第 6 回勉強会　レジュメ

照井ゼミ 3 年　藤島大輝

概要

ニューラルネットと呼ばれるプログラムは事例を参照し、自らの構造を変化させて自分で賢くなることができる。その構造は脳の神経回路を数理モデルで表現した形式ニューロンがもとになっている。学習の方法は、用意したネットワークの重みについて教師あり学習をさせるのが基本である。ニューラルネットによる学習も誤差関数の最小化問題を解くことで実行される。その手法の 1 つを勾配降下法という。

1. イントロダクション

・深層学習の研究自体は 1950 年代からブームが 2 回あり、研究が停滞していた冬の時代を 2 回迎えている。現在は第 3 次 AI ブームと位置づけることが可能。

・第 3 次 AI ブームはコンピュータなどの計算機の発達とインターネットの普及によるデータ量の増加による。膨大なデータから注目すべき特徴を見つけ、その特徴量を用いて知識を表現できることが特徴。

Ex1)2012 年の ILSVRC における Super Vision

世界的な画像認識のコンペティション ILSVRC において初参加のトロント大学が開発した Super Vision が他機関を押さえて圧勝。

→人間が特徴量を設計するのではなく、データをもとにコンピュータが自ら特徴量を作り出す

⇒新しい機械学習の方法「ディープラーニング(深層学習)」の登場

Ex2)グーグルの猫

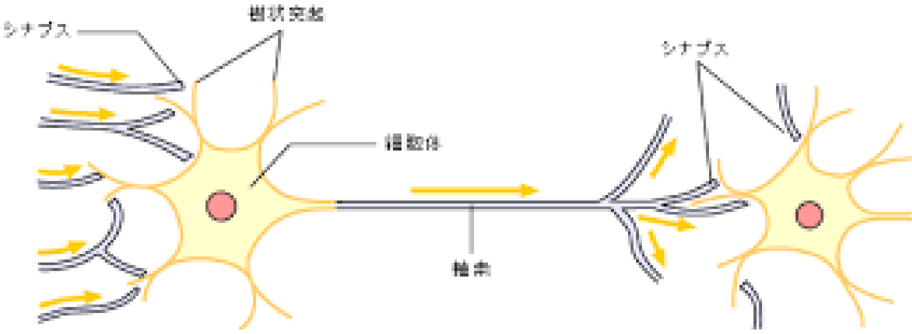
YouTube 動画から取り出した猫の画像で 9 層構造の人口ニューラルネットを教師なし学習

→様々な概念に特異的に反応する人口ニューロン(ユニット)が生まれ、それを画像に再現したもの「グーグルの猫」の発表

⇒ユニットが学び取ったものが特徴量

2. 神経細胞のネットワーク

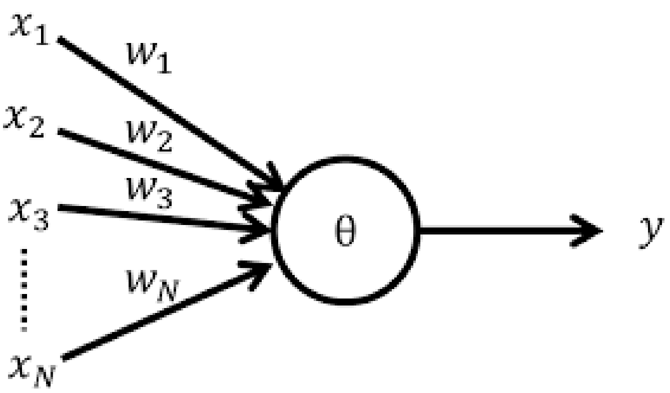
人間の脳は 1000 億個以上もの神経細胞(ニューロン)が寄り集まって構成されている



神経細胞は①細胞体②樹状突起③軸索の 3 つの部分から構成されている
①細胞体･･･演算部
②樹状突起･･･他の多くの神経細胞からの情報を受け取る入力部
③軸索･･･出力部

細胞体はある一定以上の大きさ(閾値)を越える電気信号を受け取ると軸索に向かって電気信号を出力する。電気信号は各軸索を通して他のニューロンの樹状突起に入力し、同じ電波のパターンを繰り返す。

・マッカロとピッツの形式ニューロン・・・ニューロンの活動を数理論理学的な手法によりモデル化



→形式ニューロンも他の多数の形式ニューロン i=1,2,…から入力信号 xi を受け入れる。ニューロンの出す信号はオンかオフの情報しかなく、よって xi の値は 0 か 1 のみ。シナプスごとにニューロン同士の結びつきが異なることが知られているから結合度を表す重み wi を導入して、層への出力層を

$$u = \sum wi * xi$$

と定義する。この入力を受けてニューロンは軸索へ出力を出すが、閾値を、

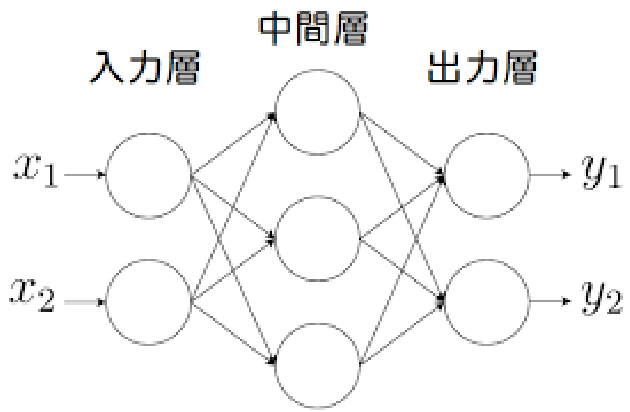
$$\theta(x+b) = \begin{cases} 1 & (x \geq -b) \\ 0 & (x < -b) \end{cases}$$

