

6.1 パラメータの更新

■NNの学習の目的

を最小化するパラメータを見つけること

⇒

NNの最適化は難題である！

理由 1 :

理由 2 :

1

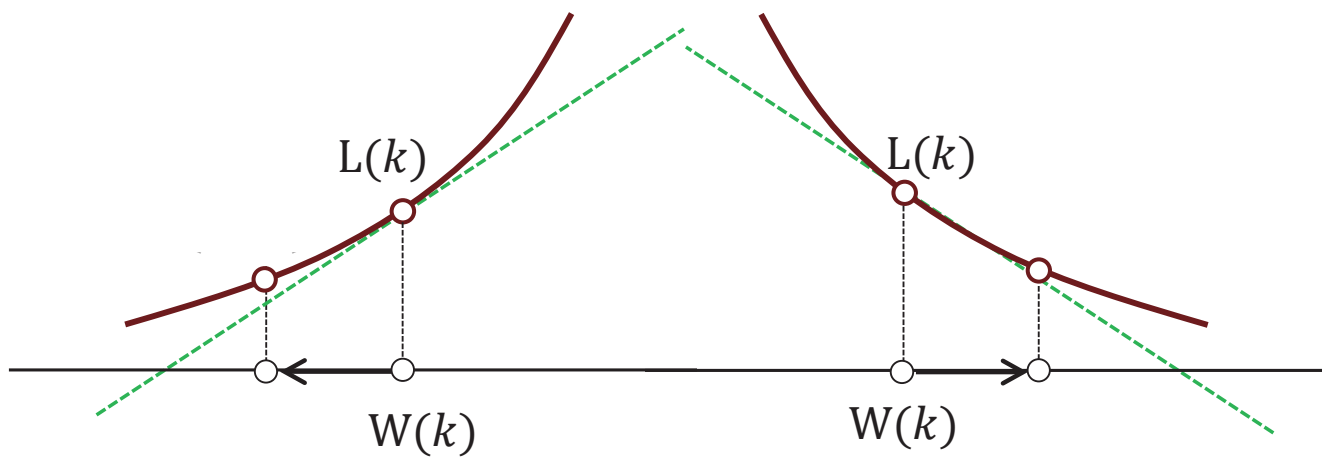
6.1.2 SGD

SGD : 確率的 勾配 降下

⇒パラメータ曲面の傾斜が へと進むことで
パラメータをアップデートする

2

6.1.2 SGD



$$= -\eta \quad \eta > 0$$

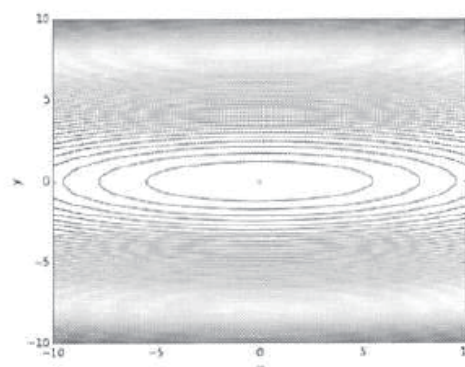
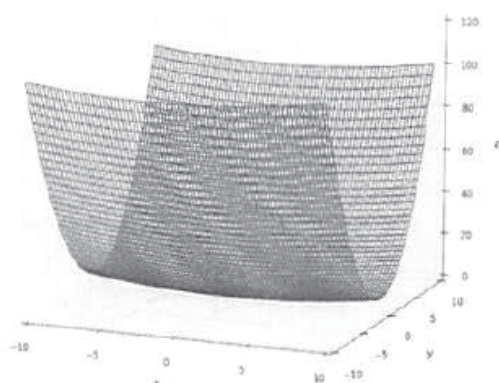
3

