

PythonとKerasによるディープ ラーニング

第1章 ディープラーニングとは何か

1.2 ディープラーニングの前史:機械学習

19ss317u 趙一

1.2.1 確率モデリング

•確率モデリング: データ分析に統計学の原理を応用するものです。

↓ (よく知られているアルゴリズムの1つ)

•ナイーブベイズアルゴリズム (単純ベイズアルゴリズム) :

入力データの特徴量がすべて独立しているものと仮定した上で、
ベイズの定理を適用することに基づく機械学習分類器です。

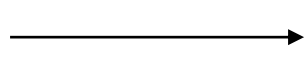
↓ (深く関連しているモデルの1つ)

•ロジスティック回帰: 回帰アルゴリズムじゃなくて、分類アルゴリズム
です。

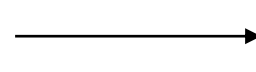
特性: 単純で融通が利く

1.2.2 初期のニューラルネットワーク

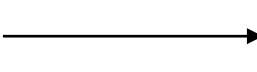
1950s



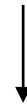
1980s中期



1989年



1990年



ニューラルネットワーク中核的な概念が模型を使って調査されていました。

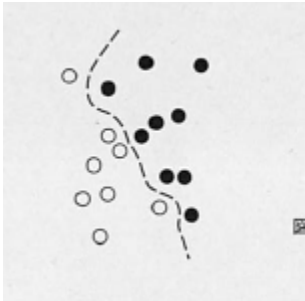
何人かの研究者によって、バックプロバゲーション（誤差逆伝播法）アルゴリズムが再評価され、ニューラルネットワークが応用されるようになりました。

Yann LeCunがニューラルネットワークとバックプロバゲーションを組み合わせ、手書きの数字の分類という問題に応用しました。

USPS(United States Postal Service)によって、導入されています。

1.2.3 カーネル法

- カーネル法: サポートベクトルマシン(SVM)をはじめとする。分類アルゴリズムの一種です。
- SVMの目的: 分類問題を解くことです。



SVMで2つ異なるカテゴリのデータ点をうまく分類する**決定境界**を見つけ出す。

本書16ページ図1-10より引用

- SVM が決定境界を見つけ出す方法:
 - 1)決定境界を超平面として表現できる新しい高次元表現へデータをマッピングする。
 - 2)マージン最大化:各クラスに最も近い点と超平面との距離を最大化することで、データ点をうまく分類する決定境界を計算する。

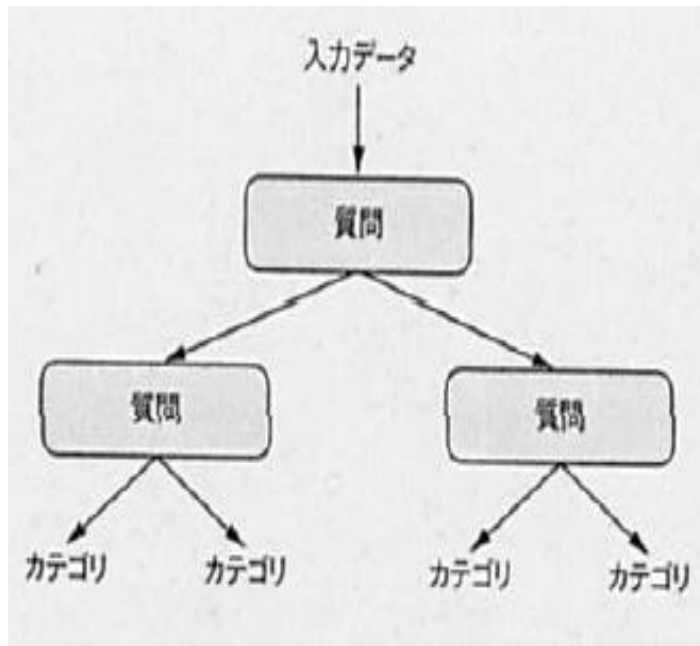
1.2.3 カーネル法

- データを高次元表現へマッピングする時、計算困難な演算になった場合は、カーネルトリックを使っています。
- SVMの長所:
 - 1)単純な分類問題において圧倒的な性能があります。
 - 2)重大な数理解析に適していたことから、よく理解されており、解釈しやすいです。
- SVMの短所:

大規模なデータセットへのスケーリングが難しく、画像分類などの知覚問題でよい結果を出せませんでした。

1.2.4 決定木、ランダムフォレスト、勾配ブースティングマシン

- 決定木:



入力データ点を分類したり、入力に基づいて出力を予測したりすることが可能なフローチャートのような構造です。

本書17ページ図1-11より引用

- ランダムフォレスト:

ランダムフォレストとは、決定木を大量に構築し、それらの出力を組み合わせるアンサンブル学習のアルゴリズムであり、幅広い問題への応用が可能です。

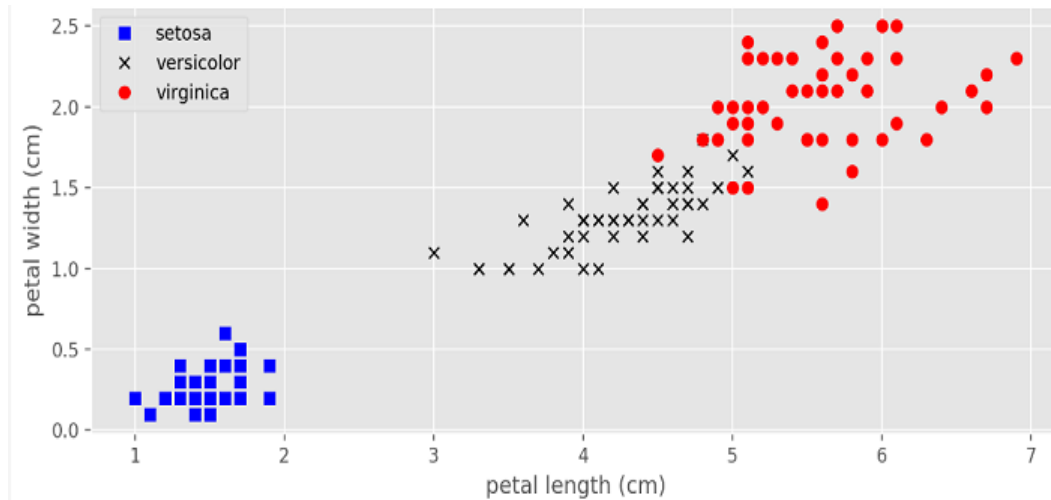
- 勾配ブースティングマシン(GBM):

弱い予想モデルのアンサンブル入力に基づく機械学習法であり、勾配ブースティングを使用します。ディープラーニングと並んで、Kaggleのコンペで最もよく使用される手法の1つです。

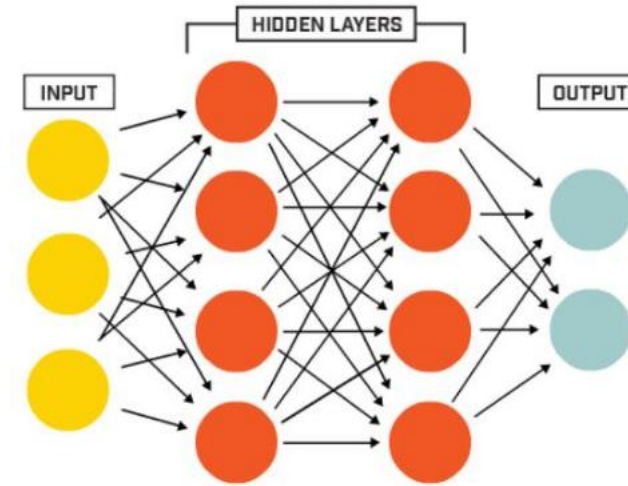
1.2.5 ニューラルネットワークの再起

- 2010年: ニューラルネットワークは見放されて、まだ取り組んでいて、重要なブレークスルーが起きた。
- 2011年:ニューラルネットワークが最初の勝利を収めました。
- 2012年以降:ディープ畳み込みニューラルネットワーク(DCNN)はあらゆるタスクにとって中心的なアルゴリズムとなっています。
- 2015年:ImageNet(難易度高いコンテスト)での分類タスクはDCNNで完全に解かれた問題と見なされています。

1.2.6 ディープラーニングの違いは何か



シャローラーニング



ディープラーニング

- 特徴エンジニアリング: 複雑な問題に必要なとなる高度な表現は、最初の入力データを変換してそうした手法に適したものにすることを作業を行わなければならない、そうした表現に適した層を手作業で設計しなければならないということです。
- 特徴エンジニアリングに対して、ディープラーニングでは完全に自動化され、全ての特徴量を1回で学習することができます。

1.2.7 現代の機械学習の情勢

- XGBoostが使用されている理由:
 - pythonとRをサポートしています。
- Kerasが使用されている理由:
 - 1)使いやすい
 - 2)柔軟性
 - 3)pythonのサポート
- 現在の応用機械学習で最も理解しておかなければならない
手法:シャローラーニング問題で勾配ブースティングマシンと知覚問題でディープラーニングです。
技術:XGBoostとKerasに精通していることです。

これらは現在Kaggleで最もよく使用されているライブラリです。