

金融市場における機械学習の活用について

目次

- I. はじめに
- II. 機械学習とは
- III. 機械学習モデルの証券投資への活用
- IV. 終わりに

資産運用部 先端金融工学運用室 シニアファンドマネージャー 徳山 相賢
ファンドマネージャー 神田 裕樹

I. はじめに

近年、IT 技術の進展やビッグデータと呼ばれる大量かつ多様なデータの活用拡大に伴い、機械学習というデータ分析手法が注目を集めている。機械学習とは、与えられた大量のデータからコンピューターが法則性を学習し、予測や識別(グループ分け)を行う手法である。例えば製造分野では不良品検知や製品の需要予測、マーケティング分野では EC サイトのレコメンデーション機能¹等で活用されており、その応用範囲は多岐に亘っている。金融の分野においても、ロボアドバイザー²や融資業務、市場分析業務などで機械学習が活用され始めており、その活用範囲は急速に広がりを見せつつある。

本稿の主旨は、金融業界におけるデータ分析の視点やデータ活用方法について、特に証券投資分野の活用事例の紹介を通して読者に広く知っていただくことである。

先ず II 章では、機械学習について馴染みの薄い読者向けに入門的解説を行う。併せて、機械学習活用者の裾野の広がりについて言及する。III 章では、金融業界における機械学習活用の一般事例について説明した後、さらに証券投資分野に焦点を絞って具体的な活用事例を紹介していく。IV 章では、機械学習モデルの金融市場への活用に関する今後の展望について述べる。

¹ 過去の購買履歴などをもとに顧客の嗜好を分析し、顧客の興味・関心がありそうな情報を提示する仕組み。

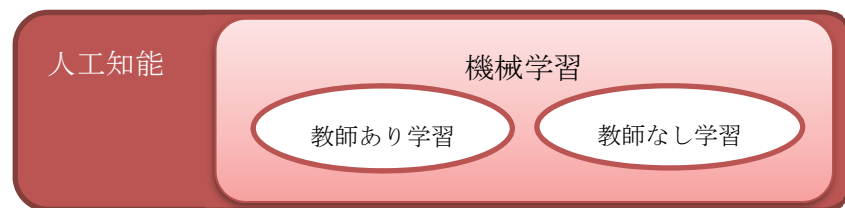
² 顧客特性に応じた運用のアドバイスや運用の代行を提供するシステム。

Ⅱ. 機械学習とは

近年、コンピューターによる計算処理速度の向上、利用可能なデータの増加、クラウドを活用した大量のデータ処理技術の発展などを背景として、「人工知能(AI)」という象徴的な言葉と共にデータの活用が急速に広がっている。人工知能の定義は様々あるが総務省によれば、「人間の思考プロセスと同じような形で動作するプログラム、あるいは人間が知的と感じる情報処理・技術」とされている。この人工知能を形作るアプローチの一つが本稿で取り上げる機械学習である。

機械学習の定義も明確なものはないが、あえて定義づけるとすれば、「与えられた大量のデータからコンピューターが法則性(ルール)を学習し、学習結果をもとに予測や識別等を行う手法」といえる。より砕けた言い方をすれば、人工知能が、スマートフォンに疑問を投げかけた際に応答を返してくれるような知性を感じさせる汎用的かつ概念的な意味を持つのに対し、機械学習はメールの自動分類や自動運転技術など特定のタスクに用いられる手段・方法を指す。人工知能と機械学習は混同して語られることもあるが、人工知能の方がより広義な概念であることに対して、機械学習は人工知能よりも狭義の概念である(図表1)。本稿では、機械学習に焦点を当てて議論する。

図表1：人工知能(AI)と機械学習の関係



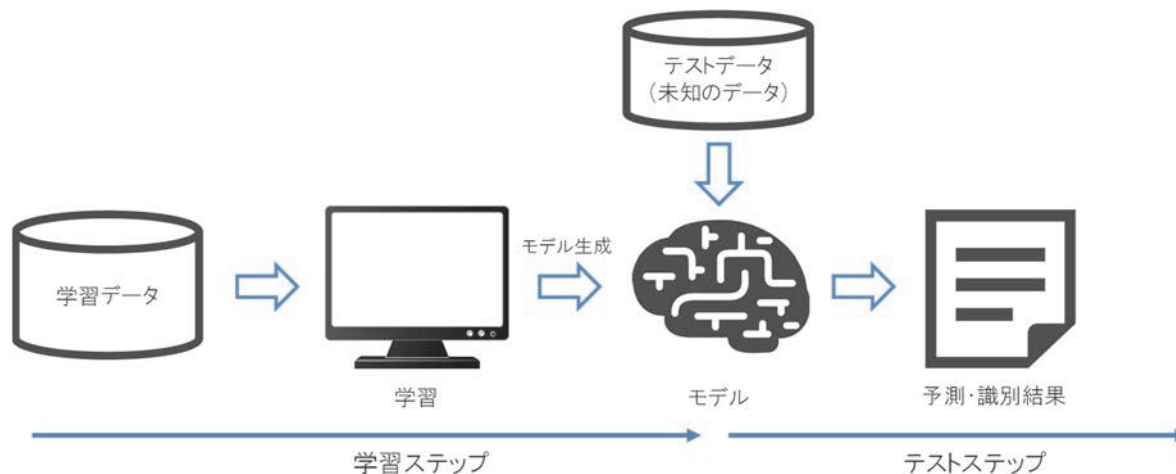
(出所)三菱UFJ信託銀行作成

機械学習の技術を使ってできることは大きく2つ、「予測」と「識別(グループ分け)」である。前者は、教師あり学習と呼ばれる手法が一般的に用いられ、後者は、教師なし学習と呼ばれる手法が用いられる。教師あり学習は、コンピューターに入力するデータ(学習データ)の中に、人間が予め用意した正解(教師)を与えた状態で学習させることが特徴である。これは受験勉強で過去問を解くプロセスに例えることができ、過去問題集(問題と解答の組み合わせ)を解きながら「こういう傾向の問題にはこういう解答パターンが当てはまる」といった学習を繰り返していくイメージである。一方、教師なし学習は、人間が正解を与えることをせず、学習データに内在する法則性をコンピューターが自ら学習することが特徴である。過去問の例でいうと、沢山の問題の中から類似した問題どうしをグループ分けし、出題傾向を掴むようなイメージである。

機械学習モデルの学習プロセスを示したのが図表2である。まず、学習データ(過去問)からコンピューターが法則性を学習し、モデルが生成される。生成されたモデルに対して、テストデータ(試験問題)を入力すると、コンピューターが学習データから得られた法則性に

従って、教師あり学習であれば予測(試験問題の解答を出す)、教師なし学習であれば識別(試験問題のカテゴリを回答する)を行うといった流れとなる。

図表 2 : 機械学習モデルの学習プロセス



(出所)三菱 UFJ 信託銀行作成

機械学習には、教師あり学習と教師なし学習以外の手法もあるが³、ここからは実務で一般的によく使われている教師あり学習と教師なし学習について、具体例を交えながら説明する。紙面の都合上、一断面のデータを用いた単純な事例説明にとどめ、学習プロセスの詳細説明については割愛する。

1. 教師あり学習と教師なし学習

ここでは、読者の皆様に教師あり学習と教師なし学習の簡単なイメージをもってもらうために、住宅ローンのデータを使った例を挙げる。住宅ローンを提案し顧客の課題解決を行いたい場合、担当している顧客全員に対して住宅ローンの提案を行うのは現実的ではない。そこで過去に担当した顧客の情報や、住宅ローン提案の成否などの情報(図表 3)に基づき、ニーズがありそうな顧客を機械学習により予想・グルーピングしてみる。