6.1.3 SGDの欠点

SGDの良い点：

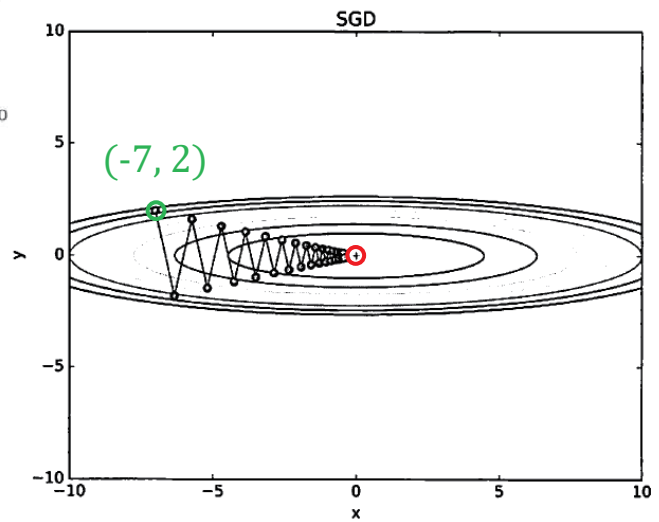
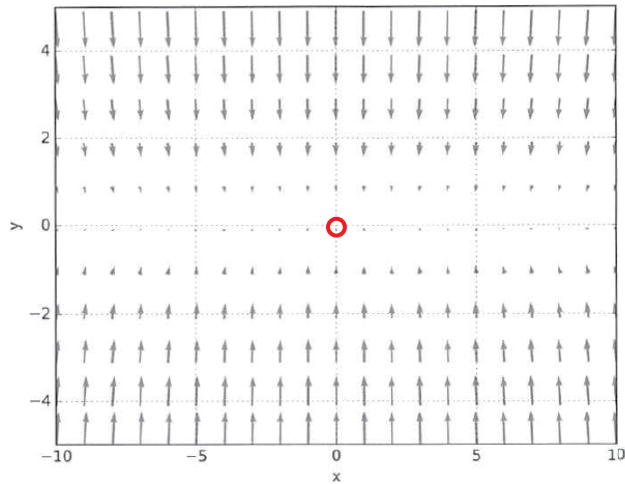
悪い点：

$$f(x, y) = \quad +$$



4

6.1.3 SGDの欠点



5

6.1.3 SGDの欠点

■ SGDが非効率的な経路探索法である理由

(1)

(2)

■ SGDの改良

(1)

の導入

(2)

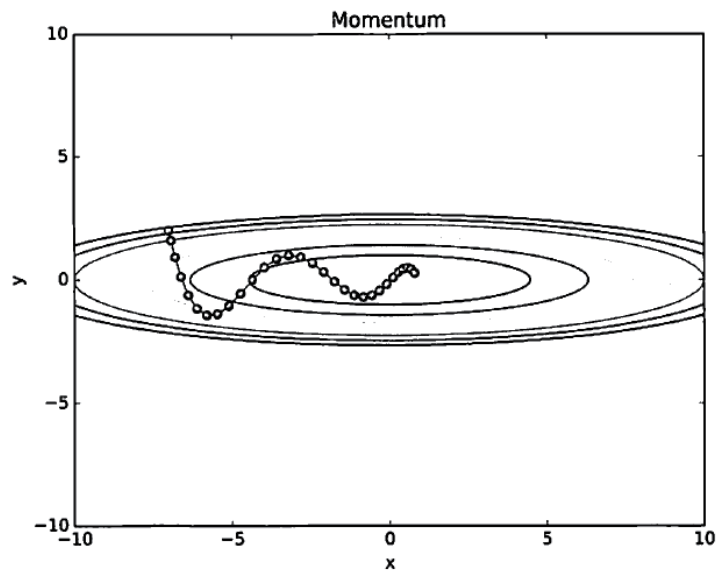
(3)

6

6.1.4 Momentum

$$= -\eta$$

$$= +$$



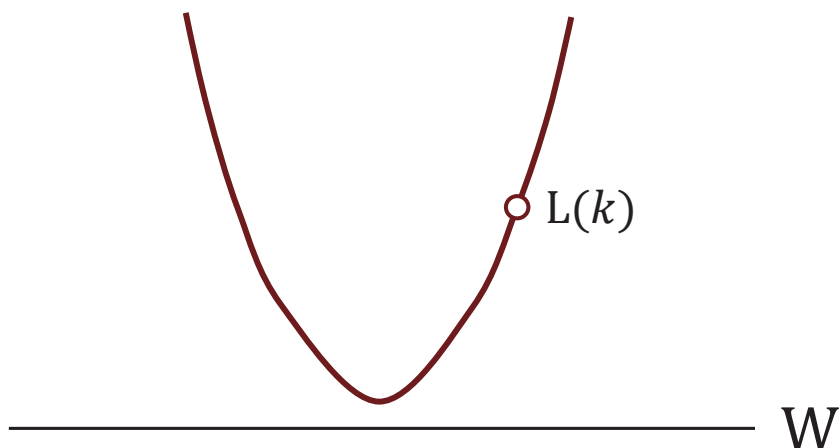
7

6.1.5 AdaGrad

学習係数 η :

