

# 資産運用と AI ～ 運用者は AI に代わられてしまうのか？ (後編)

## 目次

### I. はじめに (前編)

### II. AI ファンドの現状 (前編)

1. AI ファンドの運用状況
2. AI はどのようにマーケットを学習するのか？
3. 景気後退のシグナルとしての米国イールドカーブ
4. 米国のイールドカーブの変化が AI の投資判断となった可能性

### III. 資産運用において AI に立ちはだかる壁 (前編)

1. 実はマーケット予測に AI は向いていない？
2. 株価は本当にランダムウォークか？
3. ランダムウォークするマーケットを AI で予測するには

### IV. マーケット AI 予測モデルの実装 (後編)

1. AI に意味のないデータを入れる時代は終焉する
2. その学習データは未来予測できるだけの情報をもっているのか？
3. 為替の AI 予測モデルの実装

### V. 今後の展望 (後編)

資産運用部 債券運用第2課 兼 先端金融工学運用室  
シニアファンドマネージャー 樋口 裕之

## 【序章】

前編(10月号)では、AI が過去のマーケットデータについてどのように学習するのか、そして、実際の資産運用において AI を構成する機械学習アルゴリズムが直面する課題として、マーケットであるがゆえに“過去データの再現性の低さ”や“マーケットデータに含まれる意味のないトレンド”の問題をどのようにクリアするのか、について述べ、これらの壁をどのようにして乗り越えたのかというモデル開発者の巧拙が運用パフォーマンスに大きく影響することを説明した。

そして、これからの資産運用実務者は、これらのノウハウを AI モデルの開発という領域に留めず、今やマーケットを動かす原動力の一つとして、通常の投資判断においても AI モデルの投資行動を考慮した上でマーケットを見通すスキルが求められている、ということを述べた。

本号では、実際に複数の機械学習アルゴリズムを用いてマーケット AI 予測モデルを構築する過程について紹介する。

## IV. マーケット AI 予測モデルの実装

本章では、実際に筆者が開発した AI モデルをベースにモデルの技術的な構築方法と注意点等について解説することで、よりリアルに資産運用における AI の世界を感じていただきたい。尚、本章以降の分析は筆者の専門である為替・金利系のデータを用いて行う。

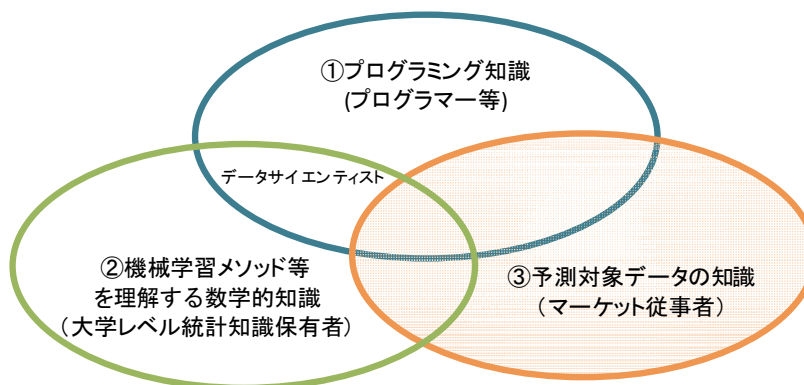
### 1. AI に意味のないデータを入れる時代は終焉する

数年前まではモデルを開発する際、「AI がすべて判断してくれるので、関係性がありそうなデータを何でも入れるべきだ」と主張する人も多かったと思う。この主張は画像・音声については正しいのかもしれない。なぜなら、画像・音声データには多少ゴミデータが入っていても、再現性の高いデータが大量にあるため、ゴミを識別しやすいからである。

また近年、関係がありそうなデータを大量に投入して AI モデルをチューニング<sup>1</sup>する方法が主流となった背景には、図表 1 の①プログラミング知識、②機械学習メソッド等を理解する数学的知識、③予測対象データの知識 を同時に持つ人材が圧倒的に少なかったことによるものであろう。①、②を持ったデータサイエンティストは一定数存在しており、彼らは画像や音声分野などの分野で成功体験を積んでいたため、マーケット知識はないが、関係がありそうなデータを大量に投入し、AI モデルをチューニングして予測結果を出しました、というのが実情かと思う。また、巷の AI の本には①について書かれている AI ハウツー本が多くある。ただ、モデルの評価を理解するには統計の知識も必要であるし、機械学習の性質に応じた学習データの加工方法など、機械学習の構造を理解する数学力も必要であろう。その意味では①、②、③の知識を同時に持つ人材を組織が意識的に育てる必要がある。そのような人材が育てば、AI に意味のないデータを入れる時代は終焉すると思われる。ただし、③のマーケット知識は一朝一夕には身につかず、最低でも3～5年程度は必要であろう。そうであれば寧ろ、マーケット知識を有するファンドマネージャーにデータサイエンスの知識を教育することが一番の近道と考える。