図表 3 : 住宅ローンの顧客データ

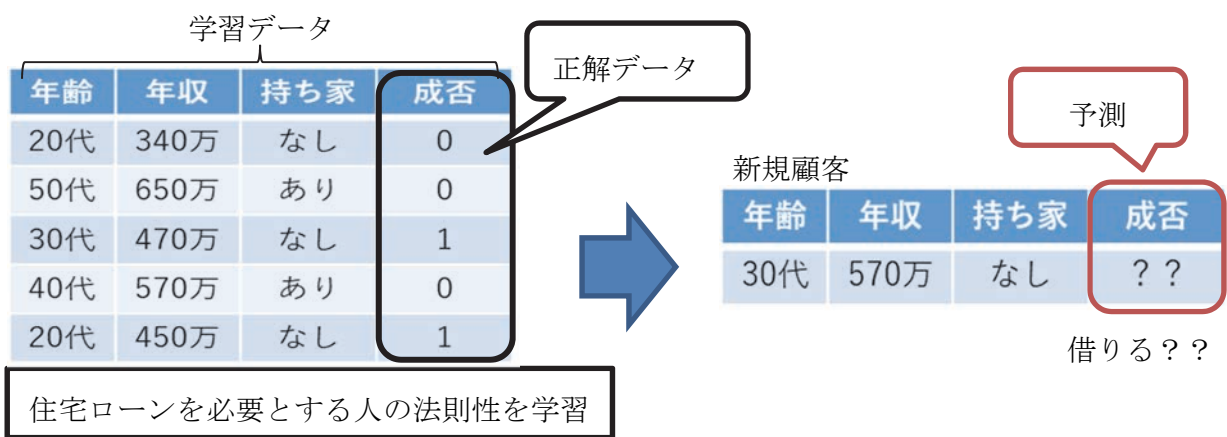
年齢	年収	持ち家	成否
20代	340万	なし	0
50代	650万	あり	0
30代	470万	なし	1
40代	570万	あり	0
20代	450万	なし	1

(出所)三菱 UFJ 信託銀行作成

³ そのほかにも正解の教師データを与えるのではなく、目的として設定した「報酬(スコア)」を最大化する方法を学習する強化学習と呼ばれる手法や、半教師あり学習と呼ばれる教師あり学習と教師なし学習の間をとったような手法もある。

まずは教師あり学習を用いて、住宅ローンのニーズがありそうな顧客を予測する問題を考える(図表4)。教師あり学習は、学習データに正解(予測したい項目)を与えた状態で法則性を学習させることが特徴であるが、ここでは正解データとして、過去の提案の「成否」、学習に利用する他のデータは「年齢・年収・持ち家」とする。ここでモデルに期待することは、何らかの法則性に基づき、提案が成功しそうな対象者を抽出することである(年収が550万以上で20代~30代の人ほど需要が高いなど)。教師あり学習によって得られた法則性を活用することで、新規顧客に対して住宅ローンを提案すべきか否かに関する意思決定を効果的に行うことができる。

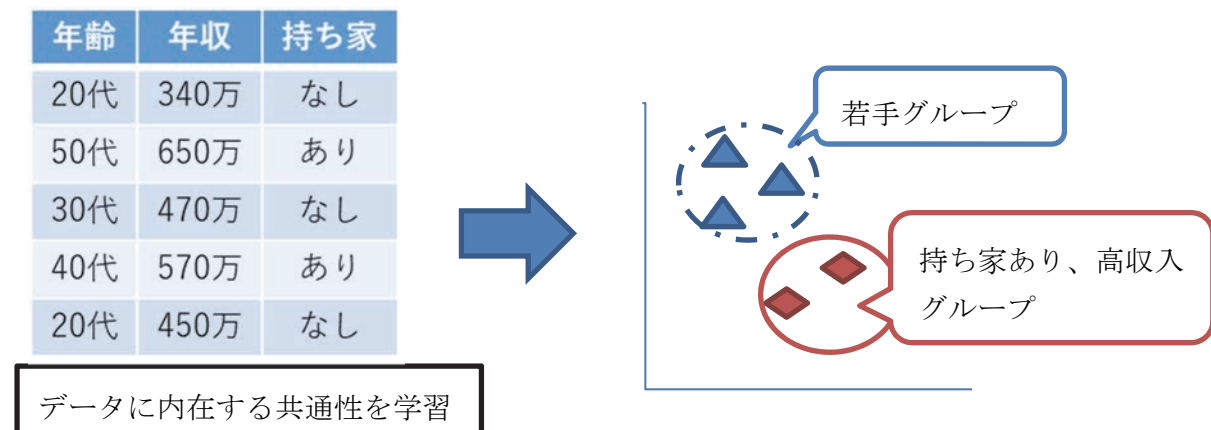
図表4：教師あり学習のイメージ



(出所)三菱UFJ信託銀行作成

次に教師なし学習を用いて、顧客をグルーピングする問題を考える(図表5)。教師なし学習は、学習データに正解(予測したい項目)を与えるのではなく、データに内在する法則性などを学習することが特徴であるため、先ほどの学習データの中から正解データである「成否」項目を除く。この事例では、「若手グループ」と「持ち家あり、高収入グループ」といったグルーピングが想定される。今回はシンプルな設定であるため、わざわざ教師なし学習にかける意義は小さいが、その他にも考慮すべき要素が多くなった場合には、教師なし学習によるグルーピング(クラスタリング)により、新たな顧客層の抽出などが可能となる。

図表5：教師なし学習のイメージ



(出所)三菱UFJ信託銀行作成

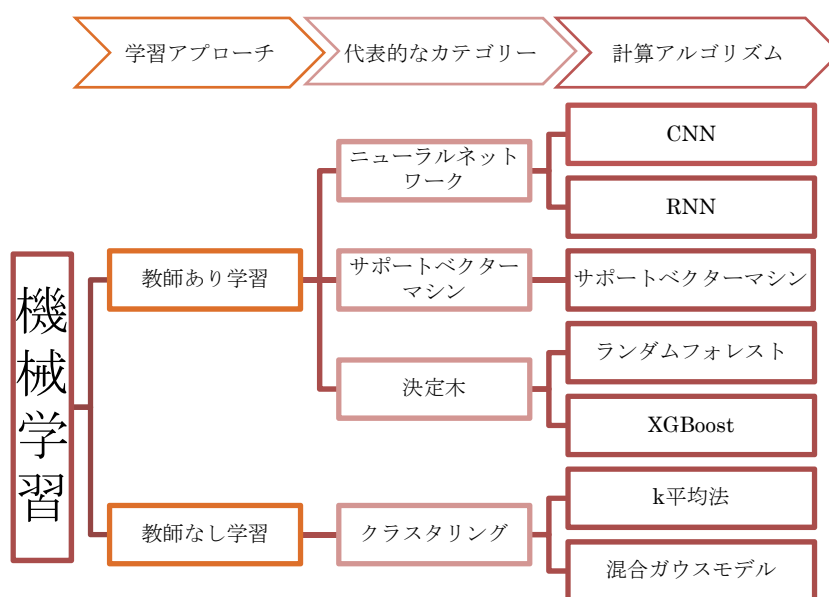
ここまで教師あり学習と教師なし学習の具体的な事例について紹介した。簡単にまとめると、教師あり学習は「予測」に適した手法で、住宅ローンの提案成否といった明確な正解データ(予測したい項目)と顧客情報(年齢・年収、持ち家状況)の間にある関係性について学習することにより、新規顧客に対する住宅ローンの提案活動を効果的に行うことが期待できる。一方、教師なし学習は「識別(グループ分け)」に適した手法で、顧客情報をグループ分け(可視化)しデータに内在する法則性を学習することで、認知していなかった顧客層の発見などが期待できる。

機械学習は使い方によっては、非常に強力なツールではあるが万能ではない。使用用途やデータの特性に応じて適切なツールを使い分ける必要がある。次節では、機械学習でよく使われるツール(アルゴリズム)について紹介する。

2. 機械学習の代表的なアルゴリズム

ここでは「教師あり学習」と「教師なし学習」の代表的な計算アルゴリズムについて概要と特徴を解説する。教師あり学習に使うアルゴリズムの代表的なカテゴリーとしてニューラルネットワーク⁴やサポートベクターマシン、決定木などが挙げられる。それぞれのなかでも、画像認識に強い計算方法や時系列のデータに適した計算方法など、計算アルゴリズムの違いで様々なモデルが存在する。教師なし学習については、クラスタリング(グループ分け)の代表的な手法であるk平均法(kmeans)、混合ガウスモデル(GMM)といったモデルを紹介する。学習アプローチ、代表的なカテゴリー、具体的な計算アルゴリズムの観点で機械学習モデルをまとめたものが図表6である。

