

2020.4.14 (火)

ビジネスサポート・BINGO 2020年4月定例会

ディープラーニングの 中核技術を識る

フルカワ技研 古川 昇

場所：福山職業能力開発短期大学校



本日の予定

I 「知る」か、「識る」か。

II ディープラーニングとは

III 「教師あり」学習と「教師なし」学習

IV 回帰 (Regression) と分類 (Classify) の相違

V 分類とクラスタリング (clustering) の相違

VI 偏微分について

VII 誤差逆伝播法と損失関数

まとめ



I 「知る」か「識る」か。

1 「知る」と「識る」の違い。

柳宗悦（やなぎ むねよし）：1889－1961

民芸運動。焼物との出会い。茶の湯の美。

柳宗悦：柳宗悦茶道論集青169-6，岩波文庫(1987)

II ディープラーニングとは

データに潜むルールやパターンを、機械が学習によって得る。機械に学習する能力を持たせることによって、機械に明示的にルールを与えなくても自ら判断をする。

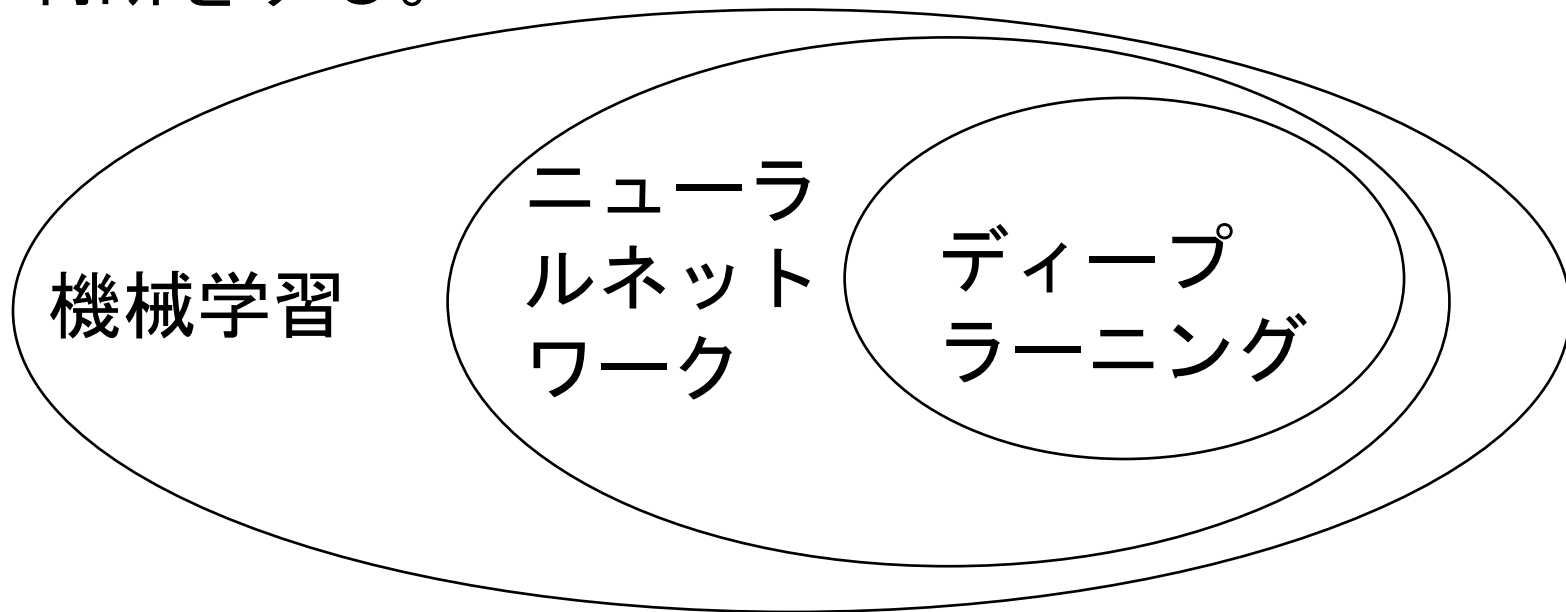


図1 機械学習、ニューラルネットワーク、ディープラーニングの関係。ニューロン：動物の脳にある神経細胞のこと。



