知能システム学特論レポート

（DL2 班）Caffe on Ubuntu

2015 年 6 月 29 日

1　報告者

15344203	有田 裕太
15344206	緒形 裕太
15344209	株丹 亮
12104125	宮本 和

2　進行状況

- 理論研究
- 順伝播型ネットワークについて

3　理論研究

3.1　ユニットの出力

各ユニットは複数の入力を受けて 1 つの出力を計算する．まず Fig. 1 よりこのユニットの総入 u は，各入力に対しそれぞれ異なる重み $w1$，$w2$，$w3$，$w4$ を入力 $x1$，$x2$，$x3$，$x4$ に掛けたものの和にバイアス値 b を

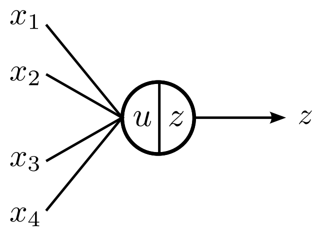

Fig.1　ユニット 1 つの入出力

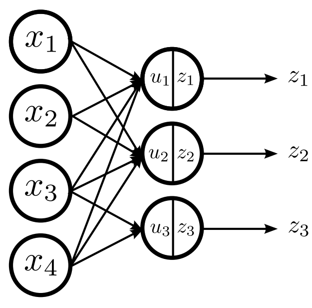

Fig.2　複数のユニットを持つネットワーク

加えたものであり以下の式で表される．

$$u = w_1x_1 + w_2x_2 + w_3x_3 + w_4x_4 + b \tag{3.1}$$

そして，ユニットの出力 z は総入力 u に対する活性関数と呼ばれる関数 f の出力は

$$z = f(u) \tag{3.2}$$

となり，順伝播型ネットワークではこのようなユニットが層状に並べられている．第 1 層のユニットを $i = 1, \cdots, I$，第 2 層のユニットを $j = 1, \cdots, J$ で表すと，次のように一般化できる．

$$u_j = \sum_{i=1}^{I} w_{ji}x_i + b_j \tag{3.3}$$

$$z_j = f(u_j) \tag{3.4}$$

3.2 活性化関数

3.2.1 シグモイド関数

入力の絶対値が大きな値をとると，出力が飽和し一定値になり，その間の入力に対して出力が徐々にかつ滑らかに変化する関数.

3.2.2 正規化線形関数

シグモイド関数は入力の変動が大き過ぎると出力が 0 か 1 の値しかとれないが，正規化線形関数はバイアスがない場合，入力が 0 以上であれば出力が入力に比例する．

3.2.3 マックスアウト

K 個 1 つ 1 つが異なる重みとバイアスを持ち，それぞれの総入力を $u_{j1}...u_{jk}$ と別々に計算したあと，それぞれの最大値をユニットの出力とする．

