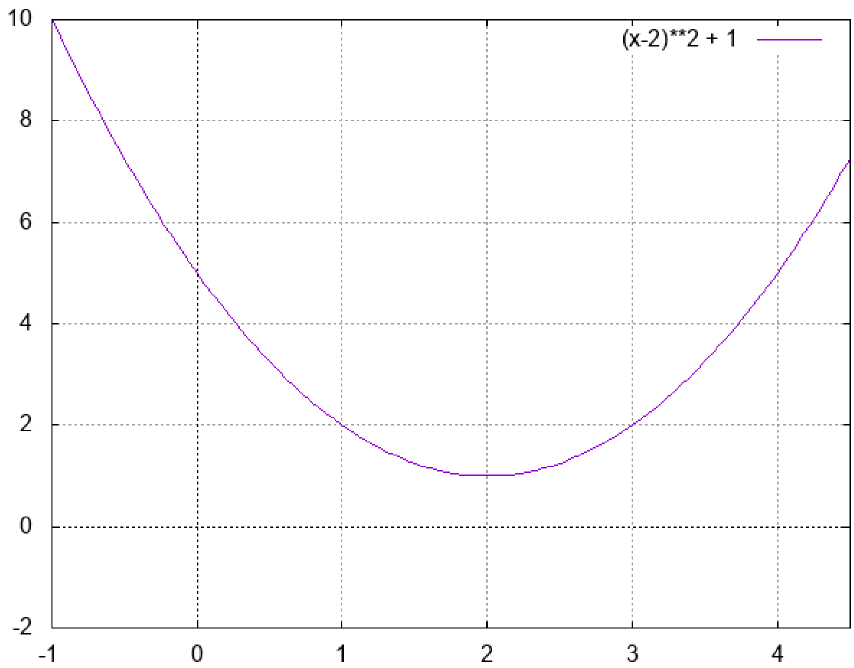

図 1　求めたい目的関数

式 (1) を微分すると式 (2) のようになる．

$$f'(x) = 2(x-2) \tag{2}$$

これが 0 になるとき，関数の傾きが 0 であることがわかるので，式 (1) は関数の形状より，下に凸で最小値を 1 つしか持たないことがわかるので，$f'(x) = 2(x-2) = 0$ より，$x = 2$ のとき最小値をとる．

2.2　最急降下法で解く

最急降下法は，関数 $f(x)$ の傾きから繰り返し x を更新し，最小値（厳密には極小値）を求める反復法である．

最急降下法の更新は以下の式で行う．

$$x_{t+1} = x_t - \eta \frac{\partial f(x_t)}{\partial x_t} \tag{3}$$

η は学習係数で $\eta = 0.01$ など小さい値をあらかじめ決めておく．η を大きくするほど学習速度は向上するが大きすぎると収束せず振動や発散をしてしまう．逆に小さいほど，

学習は安定して収束するが学習速度が低下してしまう．したがって，η は適切な値の設定が必要となる．

最初に適当に x の初期値を決める．初期値から偏微分 $\frac{\partial f(x)}{\partial x}$ を求め，式 (3) に代入し，x を更新する．このとき，x が極小値のときよりも大きければ傾きは $\frac{\partial f(x)}{\partial x} > 0$ となるため，x は小さくなり，$f(x)$ は極小値に近づく．それに対して，x が極小値のときよりも小さければ傾きは $\frac{\partial f(x)}{\partial x} < 0$ となるため，x は大きくなり，$f(x)$ はやはり極小値に近づく．このように x がどちらの場合でも更新後 $f(x)$ は極小値に近づく．また，傾きは極小値に近づくほどゆるやかになるため，修正量も小さくなり，η が適切な値の場合は，振動をせずに収束に向かう．このように最急降下法では，着実に極小値には近づくが，関数の形状や初期値によっては，最小値を求めることができない場合がある．

2.3　実際に最急降下法で解いてみる

初期値 3.0，学習係数 0.1 として，最急降下法で式 (1) の関数の最小値を求めてみよう．式 (1) を微分すると式 (2) となるので，更新式は以下のようになる．

$$\begin{aligned} x_{t+1} &= x_t - \eta \frac{\partial f(x_t)}{\partial x_t} \\ &= x_t - 0.1 \times 2(x_t - 2) \end{aligned}$$

更新式が求まったので，$x_0 = 3.0$ として，x_t を繰り返し求めてみよう．

$$\begin{aligned} x_0 &= 3.0 \\ x_1 &= 3.0 - 0.1 \times 2(3.0 - 2) \\ &= 3.0 - 0.2 \\ &= 2.8 \\ x_2 &= 2.8 - 0.1 \times 2(2.8 - 2) \\ &= 2.8 - 0.16 \\ &= 2.64 \\ x_3 &= 2.64 - 0.1 \times 2(2.64 - 2) \\ &= 2.64 - 0.128 \\ &= 2.512 \end{aligned}$$

このように，繰り返していくと，徐々に最小値をとるパラメータ $x = 2$ に近づいていく．