<sup>1</sup> AI を構成する機械学習アルゴリズムのパラメータを調節して学習データ内における適合度を上げる方法。これをあまりやりすぎると、過学習になりやすくなる。

図表 1 : 資産運用の AI モデル開発に必要な知識



出所：三菱UFJ信託銀行作成

## 2. その学習データは未来予測できるだけの情報をもっているのか？

既述のとおり、マーケットデータの予測には、未来予測力のある純度の高い学習データの使用が望ましい。ただし、機械学習アルゴリズムの多くはブラックボックスに近く、学習データが未来予測できるだけの情報をもっているか否かの判別は難しい。そこで、ここでは計量分析で予測対象と関係の高いデータを絞り込み、tSNE<sup>2</sup>と呼ばれる次元圧縮手法を用いて 10 種類のデータを 2 次元と 3 次元データに圧縮して観察してみよう。それにより再現性が高く、純度の高いデータとはどのようなものか視覚的にご覧頂こうと思う。

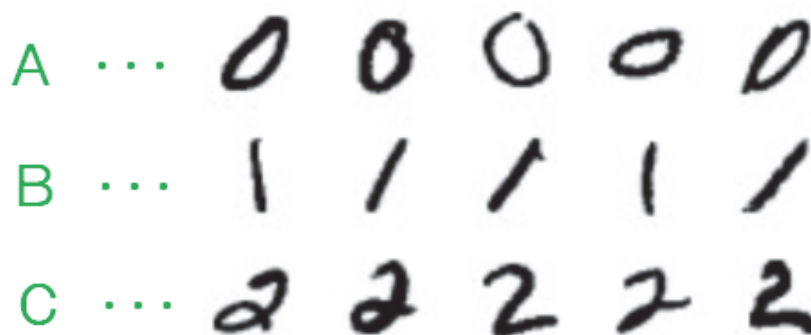
図表 2 は、ある  $t$  時点のデータ(日米金利差や為替の変動など加工したデータ 10 種類)に 1 日先 ( $t + 1$  時点)のドル円のヘッジコストが上昇：○、下落：▲、横ばい：×のラベルを付け、 $t$  時点データ 10 種類を 2 次元(Y1、Y2)と 3 次元(Y1、Y2、Y3)に圧縮し、グラフ化したものである。

2 次元データの場合は左下、3 次元データの場合は下側に、×のラベル、つまり横ばいの比率を多く観察することができる。これは、 $t$  時点のデータ 10 種類が  $t + 1$  時点のドル円ヘッジコスト横ばいを予測できるだけの情報を持っている可能性を示唆している。ちなみに、図中の黒丸(●)が今回 AI で予想しようとしているデータである。黒丸が×の近辺にいるので横ばいの可能性が高いことがデータ構造からも見て取れる。

<sup>2</sup> t 分布型確率的近傍埋め込み法 (T-distributed Stochastic Neighbor Embedding) 高次元のデータ集合を 2 次元または 3 次元へ配置する際に、高い確率で類似した集合が近傍に、異なる集合が遠方となるように表現する機械学習アルゴリズム。確率分布に基づいて分別しているため、非線形次元削減(圧縮)手法と呼ばれる。一般的な次元削減(圧縮)といえば主成分分析 (PCA) をよく用いるが、同分析は非線形データではうまく結果を得ることができない。特にマーケットデータは非線形データが多いので筆者はこの tSNE を使用している。

もちろん、すべてこの方法で視覚化できる訳ではないが、再現性の高いものに関しては図表のように抽出することができることが多い。

ちなみに画像データを tSNE で次元圧縮するとどうなるだろうか？ ピクセル化された手書き画像データ<sup>3</sup>(1ピクセルが白：0～黒：256の数字で色が表現され、それら784ピクセルで1画像データを構成)を1,000枚用意し、以下のように“A”：0、“B”：1、“C”：2というラベルを付ける。このピクセル化された数字を次元圧縮した結果が次々ページの図表3であり、“A”、“B”、“C”が非常に明確に区別されているのが分かる。これが画像データとマーケットデータの再現性の違いである。これだけデータの質が違うのである！