図表6：機械学習の手法分類



(出所)三菱UFJ信託銀行作成

⁴ 教師なし学習で利用されることもある。

(1) 教師あり学習

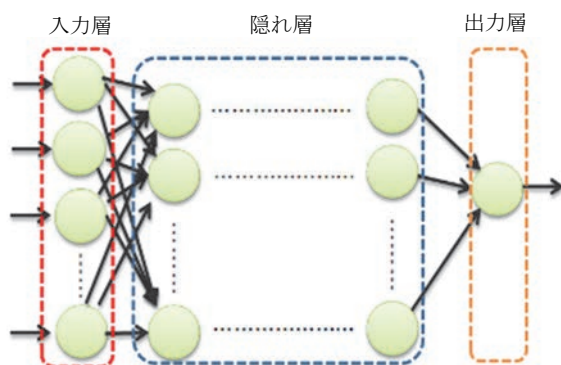
① ニューラルネットワーク

ニューラルネットワークは、人間の脳神経にある神経細胞(ニューロン)を数理モデルで表現したもので、入力層と隠れ層(受け取ったデータから特徴を抽出する層)、出力層から構成されるモデルである(図表7)。各種メディアでよく取り上げられるディープラーニングもニューラルネットワークの一種である。ニューラルネットワークには、画像認識の分野で広く使われている畳み込みニューラルネットワーク(CNN: Convolutional Neural Network)や、文章(テキストデータ)を解析する自然言語処理の分野で活用されている回帰型ニューラルネットワーク(RNN: Recurrent Neural Network)など、用途に応じて様々なモデルが考案されている。

② サポートベクターマシン

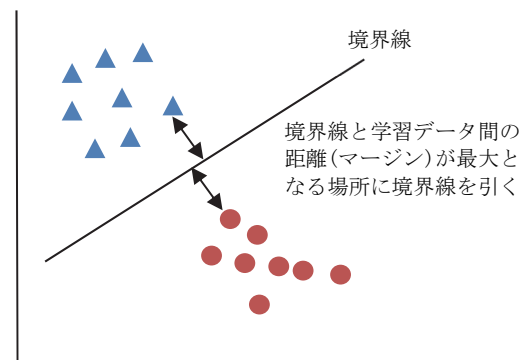
サポートベクターマシンは、パターン認識の分野で非常によく知られた手法の一つで、文章分類やスパムメールか否かの判定など様々な用途で利用される。少量の学習データでも良好な予測精度を得られやすい点が特徴であり、特に2つのデータの集合をグループ分けするような時に威力を発揮しやすいモデルである(図表8)。

図表7: ニューラルネットワーク



(出所)三菱UFJ信託銀行作成

図表8: サポートベクターマシン

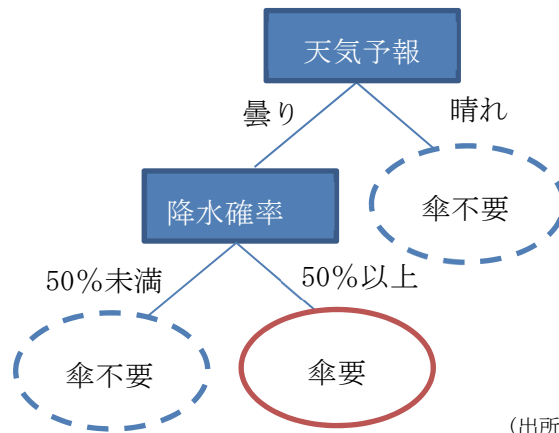


(出所)三菱UFJ信託銀行作成

③ 決定木

決定木は、その名のとおり、ある変数に基づき条件分岐を行うツリー構造のモデルで、解釈性が高いことなどから人気のあるアルゴリズムである。例えば、図表9のような、傘を持参するか否かを判断する場合であれば、天気予報や降水確率といった変数を基に条件分岐を行う。決定木には、ツリーを大量に生成するランダムフォレストや、ツリーを構築する際にこれまでに作ったツリーを活用して、新たな決定木を構築するXGBoostなど様々な計算アルゴリズムがある。

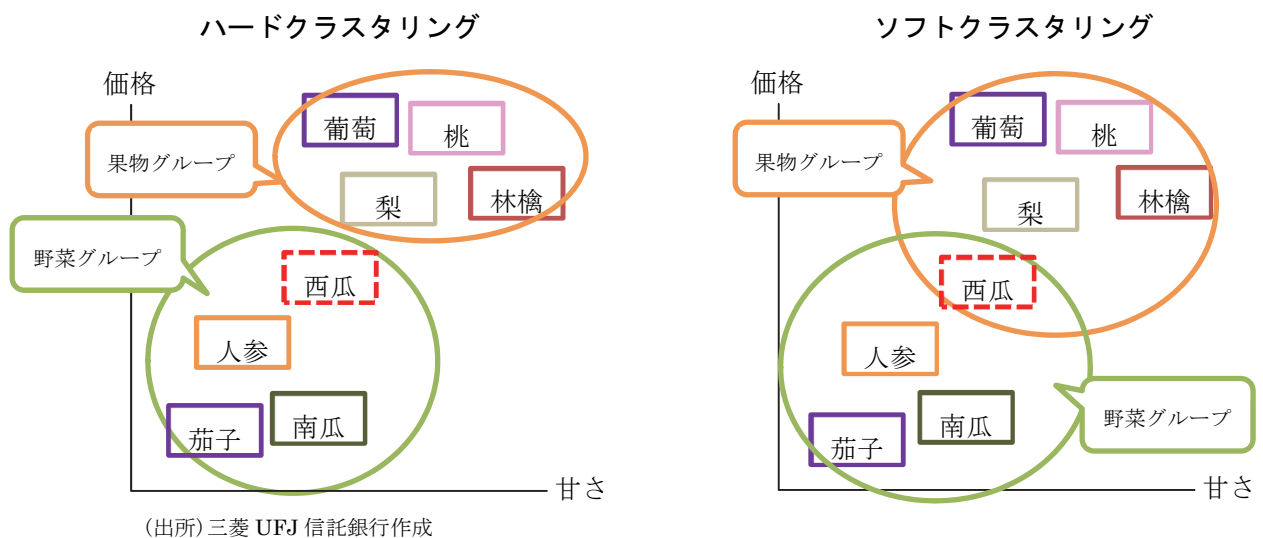
図表 9 : 決定木のイメージ



(2)教師なし学習(クラスタリング⁵)

クラスタリングは、データの類似度合いなどに従って、似ているものをグルーピングする際に使われる手法である。クラスタリングには、各データを1つのグループのみに所属するようグループ分けするハードクラスタリングと、複数のグループに所属することを許してグループ分けするソフトクラスタリングの2つがある。前者で代表的なモデルはk平均法、後者では混合ガウスモデルなどが挙げられる。例えば、果物と野菜に関するイメージデータ(価格と甘さのイメージ)がある場合、クラスタリングにより西瓜(スイカ)のグループ分けがどのようになされるかを図表 10 に例示した。ハードクラスタリングであれば、西瓜は「野菜グループ」のみに所属するようにグループ分けされ、ソフトクラスタリングであれば、「野菜グループ」でもあり「果物グループ」でもあるとグルーピングされることを許容する。

図表 10 : クラスタリングのイメージ



⁵ 教師なし学習には、主成分分析や tSNE といった次元圧縮の手法や、クラスタリングなど様々な手法が存在するが、紙面の関係上本稿では、クラスタリングについて紹介する。

3. 機械学習の民主化

ここまで、機械学習の概要や代表的なアルゴリズムについて簡単に紹介してきた。本節では、機械学習活用者の裾野が広がりつつある昨今の状況について言及しておく。一般的に、機械学習を活用して実課題を解決しようとする際に必要とされるスキルとしては、アルゴリズムを理解するための統計学や数学に関する知識、モデルを実装するためのプログラミング能力、そして業務知識が挙げられ、多くの人にとってハードルが高い。実際、かつて機械学習モデルを構築するには、プログラミングや統計学の知識に精通した人材(システムエンジニアやデータサイエンティスト)が不可欠であった。しかし近年では、IT 技術やサービス基盤の進化を背景に、技術的バックグラウンドが豊富ではないが業務知識・アイデアが豊富な人材にとってもモデル構築のハードルが着実に低下している。本節では、こうした「機械学習の民主化」と呼ばれる動きが広まりつつある背景について、簡単にまとめる。

(1) プログラミングの敷居低下

機械学習の実装には、プログラミング言語 Python などが広く利用されている。Python は機械学習に関するライブラリ⁶が充実しており、例えば `scikit-learn` という Python ライブラリを使うと、簡単なものであれば数行のプログラムコードで機械学習を実装することができる(図表 11)。また、機械学習に関する様々なプログラムコードが `GitHub`⁷や `Kaggle`⁸といったサイト上で日々公開・共有されており、世界中からプログラムの実装に必要なノウハウが比較的容易に入手可能となってきた。