ディープラーニング（深層学習）

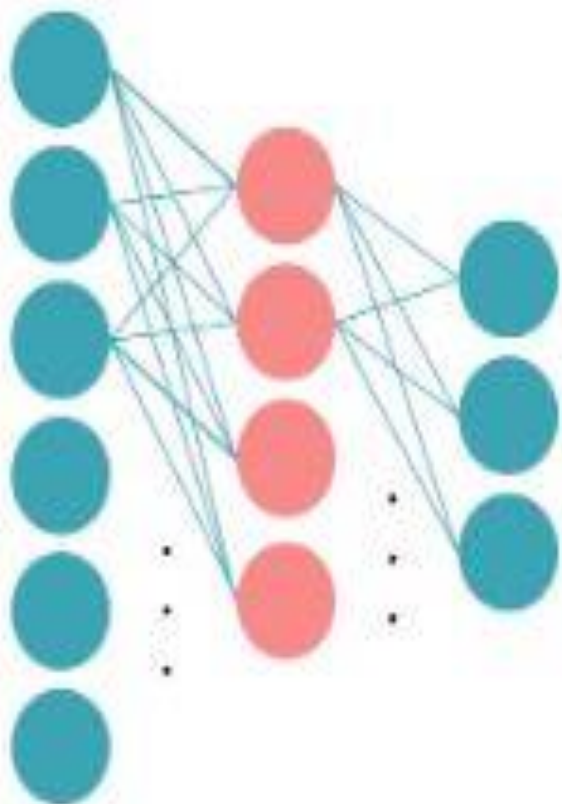
それまでの機械学習では、ある程度人間によるガイダンスが必要だったが、深層学習では機械が自ら何を学習するか決めることができる。

ディープラーニングは、極めて高度なニューラル・ネットワークを活用する最先端の高度な機械学習手法。ディープ・ラーニングと呼ばれるのは、生成されるモデルが従来型のニューラル・ネットワークと比べ非常に複雑である（＝階層が深い）から。また、ディープ・ラーニング・モデルは、従来のモデルに比べ極めて大量のデータを処理できる点も特長。

従来型のNNとディープラーニング

ニューラルネットワーク

入力層 中間層 出力層



ディープラーニング

入力層 中間層 出力層

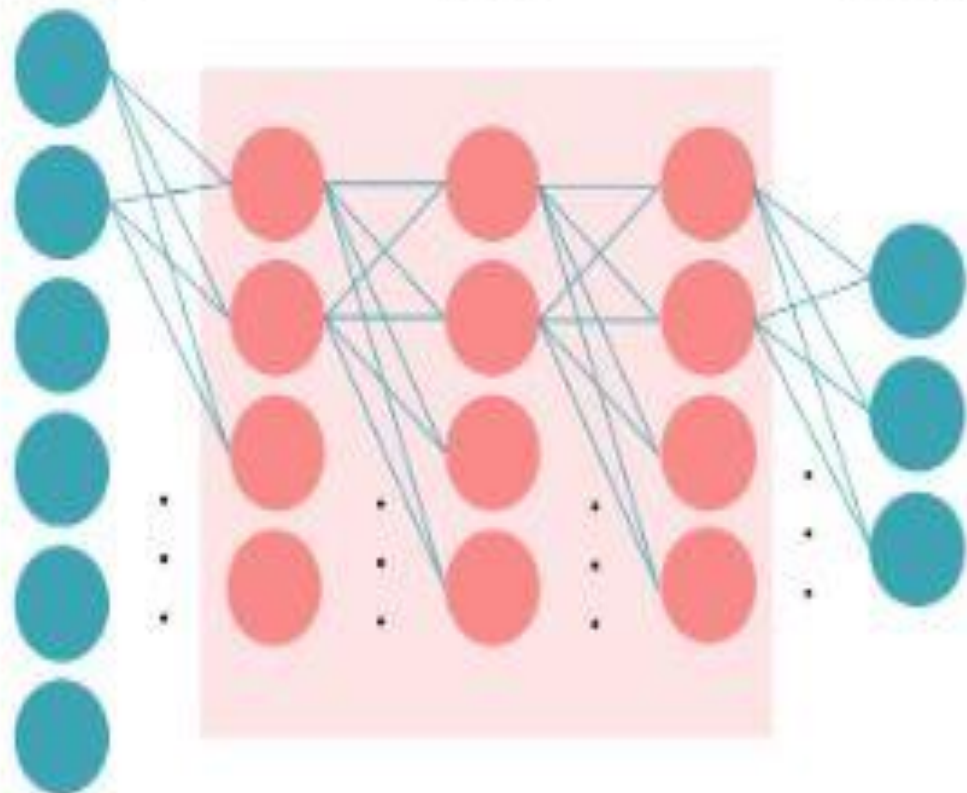


図2 従来型のNNとディープラーニング



ディープラーニング（深層学習）

Deep Learningとは、十分なデータ量があれば、人間の力なしに機械が自動的にデータから特徴を抽出してくれるディープニューラルネットワーク（DNN）を用いた学習のこと。