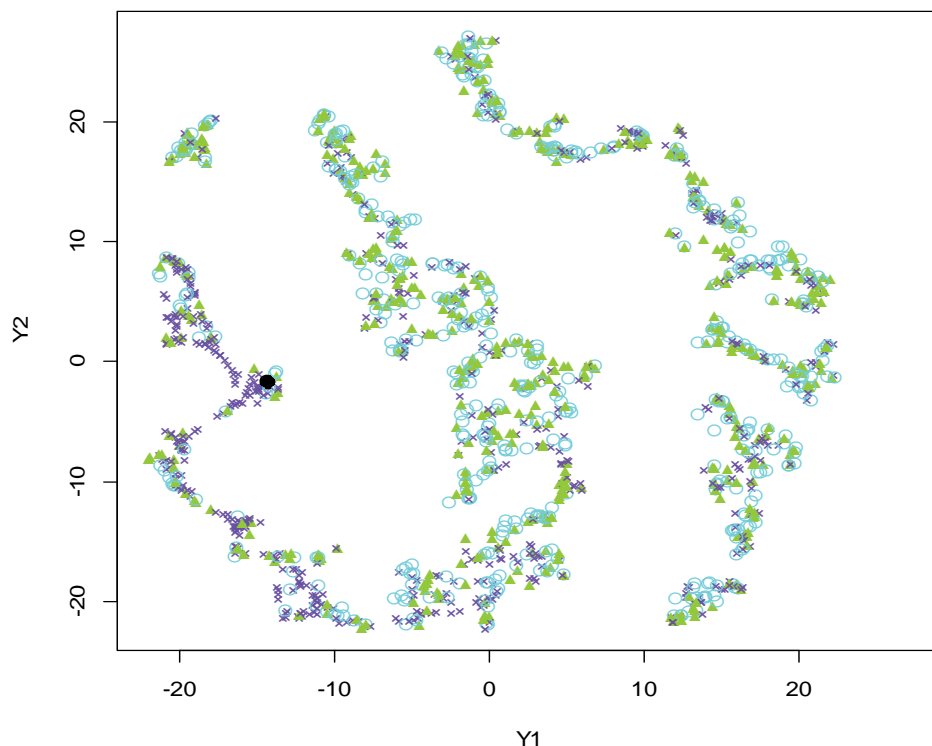


出所：Mixed National Institute of Standards and Technology database

<sup>3</sup> MNIST データと呼ばれ、米国の高校生や合衆国国勢調査局の職員が手書きした 70,000 枚の数字画像データであり、機械学習の分析をする際によく用いられるデータセットである。本分析ではそのうち 1,000 枚を無作為に抽出している。出所は MNIST database (Mixed National Institute of Standards and Technology database)。  
<http://yann.lecun.com/exdb/mnist/>

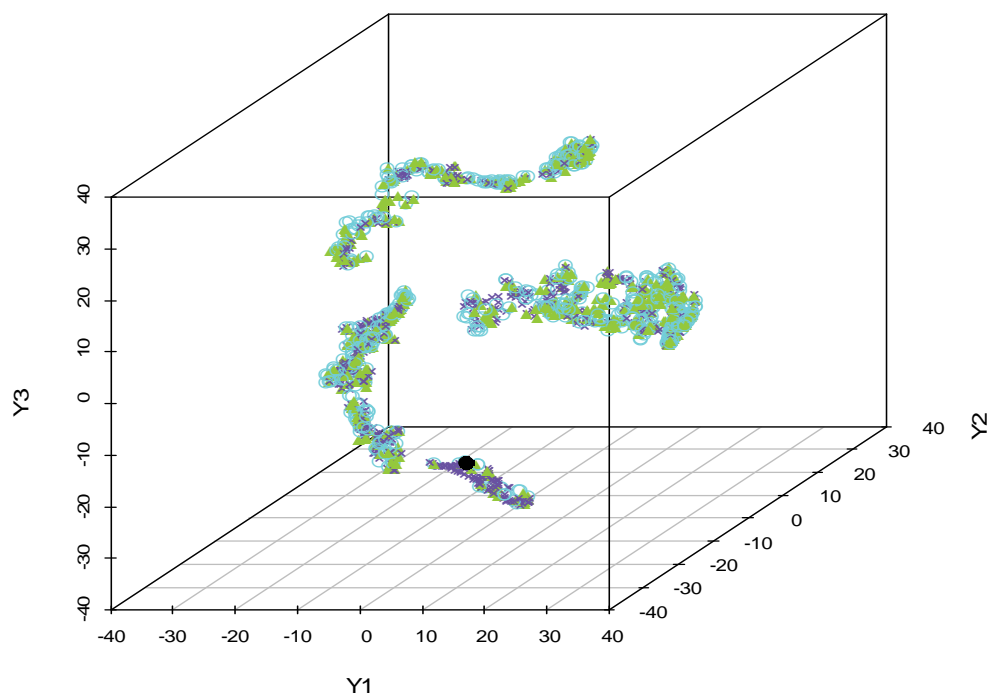
図表 2 : tSNE によるドル円ヘッジコスト次元圧縮結果(上段: 2次元 下段: 3次元)

tSNE 法による次元圧縮(2次元)