図表 11 : `scikit-learn` を活用した機械学習の実装例

```

from sklearn.ensemble import RandomForestClassifier
clf_rf = RandomForestClassifier()
clf_rf.fit(X_train, y_train)
y_pred = clf_rf.predict(X_test)

```

法則性の学習

アルゴリズム選択
(ここではランダムフォレストを選択)

法則の適用

(出所)三菱 UFJ 信託銀行作成

(2) 機械学習の自動化

機械学習モデルの構築プロセスは図表 12 に示すように、アイデア・仮説定義からモデル実装まで大きく 6 つのプロセスで構成され、すべてのステップを人間が一つ一つ進めていくのが通常である。近年の動向としては、6 つのプロセスのうち、前処理⁹から

⁶ 汎用性の高い複数のプログラムを再利用可能な形でひとまとまりにしたもの。

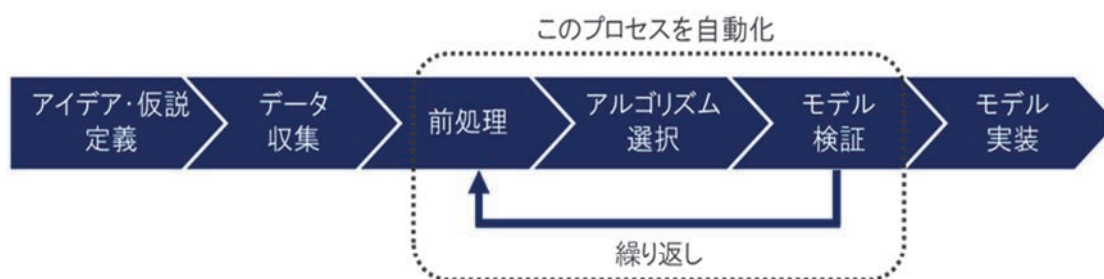
⁷ ソフトウェア開発者のためのプラットフォームであり、オープンソースのプログラムコードが多数掲載されている。

⁸ 世界規模の機械学習コンペティション。他のコンペ参加者のプログラムコードも公開されている。

⁹ 機械学習で利用するデータの品質を高めるためにデータの中から重複や欠損、誤表記などを探し出して削除や修正を行い、データの品質を上げること。データ分析は前処理の時間が 8 割を占めるともいわれる。

モデル検証までのプロセスを機械が自動処理する AutoML (Automated Machine Learning : 自動化された機械学習) という技術¹⁰が普及し始めている。機械学習モデルは人の手によって前処理、アルゴリズム選択、モデル検証して生成するのが一般的だが、AutoML を活用することによってモデル生成を自動化することが可能となる。例えば、ランダムフォレストやサポートベクターマシンなどのアルゴリズム選択についても、AutoML によって自動化が可能である。AutoML の登場によって、機械学習のアルゴリズムに関する高度な専門知識を持たない人であっても、業務知識さえあれば機械学習モデルを構築しやすい環境が整いつつある。

図表 12 : AutoML による機械学習モデル構築プロセスの自動化



(出所) 三菱 UFJ 信託銀行作成

AutoMLは、機械学習モデルの開発工数をできるだけ削減しながら、一定水準を満たす精度のモデルを作成したい場合に有効といえる。また、モデル構築作業の多くのプロセスを自動化することができるため、データ収集作業や、分析アイデア・仮説定義等にリソースを集中することができるというメリットがある。一方、アルゴリズム選択やハイパーパラメーター(アルゴリズムを使用する際に必要となる様々な設定値)のチューニングを自前で行い、高い精度を求めたい場合には不向きなのであろう。

(3) GUI ツールの登場

機械学習の民主化が進むもう一つの背景として、GUI ツールの登場が挙げられる。GUI とは、画面上でマウス操作できるインターフェースのことを指し、ここでいう GUI ツールは、マウス操作のみで機械学習モデルを作ることができるツールを意味する。例えば、米 DataRobot 社や米 Google 社などが、GUI ベースで機械学習モデルを構築できるサービスを提供している。これらのサービスでは前項で触れた AutoML も実行可能だ。このように、プログラミングを一切行うことなく、マウスによる直感的な操作のみで機械学習モデルの構築が可能となってきている。

弊社においても、ユーザーがプログラミング知識を持たなくても簡単なエクセル操作

¹⁰ 前処理を自動化の対象外としているライブラリも存在している等、自動化の対象範囲はライブラリの種類によって異なる点には留意が必要。

で機械学習モデルの構築が可能となる「多目的 AI ツール」を独自開発している。当該ツールの活用を通じて、ユーザーの機械学習リテラシー向上や、分析スキル向上を図っている。

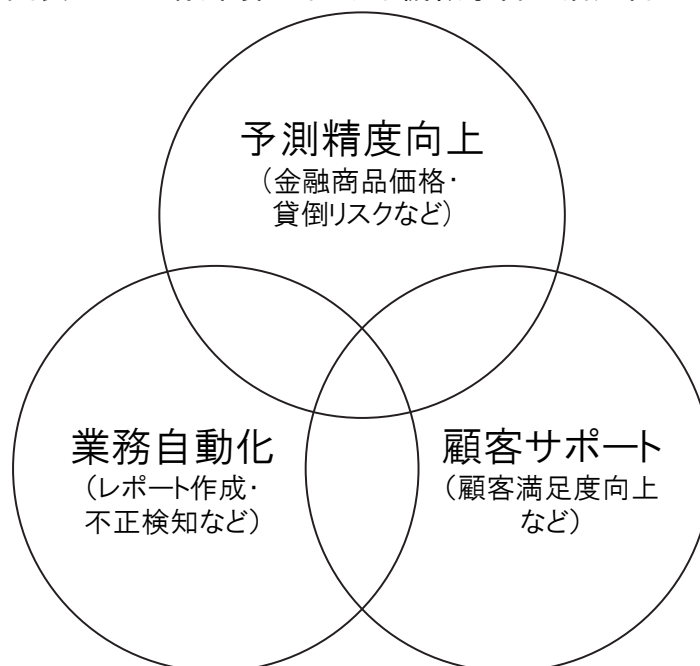
Ⅲ. 機械学習モデルの証券投資への活用

本章では、前章で説明した機械学習という手法が金融業界においてどのように活用されているのかについて紹介していく。はじめに、金融業界における機械学習の活用事例に関する一般的なトピックについて紹介する。次に、金融業界の中でも資産運用業務に焦点を当て、機械学習の証券投資への活用例として、Ⅱ章で解説した、教師あり学習と教師なし学習モデルの活用例をそれぞれ一つずつ取り上げる。

1. 金融業界における機械学習の活用事例

金融業界における機械学習の活用目的として、予測精度向上、業務自動化、顧客サポートという、大きく3つのカテゴリーに分類してみる(図表13)。

図表 13：金融業界における機械学習の活用目的



(出所)三菱UFJ信託銀行作成

予測精度向上の対象としては、金融商品価格の将来予測や、融資業務における貸倒リスクの予測などが挙げられる。財務データや非構造化¹¹データをコンピューターに大量に読み込ませ、人間では気付きにくい埋もれた法則性を見つけ出したり、ベテラン人材と同等またはそれ以上の予測精度を出すことが主な目的である。

業務自動化の事例では、レポートの自動作成や不正の自動検知などが挙げられる。具体的には、自然言語処理技術を活用した各種運用レポートの自動作成や決算短信の自動分析、異常検知技術を活用した不正取引の自動検出(証券会社であれば相場操縦、カード会社であればクレジットカードの不正使用等が該当)といった、人間であれば膨大な時間がかかる事務作業を機械化・自動化することによって生産性向上を図るのが主な目的である。

顧客サポートの事例では、ロボアドバイザーや顧客分析・マーケティングでの活用などが挙げられる。顧客属性(年代や家族構成、ローン残高など)や意向(運用や借入ニーズなど)をコンピューターに学習させ、個々に応じた最適な提案やサービスのカスタマイズ化を通じて、より一層顧客の満足度向上を図ることが目的である。

機械学習活用の大きなメリットは、ベテラン社員が長年の経験の中で培った知恵と判断力・提案力をモデル化することで、マニュアル化しにくいノウハウを社内に蓄積することが可能となる点である。また、大量のデータから様々な相関関係を把握して推論の確度を高めることで、人間の意思決定をサポートする点にもあるといえよう。

