Let’s study ‘Deep Learning’

まず、はじめに…

|  |
| --- |
| 入力データ：、出力データ：、パラメータ:(任意の行列) |



-図１-

という流れで、出力データの予測をして来た。

これが、一層の深層学習である。

ちなみに、今、黄色で囲まれた部分は、確率の範疇で扱いたいため、全ての分布関数を足して１になる必要がある。そのため、全体の和で各、確率分布を割っている。

この関数は、ソフトマックス関数と言われている。

|  |
| --- |
| ソフトマックス関数 |

次に、１層以上深層学習について説明する。

この解釈はとても簡単で、上の図において、□で囲まれた箇所を繰り返した数だけ、層数が増える。

例えば、次の例だと、



-図2-

3層の深層学習である。

より詳しく書くと、次のようになる。



-図3-

ここで、注目しなければいけないことは、**赤い□**で囲まれた、の個数である。

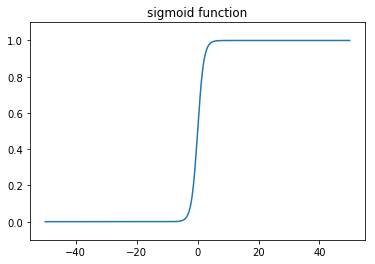
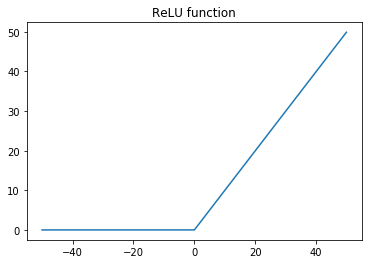
例えば、手書き文字の0~9を予測したいとたら、の個数は、10(0~9の個数)個でなければいけない。

なぜなら、0~9の確率が最も高いものを出力としなければいけないからである。

しかし、,の個数は決まっていない。その理由は、数学的にいくつの出力数が最適かまだ証明されていないからである。

また、**赤い□**で囲まれた部分は、確率を出力しなければいけなかった。そのため、確率分布として扱うため、各、確率分布を足して1にする必要があった。しかし、他の**黄色の□**の部分は確率を出力しなくても良い。

深層学習では、のような隠れ層には、Relu関数がよく多用される。



-図4-

その理由は、深層学習ではパラメータ更新する（学習）させる上で、微分を多用するからである。

上図から分かるように、’sigmoid function’では、ほとんどの場合、微分値が０となってしまう。しかし、‘Relu function’では、０より大きいところでは、必ず微分値が存在する。そのため、Relu関数がよく多用される。

さらに、の出力値を‘Laplace’で正規化することを’Batch normalization’という。

これは、「出力値は極端な外れ値はない（1次モーメント＝1、2次モーメント＝0）」という過程を行うことで、学習の速度が上がるなどの効果が得られる。

パラメータにおける微分を見ていく。

ここでは、は、スカラーとして、計算していく。

とする。

図3から、分かるように、である。

同様に、である。

ここで、分かることは、「についての微分にはが使われていること」と、「についての微分にはが使われていることである。

つまり、 「パラメータの更新　→ パラメータの更新　→ パラメータの更新」

というような流れになっている。

つまり、図で見ると、下記のようになる。



連鎖律から分かるように、パラメータの更新（学習）は入力データから出力（予測値）を出す流れと、逆の方向である。

そのため、「誤差逆伝播法」と言われる。