とモデル化。b は閾値を与えるパラメータ。すると、ニューロンの出力は、

$$y = \theta(u+b) = \theta\left(\sum wi * xi + b\right)$$

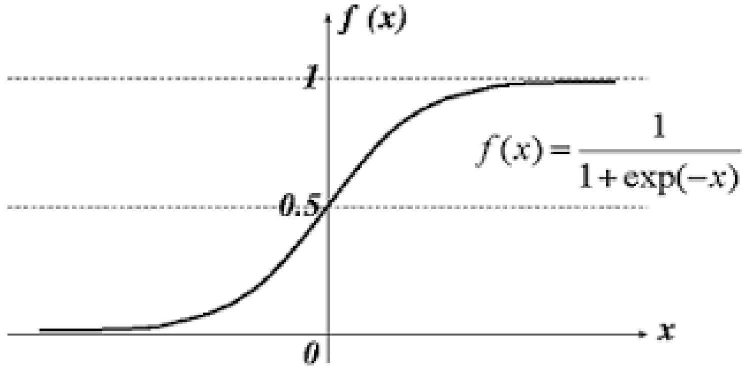
となる。総入力 u を出力 y へ変換する関数を活性化関数と呼ぶ(この場合は特に線形閾値関数、またはヘヴィサイドの階段関数とよばれる。)。

・ローゼンブラットのパーセプトロン･･･形式ニューロンのネットワークの重み w とバイアス b を固定された数ではなく、問題をよく処理できるように訓練させる
→形式ニューロンを複数組み合わせたニューロン回路(パーセプトロン)の作成



・入力層･･･入力値の成分であるベクトル *x1,x2,…*の集まりであるベクトル
x=(*x1,x2,…*)。統計学の立場からは予測変数、独立変数と呼ばれる。
・出力層･･･推定したい最終結果 **y**=(*y1,y2,…*)の各成分を出力値とする。統計学の立場からは目的変数などと呼ばれる。
・中間層(隠れ層)･･･上記以外

・順伝播型ニューラルネットワーク･･･各ユニットは実数値の入出力をもつ(離散値しかとらなかった形式ニューロンとの決定的な違い)。活性化関数として微分可能な増加関数も採用。よく用いられるのはシグモイド関数。



ユニットへの総入力 u は入力 xi にユニットごとの重みを掛け合わせた

$$u = \sum wi * xi$$

となる。ユニットからの出力 z は u にバイアスを加え活性化関数 f での変換を施した

$$z = f(u+b) = f(\sum wi * xi + b)$$

である。

中間層が行う演算処理は、

$$\mathbf{U}(l) = \mathbf{W}(l) * \mathbf{Z}(l-1),\ \mathbf{Z}(l) = f(l)(\mathbf{U}(l) + \mathbf{B}(l))$$

とまとめられる。

出力層が行う演算処理は、

$$\mathbf{Y} = \mathbf{Z}(L) = f(L)(\mathbf{U}(L)),\ \mathbf{U}(L) = \mathbf{W}(L)\mathbf{h} = \mathbf{W}(l)\mathbf{Z}(L-1)$$

とまとめられ、ここで **h** が入力 **X** の表現であり、この表現 **h** の回帰分析を通じて **x** と **y** の関係を推定するのが出力層の役割。

この表現を使ってニューラルネットにさせたいタスクごとに分けて出力層を変える

3. 勾配降下法による学習

ニューラルネットの学習は誤差関数 E(**w**)の最小化として定式化される

→深層学習では損失関数の 1 階微分の情報だけを用いる

⇒勾配降下法という

何らかの初期値 **w**(0)を用意。この初期値から始めて誤差関数の最小点を求める

勾配の逆方向(誤差関数が小さくなる方向)に進める手順は

$$\mathbf{W}(t+1) = \mathbf{W}(t) + \Delta\mathbf{W}(t),\ \Delta\mathbf{W}(t) = -\lambda\Delta E(\mathbf{W}(t))$$

λ は学習率と呼ばれ、観測者が決めるハイパーパラメータ。λ が大きすぎると最適解付近を往復。小さすぎると学習がいっこうに進まない。

→誤差関数は複雑なため、本当の最小値である大局的最小値と初期値の周りの局所的

極小値が存在
⇒局所的最小値・最大値の問題
解決策：確率的勾配法

4. 課題・理解できなかった部分
・一般線形化モデルという分野についての知識が必要
・ロジスティックシグモイド関数は何がうれしいのか
・情報量、またはエントロピーという概念が何を表しているのか分からない
・極値を求める際の変分法という数学的理解の不足

⇒まずは一般化線形モデル(統計モデルの次のステップ？)に進む必要あり(「数理統計」の知識は基本統計レベル)
⇒数学の知識は大学１年程度の理解で十分なのか？表記に慣れていないだけか？

参考図書
「人工知能は人間を越えるか　ディープラーニングの先にあるもの」松尾豊著
「これならわかる深層学習入門」瀧雅人著
「機械学習入門　ボルツマン機械学習から深層学習まで」大関真之著