金融業界においても、コロナ渦をきっかけに DX(デジタル・トランスフォーメーション)を主軸としたビジネスモデルへの転換が加速している。各業務分野にてデジタルデータを活用し、競争力を確保することが各金融機関にとって重要なテーマとなっている。機械学習の活用フィールドは今後も拡大していくものと思われる。

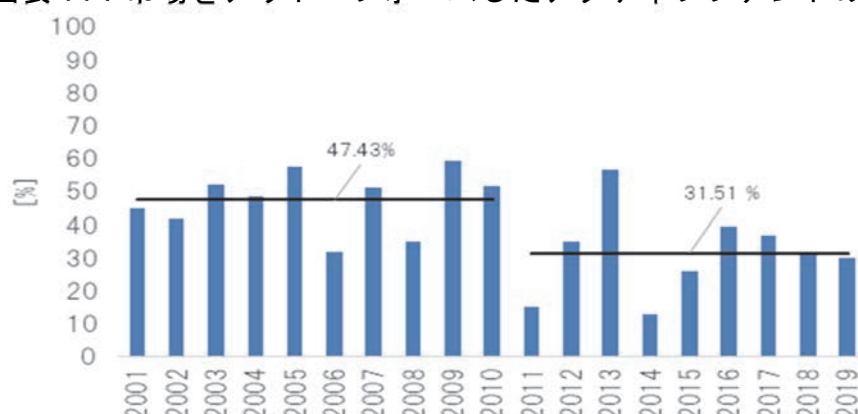
2. 証券投資における機械学習モデル活用の意義

ここからは、資産運用業務に焦点を当てる。近年、アクティブファンドが市場平均をアウトパフォームすることの難易度が高まっているといわれている。図表 14 は、米国籍の株式アクティブファンドを対象に、市場平均をアウトパフォームしたファンドの割合を年別に集計したものである。2001 年から 2010 年にかけては、平均で約半数のアクティブファンドが市場平均をアウトパフォームしていた一方、2011 年から 2019 年までの期間においてアウトパフォームしたファンドは平均で約 3 割程度にとどまっている。

背景にはどのような理由が考えられるだろうか。近年は上場企業に公平な情報公開を求めるフェア・ディスクロージャー・ルールの浸透や、インターネット普及による幅広い投資家層への情報伝達速度の向上などにより予測の差別化が難しくなっている。更にテクノロジー活用による運用の高度化を図る投資家が増加し遅延の少ないデータに基づいた投資行動が増えたことで、ミスプライスの減少及び修正速度が早まったことなどが理由と筆者は考える。

¹¹ 文書データ・電子メール・写真・動画など、定型的に扱えないデータの総称を指す。

図表 14：市場をアウトパフォームしたアクティブファンドの割合



(出所) S&P Global より三菱 UFJ 信託銀行作成

また、筆者の担当分野であるクオンツ運用という視点からは、こうした環境変化による要因に加え、方法論による要因も影響していると考えられる。伝統的なクオンツ運用で用いられるマルチファクターモデルをベースとしたアプローチは、観察データが線形性を有していることを前提としている。線形性とは、観察データ間が直線的な関係性で説明できる性質を意味する。

単純な例を用いて考えてみよう。株式市場の動きを「株価リターン＝景気指標＋金利」という2つの変数の動きだけで表現するモデルがあるとする。景気指標改善＝株価上昇、金利低下＝株価上昇など、各変数と株価リターンの関係がそれぞれ一対一の関係性で表現できることを前提としているのがマルチファクターモデルである。

しかし、近年金融市場の複雑性の高まりを背景に、金融市場の振る舞いを線形性で表現することがより難しくなっている。例えば、企業のグローバル活動の進展により、同じ市場に属する企業でも、売上やサプライチェーン構造の国内/海外依存度に大きな違いがあり、その依存関係も多様で複雑さを増してきている。同時に、国によっては制裁リスクを抱えている場合もある。産業構造に関しても、オールド・エコノミー業種、ニュー・エコノミー業種のビジネスモデルの違いなど、その他にも考慮すべき要素がいくつもある。企業活動や金融市場のグローバル化、産業構造の変化等を背景に、従来用いられてきた投資判断尺度では表現しきれない変数の種類とその相互作用により価格変動が起こっているともいえる。各変数と株価リターンの関係を一対一の関係性で表現することが難しくなっており、非線形性とはこうした複雑性を有する状態のことを指す。市場にはこうした非線形性が潜んでおり、市場の動きが複雑化する中で従来の古典的な回帰モデルでは捉えきれない要素が増してきているといえよう。

機械学習は非線形性を表現できるなどといった理由から、従来に比べて複雑性が増した市場を上手く表現できる可能性があり、機械学習モデル活用の有効性は高いと筆者は考える。アクティブのクオンツ運用ファンドにとって、スマート β ¹² ファンドとの運用差別化

¹² 従来の時価総額加重平均の株価指数ではなく、定量的なルールに基づいて構成銘柄を選定した株価指数のこと。配当やバリュウ、グロス、低ボラティリティ型などの指数がある。

や新たな超過収益の源泉獲得を追求するという観点で、機械学習活用への期待は高い。

3. 証券投資で活用するデータの種類

次に、機械学習モデルを証券投資で活用する場合に用いるデータの種類について説明する。データは主に、ファンダメンタル、マーケット、予想・評価、オルタナティブの4つに分類することができる(図表 15)。

図表 15：証券投資に使うデータの種類

ファンダメンタル データ	マーケット データ	予想・評価 データ	オルタナティブ データ
<ul style="list-style-type: none"> ・企業財務 ・マクロ経済指標 など 	<ul style="list-style-type: none"> ・金融商品価格 ・出来高 ・建玉 ・ボラティリティ など 	<ul style="list-style-type: none"> ・アナリスト推奨 ・コンセンサス予想 ・債券格付け ・バリュエーション など 	<ul style="list-style-type: none"> ・画像/動画/音声 ・テキスト ・POS データ など

(出所)各種資料より三菱UFJ信託銀行作成

- ・ファンダメンタルデータは、例えば企業財務データやマクロ経済指標などが挙げられる。月次や四半期、年度単位で更新される場合が多く、更新頻度が低いのが特徴である。また、市場関係者の注目度が高く、市場への織り込み速度が速いことから、データ単体として用いるだけでは差別化や超過収益獲得を目指す上での付加価値は小さいといえる(複数の指標・データと組み合わせて使うことで付加価値となり得る)。
 - ・マーケットデータは、主に取引所における取引活動データを指し、例えば株価や為替レート、出来高、建玉などが挙げられる。速報性の高さやデータ量が豊富である点が特徴である。
 - ・予想・評価データは、財務や市場データなどを元に加工されたデータを指し、アナリスト推奨やコンセンサス予想、格付けデータなどが挙げられる。証券会社や調査会社が付加価値情報として公表・販売していることが多く、投資のシグナルとして直接使用可能である点が特徴である。
 - ・オルタナティブデータは、POS、クレジットカードの利用履歴、経済ニュースの記事、位置情報、ウェブサイトのアクセス数、衛星画像、SNSの投稿、輸送用コンテナの領収書、商品レビューなどといったデータの総称を指す。投資家が伝統的に投資判断に活用してきた財務諸表や統計指標等といった公開情報と区別する意味でオルタナティブ(代替)と呼ばれている。
- データの多くが非構造化データであり、データ加工が困難である一方、こうした取り扱いの難しいデータほど、市場参加者が気づきにくい投資シグナルが潜んでいる