出所 : Bloomberg データより三菱 UFJ 信託銀行作成

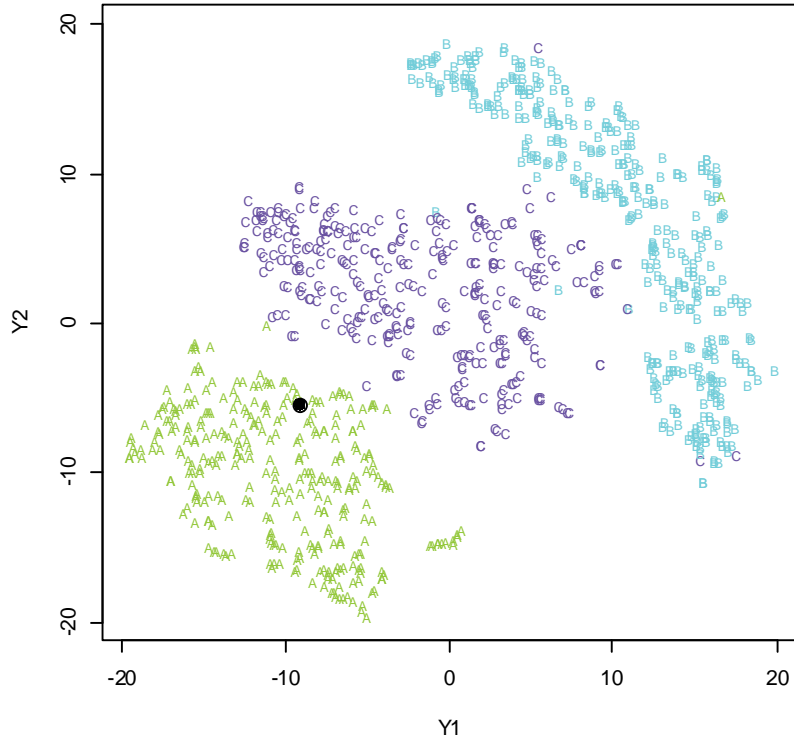
tSNE 法による次元圧縮(3次元)



出所 : Bloomberg データより三菱 UFJ 信託銀行作成

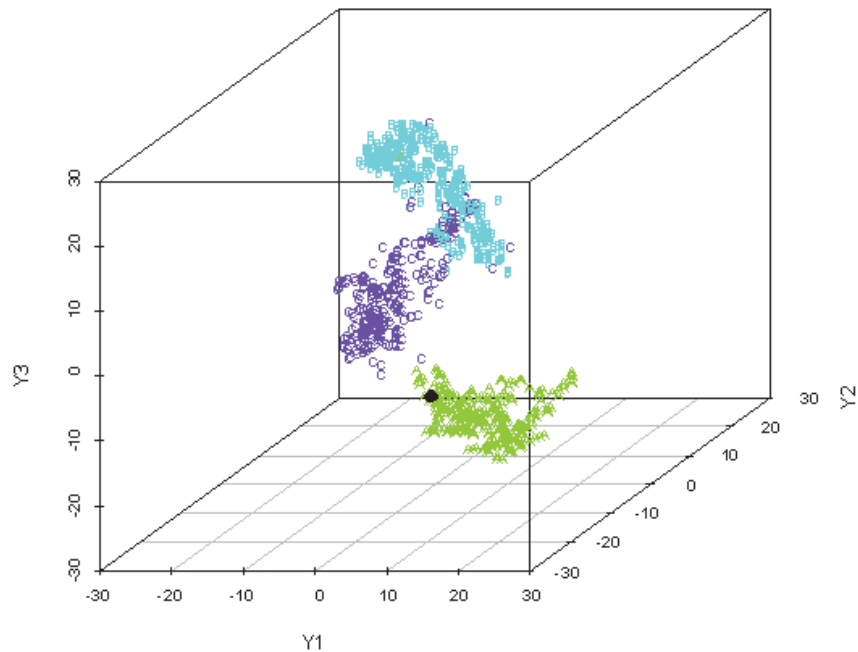
図表 3 : tSNE による画像データ次元圧縮結果(上段: 2次元 下段: 3次元)

tSNE 法による次元圧縮(2次元)