$$u_{jk} = \sum w_{jik}z_i + b_{jk} \quad (1...K) \tag{3.5}$$

$$f(u_j) = \max_{k=1...K} u_{jk} \tag{3.6}$$

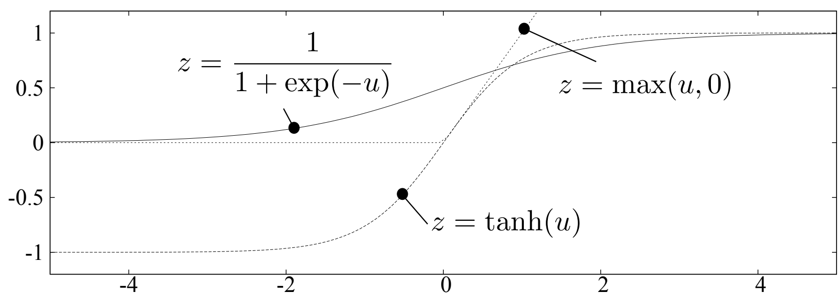

Fig.3　典型的な活性化関数とその概形

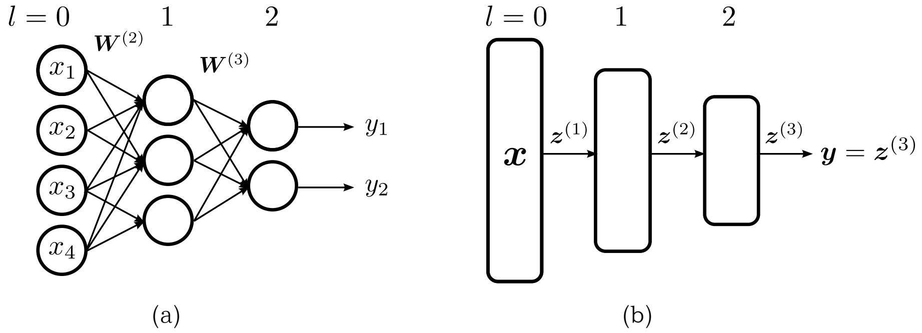

Fig.4　2 層のネットワーク

3.3　多層ネットワーク

Fig. 4 に 2 層構造のネットワークを示す．Fig. 4 (a) より各層を $l=0,\ 1,\ 2$ とすると，$l=1$ の層を入力層，$l=2$ を中間層，隠れ層，$l=3$ を出力層と呼ぶ．各層のユニットの入出力を区別するために，入力を $\boldsymbol{u}^{(l)}$，出力を $\boldsymbol{z}^{(l)}$ と定義すると，中間層（$l=2$）のユニットの出力は以下の式で表される．

$$\boldsymbol{u}^{(2)}=\boldsymbol{W}^{(2)}\boldsymbol{x}+\boldsymbol{b}^{(2)} \tag{3.7}$$

$$\boldsymbol{z}^{(2)}=\boldsymbol{f}(\boldsymbol{u}^{(2)}) \tag{3.8}$$

$\boldsymbol{W}^{(2)}$ は入力層と中間層の結合重みであり，$\boldsymbol{b}^{(2)}$ は中間層のユニットに与えられたバイアスである．同様にして $\boldsymbol{u}^{(3)}$，$\boldsymbol{z}^{(3)}$ は

$$\boldsymbol{u}^{(3)}=\boldsymbol{W}^{(3)}\boldsymbol{z}^{(2)}+\boldsymbol{b}^{(3)} \tag{3.9}$$

$$\boldsymbol{z}^{(3)}=\boldsymbol{f}(\boldsymbol{u}^{(3)}) \tag{3.10}$$

となり，任意の階層 L のネットワークに一般化すると

$$\boldsymbol{u}^{(l+1)}=\boldsymbol{W}^{(l+1)}\boldsymbol{z}^{(l)}+\boldsymbol{b}^{(l+1)} \tag{3.11}$$

$$\boldsymbol{z}^{(l+1)}=\boldsymbol{f}(\boldsymbol{u}^{(l+1)}) \tag{3.12}$$

と書ける．$l=1,\ 2,\ 3,\cdots,L-1$ の順に繰り返していくと最終的な出力 $\boldsymbol{y}$ を決定することができる．この出力を決定するのは各層間の結合重み $\boldsymbol{W}^{(l)}$ $(l=2,\cdots,L)$ とユニットのバイアス $\boldsymbol{b}^{(l)}$ $(l=2,\cdots,L)$ である．これらのパラメータを持つベクトル $\boldsymbol{w}$ を定義して，$\boldsymbol{y}(\boldsymbol{x};\boldsymbol{w})$ と表現する．

3.4　出力層の設計と誤差関数

3.4.1　学習の枠組み

順伝播型ネットワークが表現する関数 $\boldsymbol{y}(\boldsymbol{x};\boldsymbol{w})$ をネットワークのパラメータ $\boldsymbol{w}$ を変えることで変化させ，望みの関数を与えることを考える．入力 $\boldsymbol{x}$ と望みの出力 d のペアを次のように与える．

$$\{(\boldsymbol{x}_1,\mathrm{d}_1),(\boldsymbol{x}_1,\mathrm{d}_1),...,(\boldsymbol{x}_N,\mathrm{d}_N)\} \tag{3.13}$$

これらのペア $(\boldsymbol{x}, \mathrm{d})$ 1 つ 1 つを訓練サンプル (training samples) といい，その集合を訓練データ (training data) という．ネットワーク $\boldsymbol{w}$ を調整することで訓練データの入出力ペアをできるだけ再現すること学習という．