可能性があるともいえる。近年の機械学習や自然言語処理技術の進展を背景に、投資の付加価値を獲得するための新たな情報源として注目を集めている。

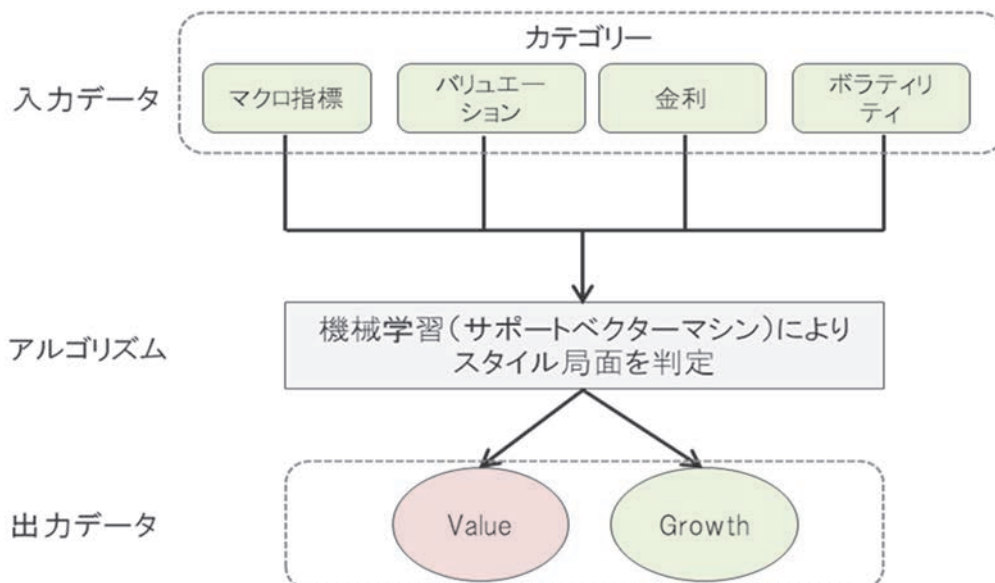
4. 教師あり学習の活用事例 ～株式市場における Value/Growth スタイル局面判断～

(1) モデル概要

ここでは、教師あり学習の機械学習モデル応用例として、米国株式市場における Value/Growth スタイル局面判断モデルについて紹介する。モデル概要は図表 16 のとおりである。モデルのアルゴリズムには、線形回帰で説明のつかない部分を表現するため、教師あり学習の一種であるサポートベクターマシンを採用している。サポートベクターマシンの特徴として、過学習¹³を起こしにくい点、比較的少ないデータであっても精度の高い予測が可能である点、2種類に分類する問題を解くのに適している点などが挙げられる。

図表 16 : Value/Growth スタイル局面判断モデルの概要

目的	・米国株式市場における、将来 1 カ月の Value/Growth スタイル優劣を予測
入力データ	・マクロ経済指標・バリュエーション・金利・ボラティリティ等の観点から採用した約 30 個の指標
出力データ	・スタイル指数 (Value 指数及び Growth 指数) の月間リターンの優劣を判定 ・ここではスタイル指数として、米国の Russell 1000 Value 指数と Russell 1000 Growth 指数を採用



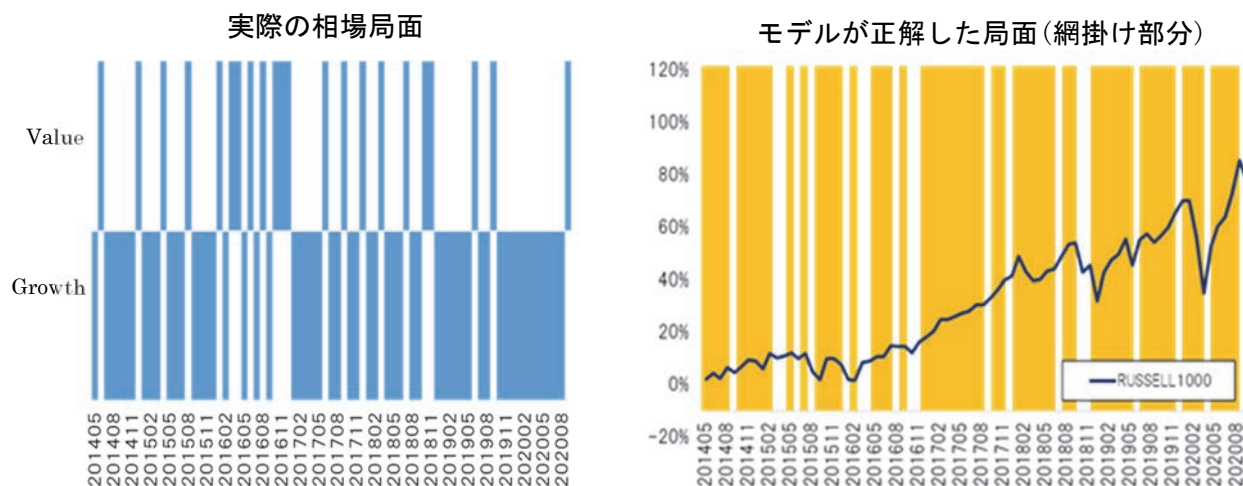
(出所)三菱 UFJ 信託銀行作成

¹³ 学習した既知データには適合する一方、その説明力を高め過ぎる(学習し過ぎる)と新たなデータでの予測精度がかえって低いモデルが出来てしまうこと。オーバーフィッティングともいう。

(2) モデルの判定結果

実際に市場で起きた Value/Growth 相場局面とモデルの事前予測が一致した局面を図表 17 に示す。市場の動きが比較的安定していた 2017 年を中心に、分析対象期間 78 ヶ月中 57 ヶ月で実際の局面と予測が一致した。一方、英国の EU 離脱を問う国民投票や米大統領選(トランプ氏が当選)が実施された 2016 年、米中貿易摩擦による米中間の政治的対立が強まった 2018 年など、政治的なイベントや突発的なヘッドラインを受けて相場が変動した局面では正解率が低下する傾向が確認された。本モデルは、図表 15 で示したオルタナティブデータを入力データとして使用していないが、例えばオルタナティブデータの一つであるニュース記事を入力情報としてモデルに取り込むことによって、予測力の向上が更に高められる可能性があると考えられる。

図表 17：モデルの判定結果

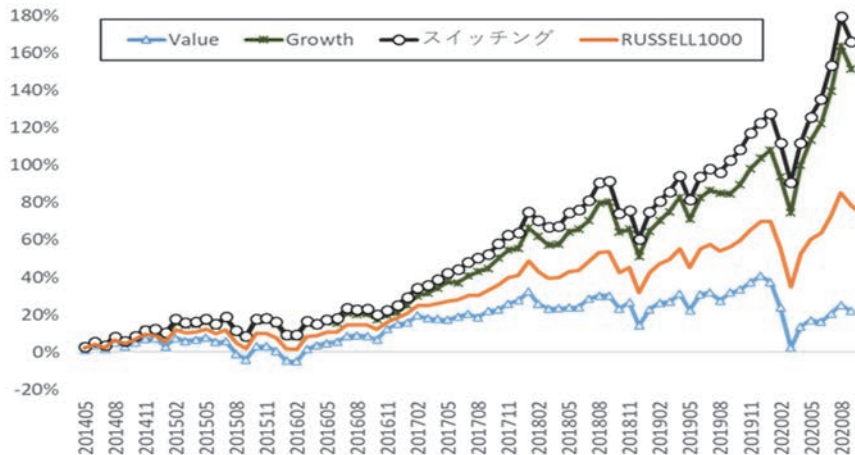


(出所)三菱 UFJ 信託銀行作成

(3) 投資戦略への適用例

本モデルの投資戦略への適用例の一つとして、スタイルのスイッチング戦略への活用が挙げられる。図表 18 に、Value/Growth のスタイルをスイッチングする投資戦略を実行した場合のバックテスト結果(アウトオブサンプル)を示す。月末時点で得られたモデルの判定結果に基づき、モデルが Value と判定した時には Value 指数を翌 1 ヶ月保有、Growth と判定した時には Growth 指数を翌 1 ヶ月保有するというシンプルな戦略ではあるが、図表のとおり、スイッチング戦略は Value 指数及び Growth 指数それぞれに対してアウトパフォームしており、その有効性が確認できる。

図表 18 : スイッチング戦略活用後の投資パフォーマンス



(出所)三菱 UFJ 信託銀行作成

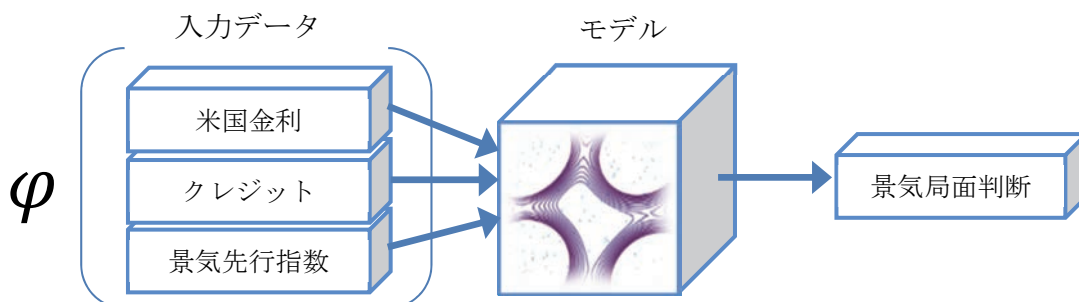
本モデルの適用例として、スイッチング戦略の他にも、マルチファクター運用でのティルト・ファクターの選択やティルト度合を判断する材料としての活用等が考えられる。

