

## 宿題 4

以下の深層学習に関する技術用語について説明する。これらを講義で説明した多層ニューラルネットワークに組み込む方法についても検討する。

### Dropout

Dropout は、訓練時にニューロンのうち一部を確率的に不活性化させて学習させることにより、汎化性能を上げる手法である。つまり、Dropout を適用する層において、重みのうち一部を確率的に選び、選ばれたものの重みを 0 として順伝搬を行い、また逆伝搬時にも重みの更新をしないようにする。ここで、Dropout する確率はハイパーパラメータとなる。推論時には、全てのノードを用いて計算を行った後、訓練時に Dropout していた割合を出力にかけることによって、用いているノードの数が訓練時より増加し大きくなった出力の値を修正している。このように学習することによって、複数のネットワークを独立に学習し、推論時にその出力を平均する、アンサンブル学習の近似になるといわれている。

これを多層ニューラルネットワークに組み込もうとするときは、単純に層の後にこれを入れればよい。なお、どのノードを不活性化したかを記憶しておく必要があることに注意である。

### momentum

パラメータの更新は SGD 以外の方法によっても行うことができる。momentum はその手法の 1 つで、直訳すると運動量である。momentum によるパラメータの更新式は以下の式で表される（ $\alpha$  はハイパーパラメータ）。

$$\mathbf{v} \leftarrow \alpha \mathbf{v} - \eta \frac{\partial E}{\partial \boldsymbol{\theta}} \quad (1)$$

$$\boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \mathbf{v} \quad (2)$$

式からわかるように、momentum では更新に用いる値を計算された勾配の値そのものからその指数平均へと変更している。これにより、振動が抑制され、損失関数が極小値に向かいやすくなるという手法である。パラメータが張る空間上での損失関数の値の動きを考えると、これは慣性が働いているかのような動きとなる。

これを多層ニューラルネットワークに組み込もうとするときは、パラメータの更新式を SGD からこれに取り換えればよい。 $\mathbf{v}$  の値を記憶しておかなければならないことに注意である。

### data augmentation

data augmentation とは、データを増やす手法のことである。これは、新規にデータを集めるということではなく、現在手元にあるデータに対し何かしらの処理を加えることでデータを増やすということである。この手法は当然データの種類によって様々であるが、例えば画像について考えると、次のようなものが挙げられる

- 上下・左右の反転
- 回転
- ノイズの付加
- 色の反転
- Random Crop（画像の一部を切り抜く）

- CutOut（画像の一部にマスクをかける）

当然，これらの手法は状況に応じて使い分けられるべきである。例えば，手書き文字認識をしたいときには，左右の反転は用いられないだろう。

これは前処理にあたる部分なので，多層ニューラルネットワークに組み込もうとするときは，その入力に対する操作となる。例えば MNIST に対しては，ノイズの付加や多少の回転は適用してみてもよい手法であると考えられる。