

1.2 デイープラーニングの前史

機械学習

17T4015S

内間けんじ

目次

1.2.0 : ディープラーニングと機械学習

1.2.1 : 確率モデリング

1.2.2 : 初期のニューラルネットワーク

1.2.3 : カーネル法

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

1.2.5 : ニューラルネットワークの再帰

1.2.6 : ディープラーニングの違い

1.2.0 ディープラーニングと機械学習

近年**ディープラーニング**が注目を集めるようになった

→しかし、ディープラーニングが常に正しいツールである訳ではない。
(データが十分でなかったり別のアルゴリズムが適切である場合)

別の機械学習のアルゴリズムによるアプローチを理解し、それらを必要に応じ実践する必要がある。

→従来の機械学習を知り、ディープラーニングがどのように誕生し、なぜ重要なのかを理解する必要。

1.2.1 確率モデリング

確率モデリングとはデータがどのような確率分布に従うかを探る試み。最も良く知られるアルゴリズムに**ナイーブベイズアルゴリズム**がある

○**ナイーブベイズアルゴリズム**

ナイーブベイズは、入力データの特徴量が全て独立している物と仮定しベイズの定理を利用したもの。

この形式のデータ分析はコンピュータが実装される前(おそらく1950年以前)より存在しており、手作業で行われていた。

1.2.2 初期のニューラルネットワーク

1950年代：ニューラルネットワークの中核的な概念は知られていたが、大規模なニューラルネットワークを効率良く訓練する方法が不明だった。

1980年代：誤差逆伝搬法がニューラルネットワークに応用されるようになり、1989年にYann LeCunが畳み込みNNと誤差逆伝搬法の概念を組み合わせることでニューラルネットワークを実現した。

ニューラルネットワークが注目され始めた。

1.2.3 カーネル法

カーネル法の登場により、1990年代に一目置かれるようになったニューラルネットワークはすぐに忘れられてしまう。

○カーネル法

カーネル法はサポートベクトルマシンを始めとする分類アルゴリズムの一種。

カーネルトリックという二点間の距離を計算することで決定超平面を見つける手法により分類を行う。

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

決定木とはフローチャートのような構造になっており、入力データ点を分類や入力に基づいて出力を予測する。決定木の研究は2000年代に盛んになり、2010年にはカーネル法よりも有線されるほどだった。

特に決定木を大量に構築した**ランダムフォレストアルゴリズム**では幅広い問題への対応が可能だった。

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

2010年にはランダムフォレストが多く使用されていたが、2014年には勾配ブースティングマシンが人気になり始める。

○勾配ブースティングマシン(GBM)

GBMは前のモデルの弱点に対処する新しいモデルの訓練を繰り返すことで機械学習モデルを改善する手法である。

決定木に適用すると、殆どの場合ランダムフォレストの性能を上回り、同じような特性を持つモデルを得ることができる。

1.2.5 ニューラルネットワークの再帰

2010年頃はニューラルネットワークはまだまだ注目されてなかった。

→2011年：IDSIAのDan CiresanがDNNによって画像分類コンペで勝利

→2012年：Alex Krizhevsky率いるチームが画像認識コンテスト
ImageNetで好成績を残す。

2012年以降、ディープ畳み込みニューラルネットワーク(DCNN)はコンピュータビジョンのあらゆるタスクの中心となった。

1.2.6 ディープラーニングの違い

ディープラーニングがより良い性能を実現したことで人気になり始めたが、問題の解決が容易であることもその理由の1つであった。

機械学習では学習を行うまえに入力データを学習するためのデータへと変換する作業が必要であった（特徴エンジニアリング）

→ディープラーニングではこの手順が自動化されすべての特徴量を一回で学習できる。