5. 教師なし学習の活用事例 ～景気局面判断と投資戦略～

(1) モデル概要

ここでは、教師なし学習の機械学習モデル応用例として、マクロ経済の景気局面判断モデルについて紹介する。モデルの目的は、景気が「良い・悪い」といった景気・経済状況をモデルに識別させることで、運用者が投資判断を行う際のアシスト材料となる指標を作成することである。モデルのアルゴリズムには、教師なし学習の一種である混合ガウスモデルを採用し、入力データには、米国金利、クレジット(CDS)、景気先行指数を採用している。モデルの概要は図表 19 のとおりである。

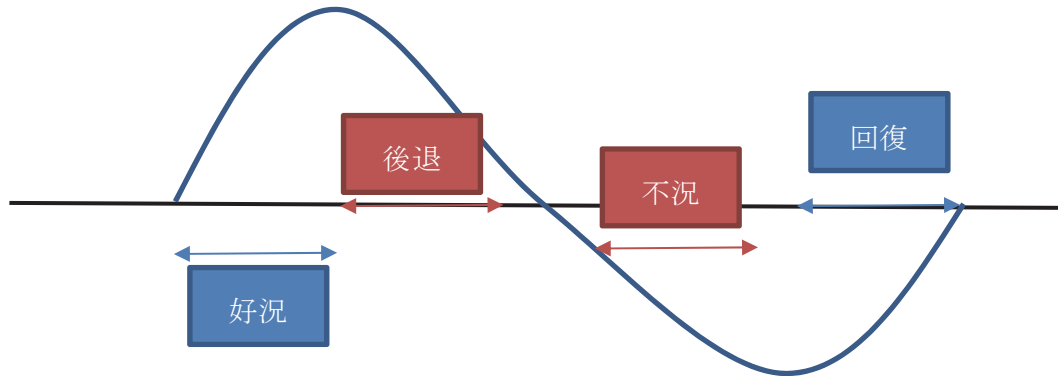
図表 19 : 景気局面判断モデルの概要



(出所)三菱 UFJ 信託銀行作成

混合ガウスモデルは、ソフトクラスタリングの手法の一つで、モデルで識別した結果を確率的に表現できる点が特徴である。例えば、景気の良し悪しの判定をする場合、景気の識別結果に加え、景気が良い確率 80%、悪い確率 20%のような確率を得ることができる。モデルを学習させる際に、識別させる景気の局面数を与える必要があるが、ここでは好況、後退、不況、回復、の4局面を設定した(図表 20)。

図表 20 : 学習させる景気局面のイメージ

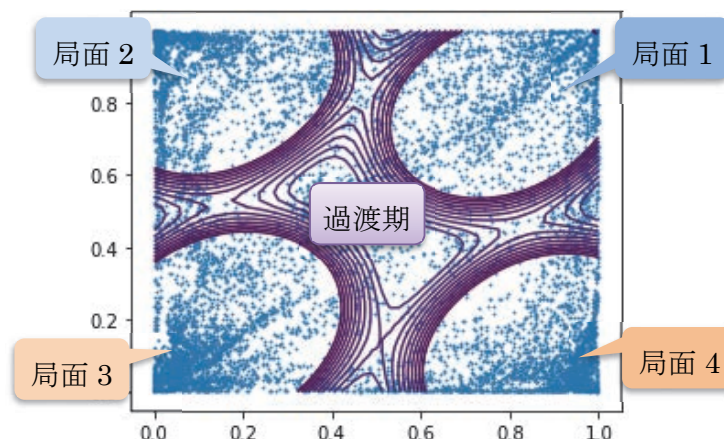


(出所)三菱 UFJ 信託銀行作成

(2) モデルの学習結果

混合ガウスモデルの出力結果は、日次の景気識別結果と確率(例：局面 1 の確率 80%、局面 2 の確率 20%)である。ここでは、ソフトクラスタリングの特徴を活かして、ある局面の確率が 60%以下であれば過渡期(局面が移り変わる局面)と定義した。出力される確率をモデルの景気局面判断に対する自信度合いと見なすことで、機械学習モデルの局面判断に柔軟性(人が景気を判断する場合も、判断に迷うような場面はある)を持たせた。混合ガウスモデルの出力結果を可視化したものが図表 21 である。図の各点は日次の局面判断データをプロットしたものであり、景気局面が大きく 4 つのグループにクラスタリングされていることがわかる。

図表 21 : 混合ガウスモデルによる景気局面判断の結果



(出所)三菱 UFJ 信託銀行作成 モデルの学習は日次ローリング

モデルの出力結果を、入力データなどをもとに解釈すると図表 22 のように整理することができる。

図表 22：景気局面判断モデルの解釈結果

局面	金利	景気	クレジット	マーケットシナリオ	マーケットの状況
1	Bear Steep	回復	Risk On	株価上昇	・投資家心理の回復 ・企業業績の持ち直し
2	Bear Flat	好況	Risk On	株価上昇	・企業業績が好調 ・景気敏感銘柄などが選好
3	Bull Flat	後退	Risk Off	株価下落	・企業業績の伸び鈍化 ・業績の先行き不透明感
4	Bull Steep	不況	Risk Off	株価下落	・投資家心理の悪化 ・業績の先行き不安
過渡期	Neutral	Neutral	Neutral	局面の移り変わり (確率的にモデルが安定しない局面)	

(出所)三菱 UFJ 信託銀行作成

次に、モデルによる景気局面判断が景気をうまく捉えられているのか、一般的な景気指標と比較し定量的に評価する。景気指標には内閣府が月次で公表している景気動向指数 CI (Composite Index) を用いた。景気動向指数とは、景気の循環的な局面を把握することを目的として作られた指数で、景気変動の大きさやテンポ(量感)を把握することができ、景気に対し先行して動く先行指数、ほぼ一致して動く一致指数、遅れて動く遅行指数の3種類がある。ここでは CI 先行指数・一致指数の前月比変化(プラスであれば景気拡大、マイナスであれば景気後退と定義)を、モデルによる景気局面判断毎に集計し比較を行った。具体的には、モデルが判断した、景気の良い局面(局面1・2)と景気が悪い局面(局面3・4)それぞれについて、CI の変化との関係性をまとめたものが図表 23 である。なお、CI はデータ公表日と基準日との間に1ヵ月以上のタイムラグが存在するが、基準日の CI を用いて、モデルによる景気局面判断結果と比較を実施した。

図表 23：景気局面判断の評価

景気局面判断	CI(先行指数)	CI(一致指数)
景気良い (局面1・2)	0.42	0.42
景気悪い (局面3・4)	▲0.34	▲0.39

(出所)内閣府データなどから、三菱 UFJ 信託銀行作成

景気が良いと判断した期間で CI(先行指数)の平均変化は+0.42 ポイント、景気が悪いと判断した期間の CI(先行指数)の平均変化は▲0.34 ポイントであることから、景気が良いと局面判断した期間には CI の変化もプラス、悪い時にはマイナス傾向となって