DNNとは、ニューラルネットワーク（NN）というパターン認識をするように設計された、人間や動物の脳神経回路をモデルとしたアルゴリズムを多層構造化したもの。

<https://leapmind.io/blog/2017/06/16/ディープラーニング（deep-learning）とは？【入門編】/>



Ⅲ 「教師あり」学習と「教師なし」学習

教師あり学習

- ・ 入力（説明変数，独立変数）と出力（目的変数，従属変数）がセットになったデータがあることが前提。「正解のラベル」をつけて学習する。
- ・ 入力データをもとに，出力データを予測，識別。
- ・ 予測には，連続する値もあれば，連続しない値（カテゴリー）もある。
- ・ 未知のデータが入力されても，正解ラベルと近い値が出ているか，を計算する。

例

- ・ 過去のデータから売上予測。
- ・ 何の画像かを分類・識別する。



教師なし学習

教師なし学習

- ・ 目的変数（ラベル）は無し。正解のラベルは無い。
- ・ 入力と出力がセットになったデータがない。
- ・ 入力データの構造や特徴，法則をつかむ。
- ・ 入力データ間の関係をつかむ。
- ・ データのクラスター（塊）や構造を見つけ出すのが，クラスタリング。似たもの同士を集めてグループ化。

例

- ・ 顧客のグループ化（セグメンテーション），ターゲティング（商品ターゲットの絞り込み）
- ・ EC（electronic commerce，電子商取引）サイト。顧客の好みを分析して，商品のレコメンデーション。顧客ごとに商品を推薦。



教師なし学習

教師なし学習

- ・主成分分析 (PCA: Principal Component Analysis)
ディープラーニングではないが、統計学では、多変量解析の手法として頻繁に応用されてきた。
- ・相関をもつ多数の説明変数から、相関の少ない少数の説明変数に変換する。
- ・次元削減の手法として、現在も頻繁に利用されている。

用途


データの圧縮：計算機資源の有効活用

データの可視化：データの見える化，グラフ，図で表現する。



半教師あり学習

- ・ 工場のIoTの現場で，得られている全データを一度に使ってしまうのではなく，一部を残しておいて検証する場合が多い。
- ・ 学習データそのものが不足していて，得られているデータでは少なすぎる場合がある。 →
データを使い回す。
新規にもっとデータをとる。



IV 回帰 (Regression) と分類 (Classify) の相違

回帰

- ・ 連続するデータの予測

例

- ・ 株価分析

分類

- ・ カテゴリーデータ（離散値）を予測。与えられたクラスに分ける（ラベリング）。
- ・ 二値分類（ある事象が起こるか、起こらないか）、

多値分類

例

- ・ 迷惑メールか否か。
- ・ アヤメの花の種類のカテゴリ分類・識別



V 分類とクラスタリング (clustering) の相違

分類

- ・クラス (class) 分け。前もって分けるクラスが分かっている。classification
教師あり学習 (Supervised Learning)

クラスタリング

- ・Clustering
- ・いくつかの類似グループ (塊, クラスタ) に分ける。いくつかのクラスタに分かれるか, 分からない。
教師なし学習 (Unsupervised Learning)