出所 : Bloomberg データより三菱 UFJ 信託銀行作成

tSNE 法による次元圧縮(3次元)



出所 : Bloomberg データより三菱 UFJ 信託銀行作成

### 3. 為替の AI 予測モデルの実装

#### (1) マーケット予測に用いる主な機械学習アルゴリズム

世の中には多くの機械学習アルゴリズムがあるが、時系列予測で使用するものとして下記リストのモデルを目にする機会が多い。機械学習の定義をインターネットで調べると線形回帰や状態空間モデル<sup>4</sup>なども入っているようだが、実務においてこれらのみを用いたモデルを AI と称するにはやや無理があるように思える。これらの手法は学習データの選別や加工など、以下の機械学習アルゴリズムと併用して活用されるケースが多いように感じる。なお、今回実装に利用する以外の機械学習アルゴリズムの内容については書籍やインターネット上に多く公開されているので、詳細は他に譲ることとする。

##### 【カーネル法】

サポートベクターマシン (SVM)

##### 【ツリーモデル、アンサンブル学習】

ランダムフォレスト

Gradient Boosting

LightGBM

XGBoost

##### 【ニューラルネットワーク】

ディープラーニング

畳み込みニューラルネットワーク (CNN)

再起型ニューラルネットワーク (RNN)

Long short-term memory (LSTM)

#### (2) ユーロドル<sup>5</sup> AI 予測モデルの構築

筆者は AI モデルを開発する際、まずは計量分析手法で学習データの選別を行い、その後比較的ベーシックな機械学習アルゴリズムであるサポートベクターマシン (SVM)、ランダムフォレスト、ディープラーニングで簡易予測を行う (これら 3 つのモデルの説明については巻末の付録参照)。これらのアルゴリズムは全く構造が異なるため、同じ誤りを犯しづらい。また、学習データの形状 (データ間のスケールの差、総データ数など) によって、それぞれ得意分野が異なる。同じ為替データでもディープラーニングが良い場合もあればランダム

<sup>4</sup> 必ずしも観測できるとは限らない状態を表す変数と実際に観測できる変数が異なるような系を数式で表現したモデルで工学分野において広く活用されている。

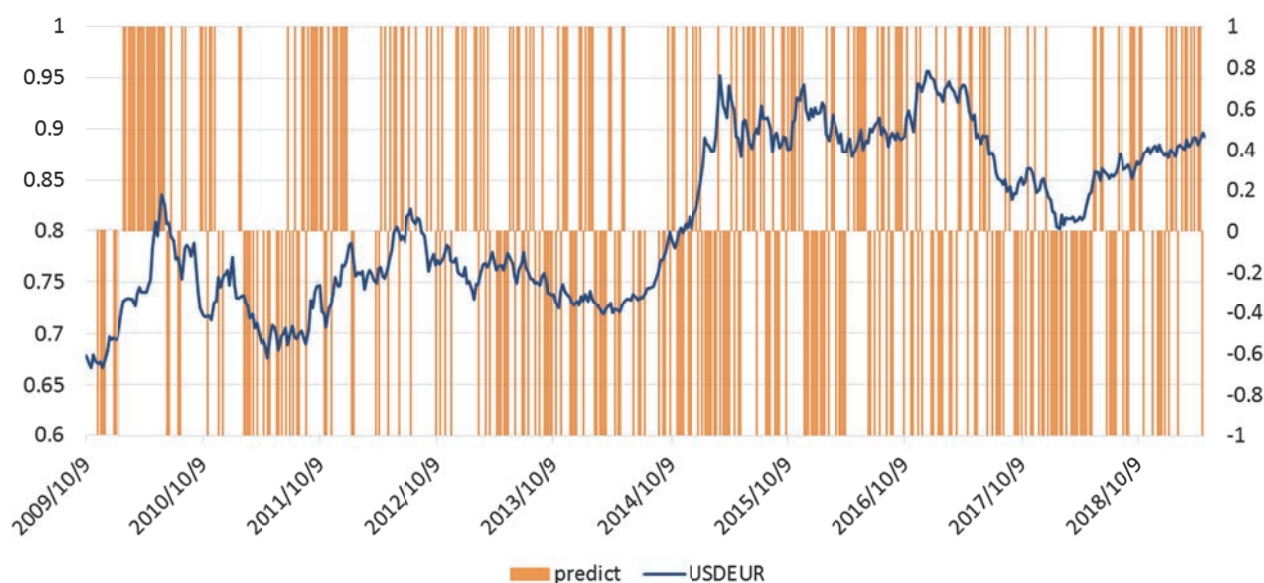
<sup>5</sup> 実際のモデルはドルユーロで表示。



フォレストの方が良い場合もある。また、予測結果などから機械学習アルゴリズムに学習データの情報をより適正に伝えられるよう、データ加工等も行う。この辺りの感覚は多くのマーケットデータを分析する過程で身に付くものであり、本などに書きづらい AI 技術者のノウハウといえる。