小さすぎる→

大きすぎる→パラメータのアップデート幅が大きすぎて
にたどりつけない



6.1.5 AdaGrad : Adaptive Subgradient

■ 対策 学習係数を

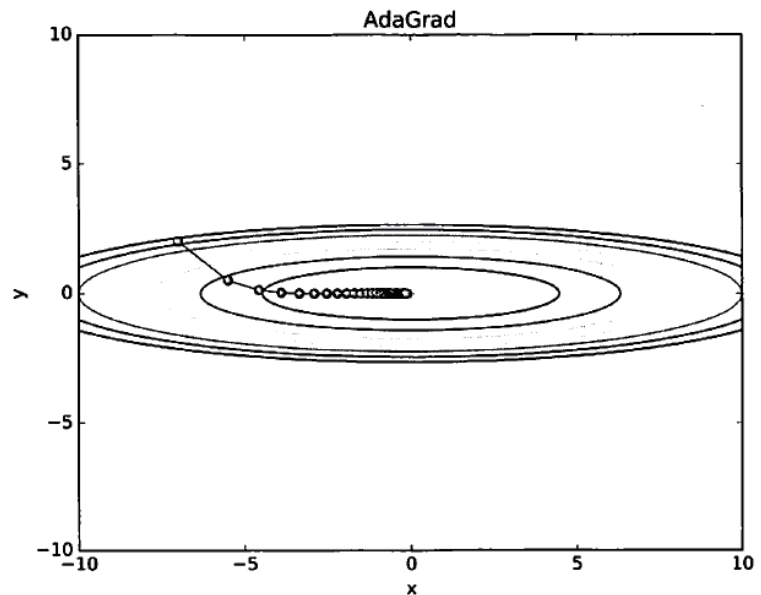
小さくする

AdaGrad → 個々のパラメータに対して

を調整する

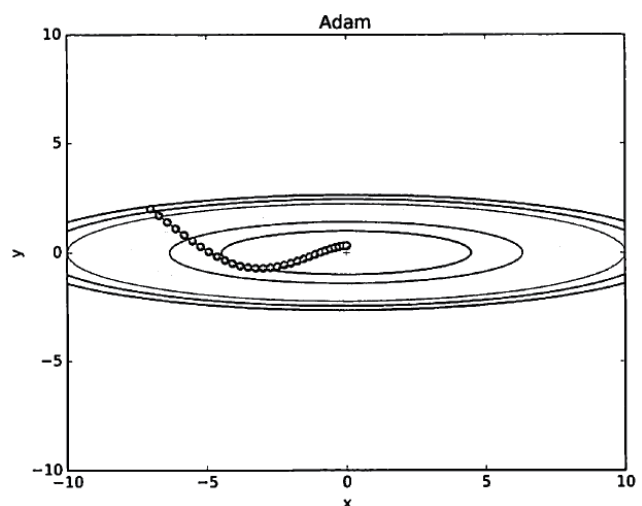
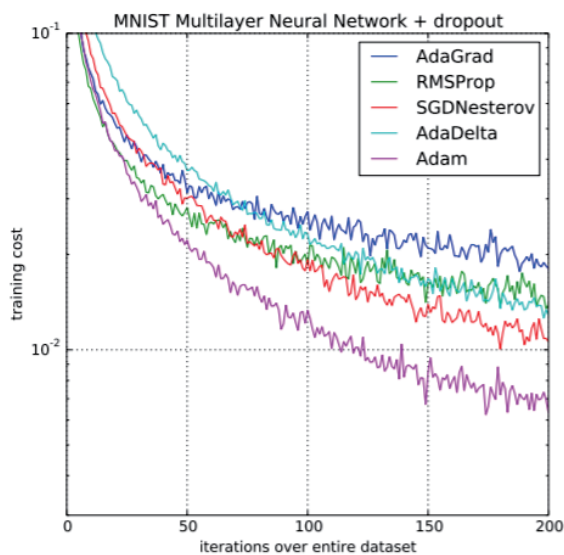
$$\mathbf{h} \leftarrow \mathbf{h} + \odot$$

