ディープラーニングの仕組みを知ろう！

第 2 回　人工知能勉強会

Shion MORISHITA

June 25, 2024

目次

はじめに

勾配降下法

- 勾配降下法の基本概念
- 勾配降下法の式

ニューラルネットワークと勾配降下法

- ニューラルネットワークのパラメータと変数

はじめに

目的

-

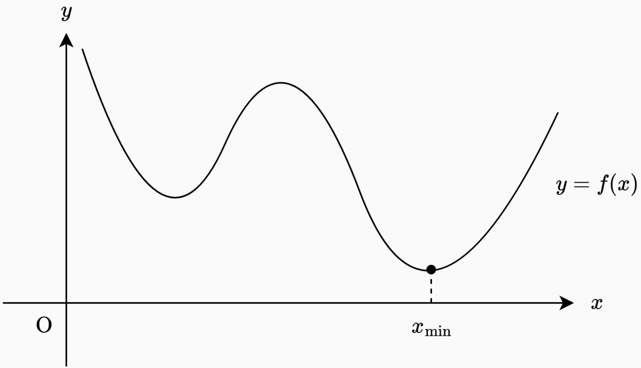
勾配降下法

勾配降下法

勾配降下法の基本概念

勾配降下法とその目的

- 機械学習や最適化の分野で広く用いられる最適化アルゴリズム
- 目的：最小化（または最大化）したい関数の最適なパラメータを見つけること



勾配降下法のアイデア

どのように関数が最小となるパラメータを見つけるか？

- 【重要】多変数関数の最小条件 を利用（第 1 回）
- 斜面を転がるボールのイメージ

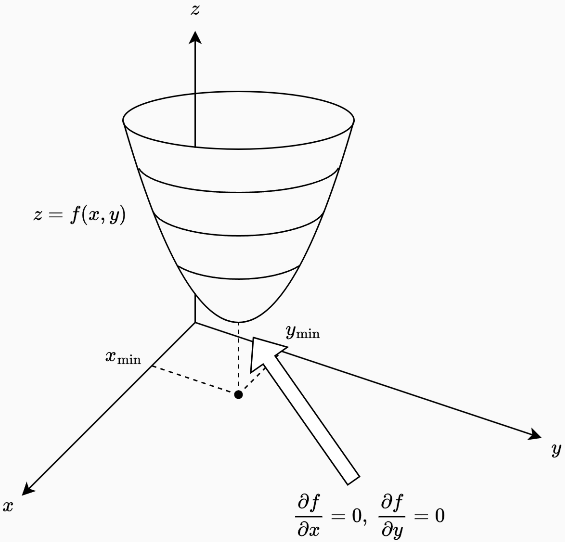
【重要】多変数関数の最小条件（第 1 回）

関数 $z = f(x, y)$ が最小になる必要条件は、$\dfrac{\partial f}{\partial x} = 0$ かつ $\dfrac{\partial f}{\partial y} = 0$