さて今回についても、まず既述の3種類のモデルで予測を実施<sup>6</sup>し、その結果を図表4～6で示す。その中で一番引っ掛かりが良かったモデルが図表4のランダムフォレストモデルである。“引っ掛かりが良い”とは、決して精度が高いという意味ではなく、自己の相場観において、ポイントとなる状況変化の局面で AI がしっかり投資判断ができているかという点である。例えば、シニアファンドマネージャーがジュニアファンドマネージャーの力量をみる際の目線にかなり近いと思う。以下の3つの図表は、ドルユーロの推移について、AI モデルの予測結果(predict の棒グラフが上に出ていれば上昇、下に出ていれば下落、空白なら横ばいを予測)と実績(折れ線グラフ)を比較したものである。

図表4：ランダムフォレストによるドルユーロの AI 予測



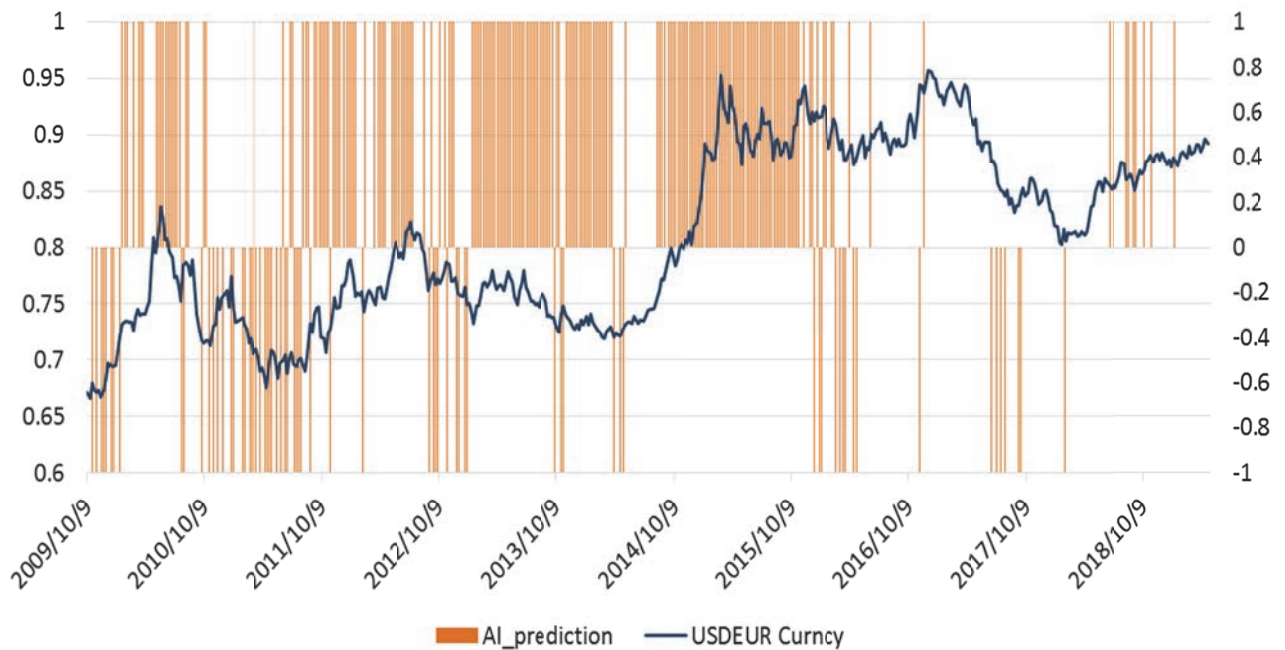
出所：Bloomberg データより三菱UFJ信託銀行作成

一方、図表5のディープラーニングでは2016年以降、相場が大きく動いているにも関わらず殆ど投資行動を取っていないことが分かる。また、図表6のSVMについても、下落中心の予測になってしまっていて、変化の局面に対しうまく反応できていない。

<sup>6</sup> 1990年1月～予測対象の一つ手前のデータまでを学習データとし、未学習データの翌週の方向性を予測。同手法で2009年10月から逐次推定を実施したもの。すなわちアウトオブサンプルの予測を行っている。