$$\mathbf{W} \leftarrow \mathbf{W} - \eta$$



9

6.1.6 Adam (2015)



10

6.2.1 重みの初期値を 0 にする？

(1) 初期値はできる限り

理由： を防ぐため

(2) 初期値は が良い

理由： の値で重みが更新されてしまう

→重みが 値になる

→学習できなくなる

11

6.4.1 過学習

■過学習とは？

訓練データのみに 結果, 以外の

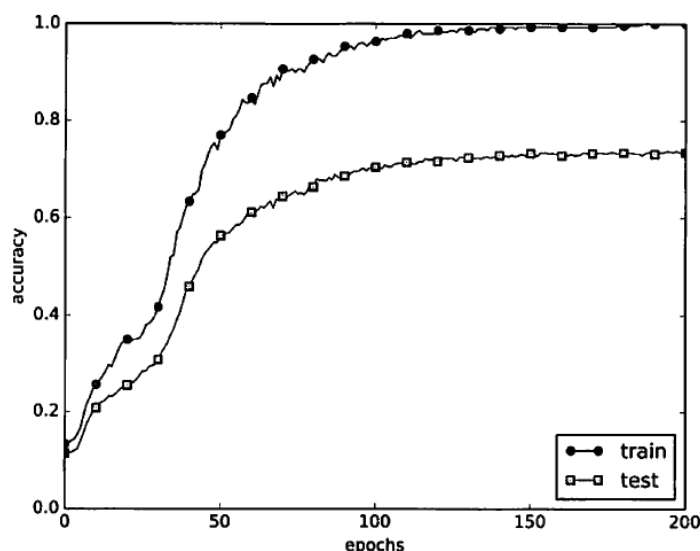
データに対して を得ることができない学習状態

主原因) 訓練データが少なく を表せていない



使用訓練データ300

MNIST訓練データ60,000



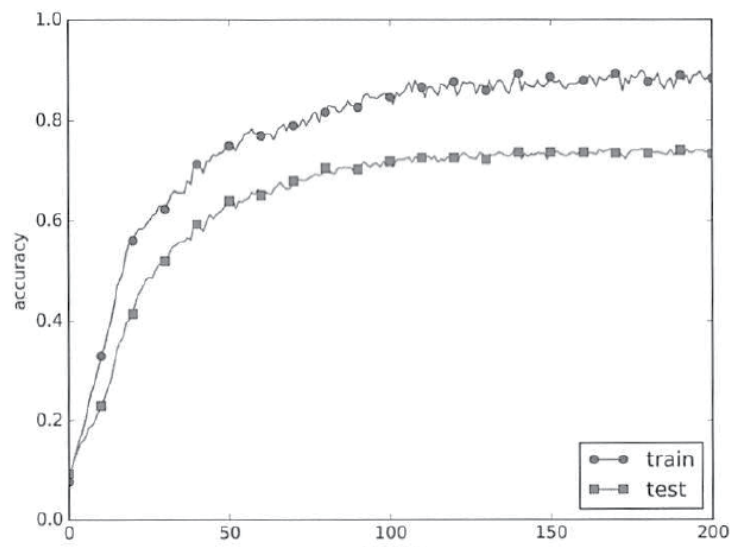
12

6.4.2 Weight decay

重みが ∞ になることでも過学習が発生する！

$$\text{損失関数: } L \rightarrow L + \frac{1}{2}\lambda \mathbf{W}^2 \quad \text{変化分: } \frac{\partial L}{\partial \mathbf{W}} \rightarrow \frac{\partial L}{\partial \mathbf{W}} + \lambda \mathbf{W}$$

使用訓練データ300
MNIST訓練データ60,000



13

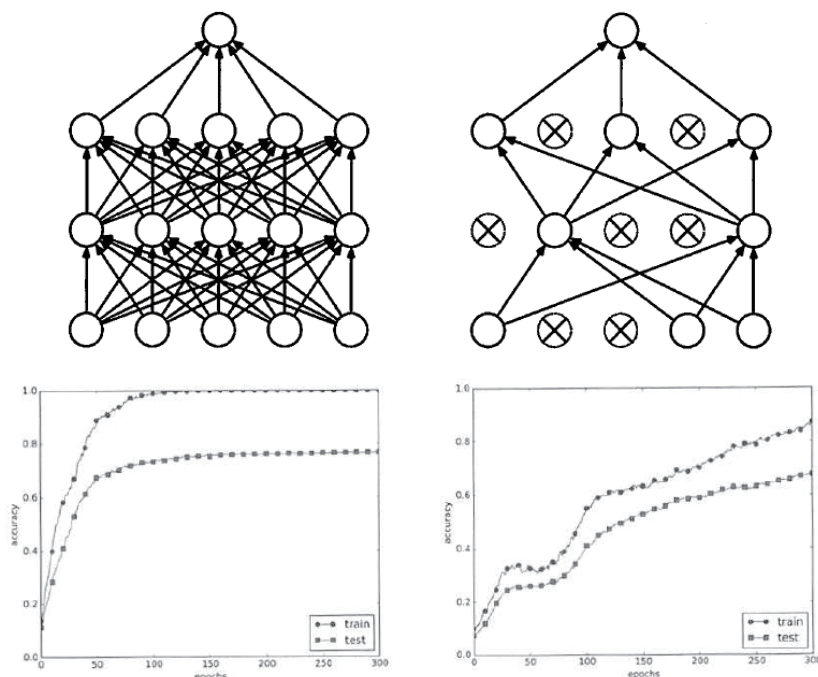
6.4.3 Dropout

NNが複雑になると

では対応は困難

→訓練時に

Dropoutが必要



14