

最急降下法 (Gradient descent)

定義 1 : 最急降下法 (Gradient descent) ある関数

$$f: \mathbb{R}^n \rightarrow \mathbb{R} \quad (n \in \mathbb{N})$$

が \mathbb{R}^n 上で微分可能であるとする。最急降下法はこの f の極小値 (極大値) を探索するアルゴリズムである。

以下のような手順で実行される。

1. 探索における初期値 $\mathbf{x}^0 \in \mathbb{R}^n$ を定める
2. 学習に依存する正の実数によるパラメータ α を定める
3. k 回目の解による $f(\mathbf{x}^k)$ の勾配 $\nabla f(\mathbf{x}^k)$ が $\mathbf{0}$ となれば終了する (実際にはある程度小さい値で終了する)
4. k 回目の解 \mathbf{x}^k から $k+1$ 回目の解を

$$\mathbf{x}^{k+1} = \mathbf{x}^k - \alpha \nabla f(\mathbf{x}^k)$$

によって得る

5. item 3 へ戻る

最急降下法の正当性. 対象とする関数を f とし、 k 回目の解を \mathbf{x}^k とする。このとき、 f を 1 次の項までテイラー展開すると

$$f(\mathbf{x} + \mathbf{d}) = f(\mathbf{x}) + (\mathbf{d} \cdot \nabla) f(\mathbf{x}) + o(\|\mathbf{d}\|) \quad (0.1)$$

となる。ここで式 (0.1) を

$$\begin{aligned} f(\mathbf{x} - \alpha \mathbf{d}) &= f(\mathbf{x}) - \alpha (\mathbf{d} \cdot \nabla) f(\mathbf{x}) + o(\|\alpha \mathbf{d}\|) \\ &= f(\mathbf{x}) - \alpha \mathbf{d} \cdot \nabla f(\mathbf{x}) + \alpha^2 o(\|\mathbf{d}\|) \end{aligned}$$

と書くと

$$f(\mathbf{x} - \alpha \mathbf{d}) - f(\mathbf{x}) = -\alpha \mathbf{d} \cdot \nabla f(\mathbf{x}) + \alpha^2 o(\|\mathbf{d}\|)$$

が成立する。この $f(\mathbf{x} - \alpha \mathbf{d}) - f(\mathbf{x})$ が負であるためには

$$-\alpha \mathbf{d} \cdot \nabla f(\mathbf{x}) + \alpha^2 o(\|\mathbf{d}\|) \leq 0$$

である必要がある。ここで $\mathbf{d} \cdot \nabla f(\mathbf{x}) \geq 0$ とすると

$$-\alpha \mathbf{d} \cdot \nabla f(\mathbf{x}) + \alpha^2 o(\|\mathbf{d}\|) \leq 0 \Rightarrow \alpha \mathbf{d} \cdot \nabla f(\mathbf{x}) \geq \alpha^2 o(\|\mathbf{d}\|)$$

となる。 $b = o(\|\mathbf{d}\|)$ としたとき

$$\begin{aligned} \alpha \mathbf{d} \cdot \nabla f(\mathbf{x}) &\geq \alpha^2 |b| \\ \Rightarrow \mathbf{d} \cdot \nabla f(\mathbf{x}) &\geq \alpha |b| \\ \Rightarrow \alpha &\leq \frac{\mathbf{d} \cdot \nabla f(\mathbf{x})}{|b|} \end{aligned}$$

となるため、 $\mathbf{d} \cdot \nabla f(\mathbf{x}) \geq 0$ であるとき $f(\mathbf{x} - \alpha \mathbf{d}) - f(\mathbf{x}) \leq 0$ を成立させる α は必ず存在する。

最急降下法では $\mathbf{d} = \nabla f(\mathbf{x})$ であるため

$$\begin{aligned} \mathbf{d} \cdot \nabla f(\mathbf{x}) &= \nabla f(\mathbf{x}) \cdot \nabla f(\mathbf{x}) \\ \mathbf{d} \cdot \nabla f(\mathbf{x}) &= \|\nabla f(\mathbf{x})\|^2 \end{aligned}$$

となり $\mathbf{d} \cdot \nabla f(\mathbf{x}) = \|\nabla f(\mathbf{x})\|^2 \geq 0$ が成立する。したがってある α が存在し、 $f(\mathbf{x} - \alpha \nabla f(\mathbf{x})) - f(\mathbf{x}) \leq 0$ が成立する。

よって適切な α を定めれば繰り返し計算するごとに f の値は小さくなる。

□