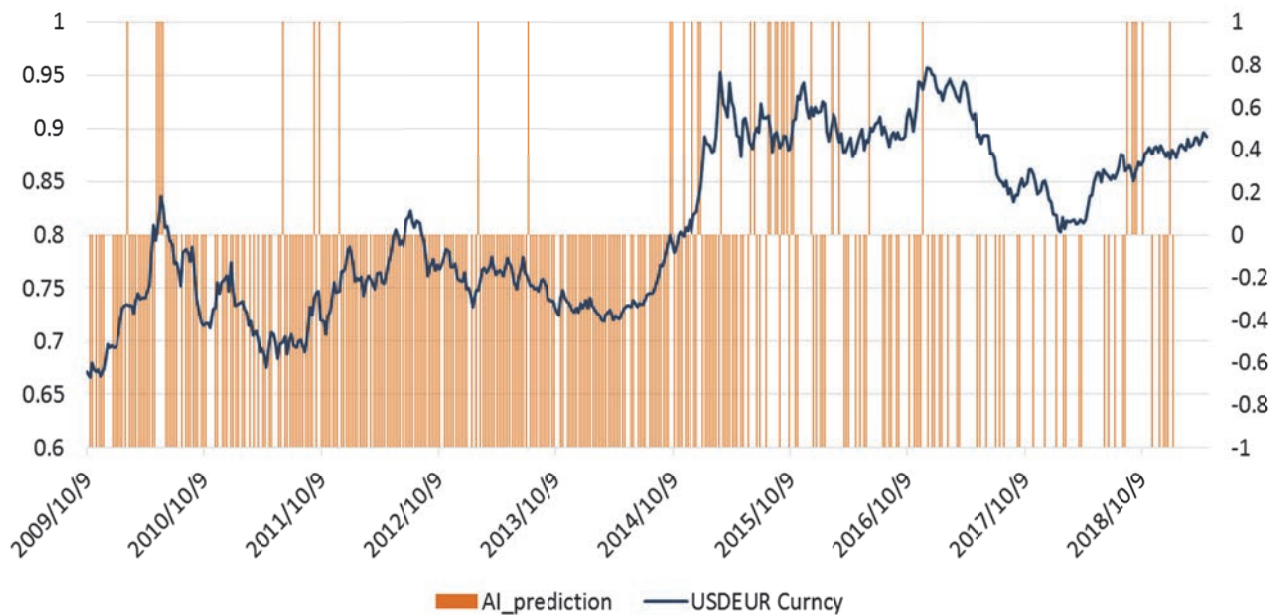


図表5：ディープラーニングによるドルユーロのAI予測



出所：Bloomberg データより三菱UFJ信託銀行作成

図表6：SVMによるドルユーロのAI予測



出所：Bloomberg データより三菱UFJ信託銀行作成

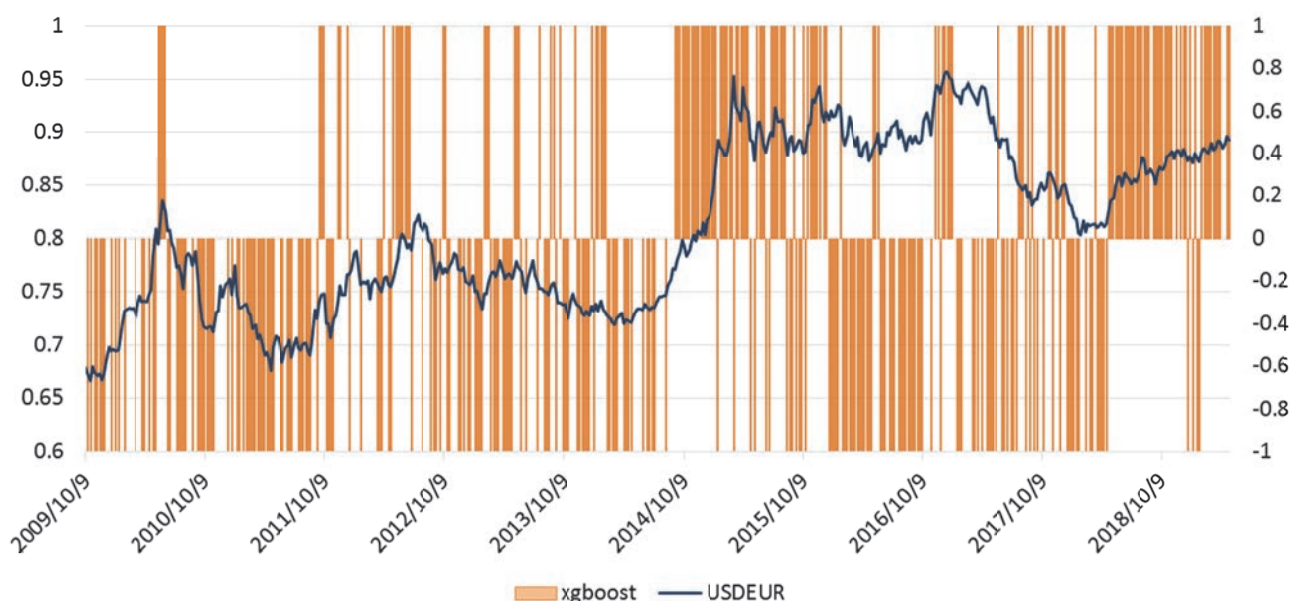
図表4のランダムフォレストモデルは正答割合について改善の余地があるものの、ポイントとなる変化の局面で反応できている。そこで、更に予測精度を向上させる為、ランダムフォレストモデルの上位モデルである **XGBoost**<sup>7</sup> を用いてみる。同モデルはチューニングが必要なデータ項目は増えるものの、全般的に精度が高く近年注目を集めているモデルである。世界のデータサイエンティストが賞金を狙ってデータ分析をする **Kaggle**<sup>8</sup> においても、**LightGBM** と呼ばれるモデルと並んで非常に多く使用されている人気の機械学習アルゴリズムである(実は人気上位はディープラーニングではない。ディープラーニングは必ずしも世の中で考えられているような万能なモデルではないのだ! )。