VI 偏微分について

微分はdifferentiation , differential。

偏微分では, 多変数関数を「特定の文字以外は定数だとみなして」微分する。

xとyについての関数

$$f(x, y) = x^3 + y^2 + 6y + xy$$

xについての偏微分は

$$\frac{\partial f(x, y)}{\partial x} = 3x^2 + y$$

VII 誤差逆伝播法と損失関数

○で表わされるノード間は、エッジでリンクされている。

エッジは「重み」をもつ。ノードの値は、1つ前の層のノードの値と、エッジの重みを乗算し、それらを合算して求めた上で、活性化関数で変換する。

入力層から右へ計算していくのが、順伝播。

逆に右から左へ計算していくのが、逆伝播。

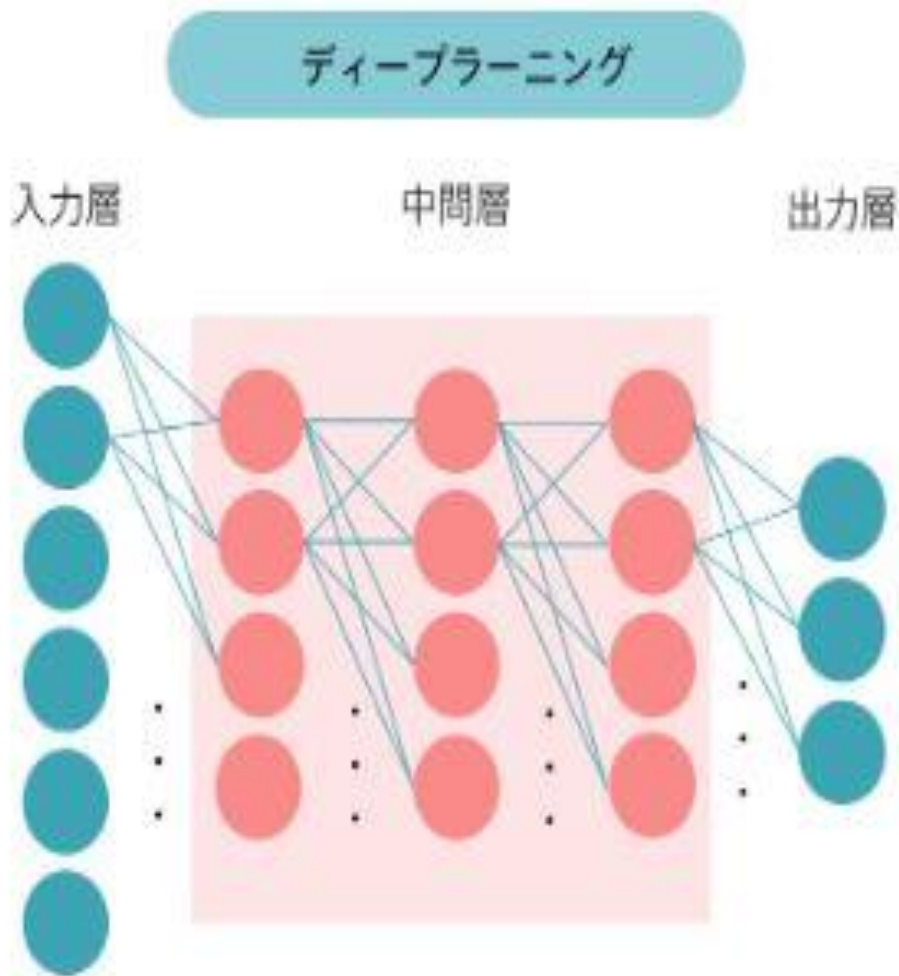


図3 ディープラーニングの模式図



損失関数とは

損失関数は誤差関数ともいう。

- ・ 順伝播のあと，出力層では正解ラベル（目的変数）の値と比較して，損失関数により，誤差を求める。

誤差が小さくなるように，エッジの重みを更新する。

- ・ 誤差が最小になるように，エッジの重みを更新するのが，「勾配降下法」。

ここに偏微分が必要となる。

なお誤差の最小値でなく，極小値である局所最適点で計算を打ち切ることになる可能性もある。



まとめ

- 1 従来のニューラルネットワークは中間層が1個、ディープラーニングは複数。
- 2 「教師あり学習」「教師なし学習」「半教師あり学習」を状況に応じて使い分ける必要がある。
- 3 順伝播と誤差逆伝播, そして損失関数の関係。誤差が最小になるように, エッジの重みを更新する。



参考文献

- 1 足立悠:Neural Netwok Console入門, リックテレコム(2018)
- 2 浅川伸一ほか:ディープラーニングG検定公式テキスト, 翔泳社(2019)
- 3 秋庭伸也ほか:機械学習図鑑, 翔泳社(2019)
- 4 塚本邦尊ほか:東京大学のデータサイエンティスト育成講座, マイナビ出版(2019)

Windows は米国Microsoft Corporationの、米国およびその他の国における登録商標または商標です。



謝辞

ご清聴ありがとうございました。