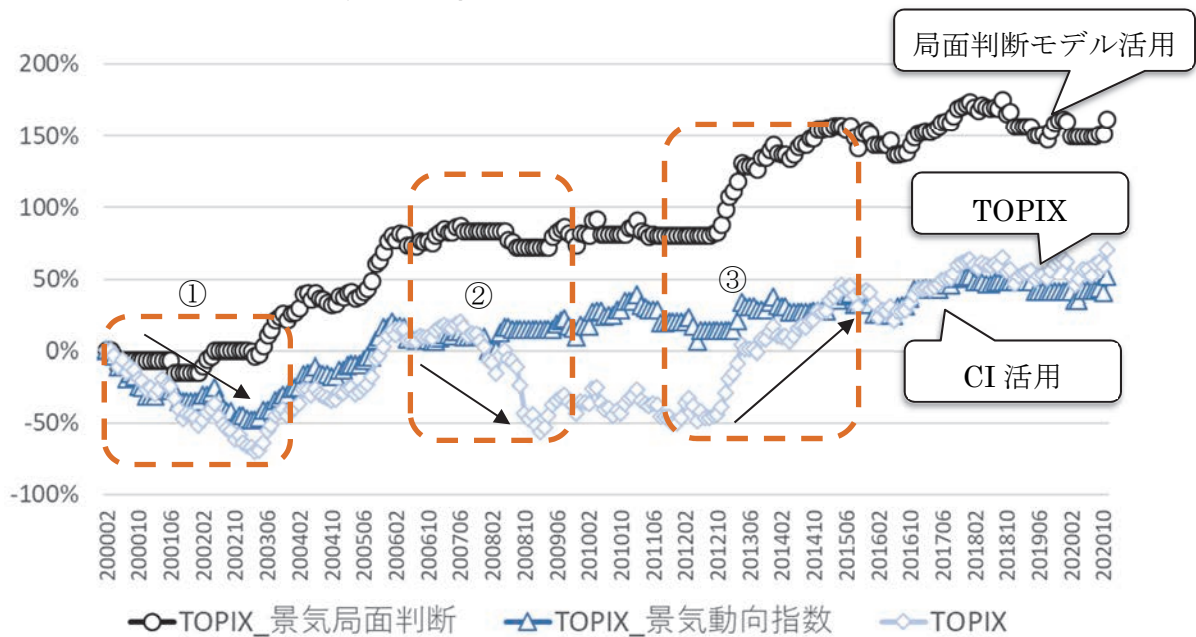
いることが確認できる。CI(一致指数)においても同様の傾向がみられ、モデルによる景気局面判断が一般的な景気指標と類似した情報を得られていることが確認できる。

ここで景気局面判断モデルのポイントをまとめると、①一般的な景気指標と類似した情報をつかめている点、②景気指標の発表日を待たずにリアルタイムに景気動向を捉えることができる点、である。次節では景気局面判断モデルを活用した投資戦略について紹介する。

(3) 投資戦略への適用例

景気局面判断モデルを株式投資へ応用する事例を簡単に紹介する。景気が良くなれば株価は上昇し、景気が悪くなると株価は下落するという投資アイデアのもと、景気が良い(回復・好況)と判断されたときのみ TOPIX(東証株価指数)に投資し、景気が悪い(悪化・不況)と判断されたときには現金で保有という戦略の有効性を検証する。検証では、景気局面判断モデルを活用した場合、景気動向指数 CI を活用した場合、TOPIX を保有し続けた場合、の3つの比較を行う。シミュレーションは、月末時点で得られる情報に基づき翌1カ月の TOPIX に投資した場合の収益率を計測した(図表 24)。

図表 24 : 景気局面判断と株式投資



(出所)三菱 UFJ 信託銀行作成

景気局面判断モデルを活用した場合は、TOPIX を保有し続けた場合と比較して、景気が悪化し TOPIX が大きく下落するような局面(①、②)において、パフォーマンスの下落幅を抑制できていることがわかる。また景気動向指数 CI を活用した場合と比較すると、TOPIX 上昇時(③)において、上昇トレンドを早い段階で捕捉できていることがわかる。これはモデルが日次で景気局面判断を行っているため、月次で公表されている景気動向指数よりも、いち早く景気の変化を捉えられていることが要因と推察される。

ここでは景気局面判断モデルを活用して TOPIX へ投資する事例を簡単に紹介したが、その他の応用事例としては、景気の局面に応じて魅力的なファクターに投資するファクターローテーション戦略や、ポートフォリオの最適化(リスクの管理)等、様々な用途への応用が期待される。

IV. 終わりに

本稿では、金融業界における機械学習の活用事例や、機械学習モデル構築者の裾野拡大に関する一般的なトピックについて触れた後、証券投資分野を対象とした機械学習の具体的活用事例について紹介してきた。本稿では紙面の関係上紹介できなかったが、Ⅲ章で紹介した、スタイル局面判断や景気局面判断の他にも、企業収益予測、市場センチメント判断などの領域においても機械学習を活用しており、今後も応用範囲を広げていく予定である。証券投資分野における機械学習の活用範囲はアイデア次第で今後も広がりを見せていくと思われる。

本稿で述べたとおり、機械学習の民主化が進む中、今後、機械学習を証券投資分析に活用していくにあたって実務担当者に求められる能力とはどういったものであろうか。それは、インプットすべきデータとそうでないものを見極め、取捨選択する、ファンドマネージャーであれば古くから求められている能力に他ならない。つまり、投資戦略としての超過収益の源泉を明確にした上で、戦略を構築する力である。金融市場は、世界中の政治・経済・天災などの影響を受けることから、考慮すべき要素が膨大である一方、世の中のあらゆるデータを全てインプットすることは現実的には困難である。また、投資としてのシグナルよりもノイズの方が溢れかえっているという側面も多いことから、人間が情報の重要性を判断し、入力データを決定する必要がある。ここは当面機械任せにはできない領域であろう。

機械学習モデルは万能薬ではないことにも留意しておく必要がある。市場特性や人々の行動様式に変化が生じれば、市場が過去と同じような振る舞いをしなくなることから、永続的に有効性を発揮し続けるモデルを構築するのは困難といえる。アフターコロナの社会構造変化により、これまで有効であった投資戦略が今後は有効に機能しなくなることも考えられる。そのため、モデルは一度構築すれば終わりというものではなく、環境変化に応じて使用するアルゴリズムの変更や、入力データの入れ替え、学習頻度の変更等を要することになる。こうしたプロセスは人間が対応する領域となるが、その対応範囲が極力狭くて済むよう頑強性の高いモデルを生成するのが実務担当者の腕の見せ所であるともいえよう。

最後に、金融市場におけるデータ分析・活用に関するアプローチ手法の一つである機械学習について、本稿が読者の理解に少しでも貢献できたら幸いである。

(2020年12月22日 記)

※本稿中で述べた意見、考察等は、筆者の個人的な見解であり、筆者が所属する組織の公式見解ではない

【参考文献】

- ・『資産運用とAI ～ 運用者はAI に代わられてしまうのか？(前編)(資産運用情報2019年10月号)』, 三菱UFJ信託銀行, [2019]
- ・『資産運用とAI ～ 運用者はAI に代わられてしまうのか？(後編)(資産運用情報2019年11月号)』, 三菱UFJ信託銀行, [2019]
- ・『ディープラーニング活用の教科書 実践編』, 日経クロストrend, [2019]
- ・『データサイエンティスト養成読本 機械学習入門編』, 技術評論社, [2015]
- ・『人工知能は人間を超えるかディープラーニングの先にあるもの』, 松尾豊, KADOKAWA, [2015]
- ・『Rによる機械学習』, Brett Lantz, 翔泳社, [2017]
- ・『Advances in Financial Machine Learning』, Marcos López de Prado, Wiley, [2018]
- ・『情報通信白書』, 総務省, <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/r01/html/nd113210.html>(最終閲覧日: 2020年11月27日)
- ・『景気動向指数の利用の手引』, 内閣府, <https://www.esri.cao.go.jp/jp/stat/di/di3.html>(最終閲覧日: 2020年12月16日)

本資料について

- 本資料は、お客さまに対する情報提供のみを目的としたものであり、弊社が特定の有価証券・取引や運用商品を推奨するものではありません。
- ここに記載されているデータ、意見等は弊社が公に入手可能な情報に基づき作成したのですが、その正確性、完全性、情報や意見の妥当性を保証するものではなく、また、当該データ、意見等を使用した結果についてもなんら保証するものではありません。
- 本資料に記載している見解等は本資料作成時における判断であり、経済環境の変化や相場変動、制度や税制等の変更によって予告なしに内容が変更されることがありますので、予めご了承下さい。
- 弊社はいかなる場合においても、本資料を提供した投資家ならびに直接間接を問わず本資料を当該投資家から受け取った第三者に対し、あらゆる直接的、特別な、または間接的な損害等について、賠償責任を負うものではなく、投資家の弊社に対する損害賠償請求権は明示的に放棄されていることを前提とします。
- 本資料の著作権は三菱UFJ信託銀行に属し、その目的を問わず無断で引用または複製することを禁じます。
- 本資料で紹介・引用している金融商品等につき弊社にてご投資いただく際には、各商品等に所定の手数料や諸経費等をご負担いただく場合があります。また、各商品等には相場変動等による損失を生じる恐れや解約に制限がある場合があります。なお、商品毎に手数料等およびリスクは異なりますので、当該商品の契約締結前交付書面や目論見書またはお客さま向け資料をよくお読み下さい。

編集発行：三菱UFJ信託銀行株式会社 アセットマネジメント事業部
東京都千代田区丸の内1丁目4番5号 Tel. 03-3212-1211（代表）