

# Momentum

momentum采用上一步的惯性减少陷入local minimum和saddle point的概率。

$$\text{Movement } \mathbf{m}^2 = \lambda \mathbf{m}^1 - \eta \mathbf{g}^1$$

$$\text{Move to } \boldsymbol{\theta}^2 = \boldsymbol{\theta}^1 + \mathbf{m}^2$$

m为上一步的移动方向，g为模型计算出的梯度。