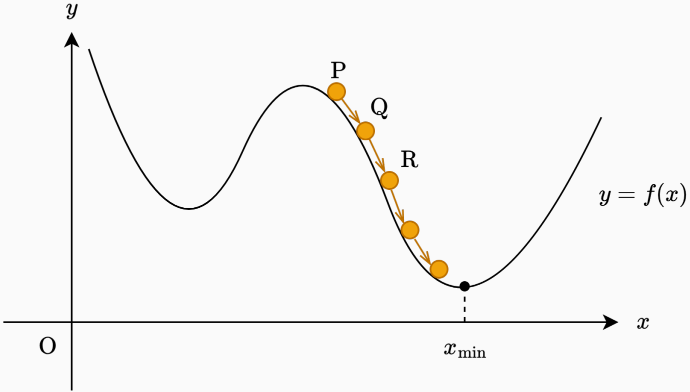
ポイント

どの成分から見ても傾きが 0 なら、最小値の可能性あり！

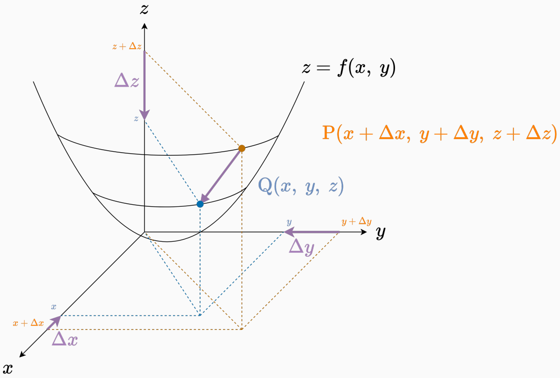
多変数関数の最小条件のイメージ



斜面を転がるボールのイメージ



斜面を転がるボール（多変数関数 ver.）



- 最速で転がる（Δz が最小になる）には $\Delta x,\ \Delta y$ をどう決める？

勾配降下法

勾配降下法の式

【重要】勾配降下法の基本式（2 変数関数）

η を正の小さな定数として、変数 $x,\ y$ が $x+\Delta x,\ y+\Delta y$ に変化するとき、関数 $z=f(x,y)$ が最も減少するのは次の関係を満たすときである：

$$\begin{bmatrix}\Delta x\\ \Delta y\end{bmatrix} = -\eta \begin{bmatrix}\dfrac{\partial z}{\partial x}\\ \dfrac{\partial z}{\partial y}\end{bmatrix}.$$

勾配降下法の基本式の導出 i

「関数の近似公式　簡潔 ver.」（第 1 回）より、

$$\Delta z \simeq \left\langle \begin{bmatrix} \dfrac{\partial z}{\partial x} \\ \dfrac{\partial z}{\partial y} \end{bmatrix}, \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \right\rangle .$$

「ベクトルの基本公式」（第１回）の内積の式

$$\langle \boldsymbol{a}, \boldsymbol{b} \rangle \triangleq \|\boldsymbol{a}\|\|\boldsymbol{b}\| \cos\theta$$

および、コーシー・シュワルツの不等式

$$-\|\boldsymbol{a}\|\|\boldsymbol{b}\| \leq \langle \boldsymbol{a}, \boldsymbol{b} \rangle \leq \|\boldsymbol{a}\|\|\boldsymbol{b}\|$$

勾配降下法の基本式の導出 ii

より、内積が最小となるのは $\cos\theta = -1$ のとき、すなわち、ベクトルの向きが反対のとき（$\theta = 180°$）。

ベクトルの向きが反対というのは、ベクトルの符号が異なるという意味なので、

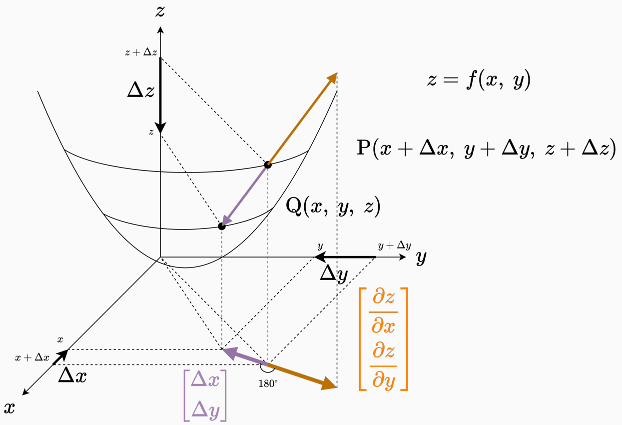
$$\boldsymbol{a} = -k\boldsymbol{b} \quad (k\text{：正の定数})$$

と表せる。今回の表記に合わせれば、

$$\begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = -\eta \begin{bmatrix} \dfrac{\partial z}{\partial x} \\[2ex] \dfrac{\partial z}{\partial y} \end{bmatrix} \quad (\eta\text{：正の定数})$$

と導かれる。 □

結局どういうこと??