この場合，ネットワークが表す関数と訓練データとの近さ $(\boldsymbol{y}(\boldsymbol{x}_n; \boldsymbol{w}))$ を誤差関数 (error function) で定義する．誤差関数は問題の種別や活性化関数によって異なる．Tab.1 に問題の種別ごとの活性化関数と誤差関数の一覧を示す．

Tab.1　問題の種別ごとの活性化関数と誤差関数

問題の種別	出力層の活性化関数	誤差関数
回帰	正接双曲線関数や恒等写像	二乗誤差 式 (3.14)
二値分類	ロジスティック関数	式 (3.15)
多クラス分類	ソフトマックス関数	交差エントロピー 式 (3.17)

3.4.2　回帰

回帰 (regression) とは出力連続値をとる関数を対象に訓練データを良く再現する関数を求めることをいう．回帰では活性化関数に正接双曲線関数や恒等写像を用い，評価関数は次式が良く用いられる．

$$E(\boldsymbol{w}) = \frac{1}{2}\sum_{n=1}^{N} ||\boldsymbol{d}_n - \boldsymbol{y}(\boldsymbol{x}_n; \boldsymbol{w})||^2 \tag{3.14}$$

3.4.3　二値分類

二値分類では入力 $\boldsymbol{x}$ に応じて 2 種類に区別する問題を考える．すなわち，$d \in \{0, 1\}$ とする．このとき，活性化関数はロジスティック関数 $y = 1/(1 + \exp(-u))$ とし，誤差関数は次式で与える．

$$E(\boldsymbol{w}) = -\sum_{n=1}^{N} [d_n \log y(\boldsymbol{x}_n; \boldsymbol{w} + (1 - d_n) \log\{1 - y(\boldsymbol{x}_n; \boldsymbol{w})\})] \tag{3.15}$$

3.4.4　多クラス分類

多クラス分類とは入力 $\boldsymbol{x}$ に応じて有限個のクラスに分類する問題である．一例として Fig.5 に手書き文字認識の例を示す．この問題では活性化関数にはソフトマックス関数 (softmax function) が良く用いられる．出力相 $l = L$ の k 番目 $(k = 1, ..., K)$ のユニットの出力は $l = L - 1$ 層の出力を元に次式で与えられ，これをソフトマックス関数という．

$$y_k \equiv z_k^{(L)} = \frac{\exp(u_k^{(L)})}{\sum_{j=1}^{K} exp(u_j^{(L)})} \tag{3.16}$$

また，誤差関数は次式で与える．

$$E(\boldsymbol{w}) = -\sum_{n=1}^{N}\sum_{k=1}^{N} d_{nk} \log y_k(\boldsymbol{x}_n; \boldsymbol{w}) \tag{3.17}$$

なお，この関数は交差エントロピー (cross entropy) と呼ばれる．

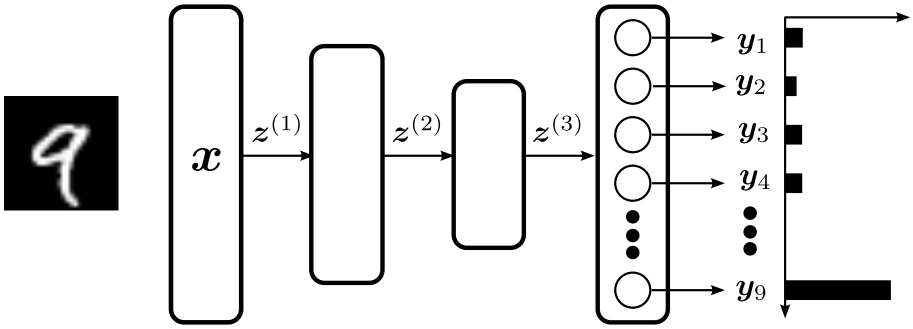

Fig.5　手書き文字認識の例

4　今後の課題

- 理論研究を進める．
- Caffe を使いこなす

参考文献

[1] 岡谷貴之，“機械学習プロフェッショナルシリーズ　深層学習”，講談社，2015.