

# BengioらのNesterov Accelerated Gradient

---

- Notation

- パラメータ  $\theta$
- モーメントム  $v$
- 損失  $L$
- 学習率  $\eta$
- モーメントム係数  $\alpha \in [0, 1)$
- 暫定的な1ステップ先のパラメータ  $\Theta$

- スライド p.10 の更新式

$$\theta_{t+1} = \theta_t + v_{t+1} \quad (1)$$

を $\Theta$ を使って書き直すことを考える

- ある時点  $t$  におけるパラメータ  $\theta_t$  がモーメントム  $v_t$  の方向にそのまま進んだ結果を

$$\Theta_t = \theta_t + \alpha v_t \quad (2)$$

とおく.

- このとき,

$$\theta_t = \Theta_t - \alpha v_t \quad (3)$$

となる

- ある時点  $t + 1$  のケースも同様に  $\Theta_{t+1} = \theta_{t+1} + \alpha v_{t+1}$  とかけることから,

$$\theta_{t+1} = \Theta_{t+1} - \alpha v_{t+1} \quad (4)$$

となる

- よって, 式(1)に式(3)と式(4)を代入すると

$$\Theta_{t+1} - \alpha v_{t+1} = (\Theta_t - \alpha v_t) + v_{t+1}$$

$$\Theta_{t+1} = \Theta_t - \alpha v_t + \alpha v_{t+1} + v_{t+1} \quad (5)$$

$$\Theta_{t+1} = \Theta_t - \alpha v_t + (\alpha + 1)v_{t+1} \quad (6)$$

を得る. 式(5)はスライドp.12の  $\Theta_{t+1}$  の式1行目, 式(6)はスライドp.12の  $\Theta_{t+1}$  の式2行目にそれぞれ対応する.

- このとき, スライドp.10のモーメントムの更新式

$$v_{t+1} = \alpha v_t - \eta \frac{\partial L}{\partial (\theta_t + \alpha v_t)}$$

は,

$$v_{t+1} = \alpha v_t - \eta \frac{\partial L}{\partial \Theta_t} \quad (7)$$

と書くことができる.

- 式(7)を式(6)に代入し, 変形を行うと

$$\Theta_{t+1} = \Theta_t - \alpha v_t + (\alpha + 1) \left( \alpha v_t - \eta \frac{\partial L}{\partial \Theta_t} \right) \quad (8)$$

$$= \Theta_t - \alpha v_t + \alpha(\alpha + 1)v_t - (\alpha + 1) \left( \eta \frac{\partial L}{\partial \Theta_t} \right)$$

$$= \Theta_t - \alpha v_t + (\alpha^2 + \alpha)v_t - (\alpha + 1) \left( \eta \frac{\partial L}{\partial \Theta_t} \right)$$

$$= \Theta_t - \alpha v_t + \alpha^2 v_t + \alpha v_t - (\alpha + 1) \left( \eta \frac{\partial L}{\partial \Theta_t} \right)$$

$$= \Theta_t + \alpha^2 v_t - (\alpha + 1) \left( \eta \frac{\partial L}{\partial \Theta_t} \right) \quad (9)$$

$$= \Theta_t + \alpha^2 v_t - \left\{ \alpha \left( \eta \frac{\partial L}{\partial \Theta_t} \right) + \left( \eta \frac{\partial L}{\partial \Theta_t} \right) \right\}$$

$$= \Theta_t + \alpha^2 v_t - \alpha \left( \eta \frac{\partial L}{\partial \Theta_t} \right) - \left( \eta \frac{\partial L}{\partial \Theta_t} \right)$$

$$= \Theta_t + \alpha \left( \alpha v_t - \eta \frac{\partial L}{\partial \Theta_t} \right) - \eta \frac{\partial L}{\partial \Theta_t} \quad (10)$$

となる。ここで、式(8)はスライドp.12の  $\Theta_{t+1}$  の式3行目、式(9)はスライドp.12の  $\Theta_{t+1}$  の式4行目、式(10)はスライドp.12の  $\Theta_{t+1}$  の式5行目にそれぞれ対応する。