<u>ポイント</u>　$\Delta x,\ \Delta y$ の値を、偏微分の値で決定できる！

【重要】勾配降下法の基本式（n 変数）

η を正の小さな定数として、変数 $x_1,\ x_2,\ \ldots,\ x_n$ が $x_1+\Delta x_1,\ x_2+\Delta x_2,\ \ldots,\ x_n+\Delta x_n$ に変化するとき、関数 f が最も減少するのは次の関係を満たすときである：

$$\begin{bmatrix} \Delta x_1 \\ \vdots \\ \Delta x_n \end{bmatrix} = -\eta \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix}.$$

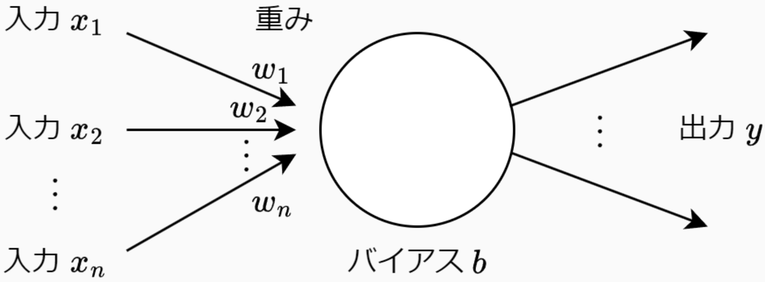
この式に従って点 $(x_1,\ \ldots,\ x_n)$ を次々と移動させることで、関数 f が最小になるパラメータを探索する方法を**勾配降下法**という。

ニューラルネットワークと勾配降下法

ニューラルネットワークと勾配降下法

ニューラルネットワークのパラメータと変数

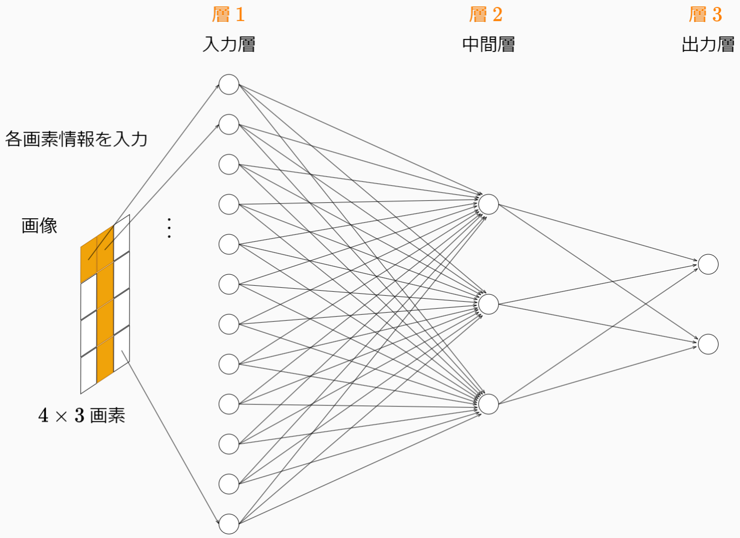
ユニットの復習



重み付き入力：$z = w_1x_1 + w_2x_2 \cdots + w_nx_n + b$

出力：$y = \sigma(z)$

層に番号付けする



変数名・パラメータ名

- x_i
  - 入力層（層 1）にある i 番目のユニットの入力を表す変数。入力層では、出力と入力は同一値なので、出力の変数にもなる。また、該当するユニットの名称としても利用。
- w^l_{ji}
  - 層 $l-1$ の i 番目のユニットから層 l の j 番目のユニットに向けられた矢の重み。i と j の順序に注意。ニューラルネットワークを定めるパラメータ。
- z^l_j
  - 層 l の j 番目にあるユニットが処理する重み付き入力を表す変数。
- b^l_j
  - 層 l の j 番目にあるユニットのバイアス。ニューラルネットワークを定めるパラメータ。
- a^l_j
  - 層 l の j 番目にあるユニットの出力変数。また、そのユニットの名称としても利用。

変数名・パラメータの図解

（図で示す）

【参考】ギリシャ文字一覧

文字	名称	文字	名称
α	アルファ	ν	ニュー
β	ベータ	ξ	グザイ
γ	ガンマ	o	オミクロン
δ	デルタ	π	パイ
ϵ	イプシロン	ρ	ロー
ζ	ゼータ	σ	シグマ
η	イータ	τ	タウ
θ	シータ	υ	ウプシロン
ι	イオタ	ϕ	ファイ
κ	カッパ	χ	カイ
λ	ラムダ	ψ	プサイ
μ	ミュー	ω	オメガ