結果は図表7のとおりであり、図表4と見比べて欲しい。**XGBoost** でかなり精度が上がったのが分かるはずだ。実際も、このようなプロセスでモデルの選択を行っていく。もちろん、今回お見せしたモデル構築プロセスは、あくまで一例であり、扱うデータ次第では様々なパターンがある。また、紹介したプロセスは筆者のやり方であるため、開発者によって様々なやり方があるのであろう。大切なのは自分なりに上手くいくモデル構築方法を多く持つことであると思う。このように、開発する人それぞれの異なるモデルになるのも AI モデルの特徴なのである。また、AI モデルの開発に多くのマーケット知識やデータ分析技術、AI 技術者のノウハウが使われていることも、お分かりいただけたと思う。

<sup>7</sup> XGBoost は Gradient Boosting と呼ばれるモデルとランダムフォレストを組み合わせランダムフォレストの欠点を補っているため、上位モデルと既述。Gradient Boosting とは学習データを分類した際、当てはまりの悪い部分の影響を抑制させる工夫が施されたモデル。

<sup>8</sup> 機械学習モデルを用いた世界最大規模コンペティション。データを元にしてスポンサー企業が問題を出し、その問題に対して参加者である Kagglers が予測結果を出す分析モデルを作成、分析モデルのベストスコアで順位を競う。上位者には賞金やメダルが授与され、Kaggle のランクにもその内容が反映される。

図表7：XGBoostモデルによるドルユーロのAI予測



出所：Bloomberg データより三菱UFJ信託銀行作成

## V. 今後の展望

これまで2ヵ月に亘り、資産運用におけるAIの現状や、AIを用いたマーケット予想モデルの構築について紹介してきた。

冒頭の疑問「運用者はAIに代わられてしまうのか？」に対し、今更言及する必要はないのかもしれないが、“NO”というのが筆者の答えである。ただ、AI時代に突入し、マーケット形成にも変化が起こりつつある現在、我々ファンドマネージャーも転機を迎えようとしている。時代の変化に対応できないファンドマネージャーはマーケットから退場しなければならない時代がもうそこまで来ているのかもしれない。

IV章で構築したAIモデルも予測精度は高いが、自動運転までできるかというところではない。インプットデータではカバーしきれない事象が起こった場合に当該モデルの信頼度が当然下がるため、AIがどのようなデータで予測しているかをファンドマネージャーが把握し、そこから逸脱したと判断されれば、AIによる運用を停止したり、学習データを更新したりする必要があるのだ。つまり、AIとファンドマネージャーはお互い補完し合う関係なのである。今の技術では、AIにインプットデータでカバーしきれない事象かどうかを判断させ、自らそのデータを見つけ、学習していくようないわゆる完全自動化にはまだ距離を感じる。ただ、将来、各ファンドマネージャーが自分の相場観を反映させたAIモデルとともに運用する時代はそう遠くないかもしれない。つまり、人間はどのような材料でマーケットが動いているか判断し、そのデータをAIが学習し、人間では検知できないような真理を我々に教えてくれるような補完体制だ。その時にいえるであろうことは、AIは敵というよりは寧ろ、非常に強力な仲間のような存在になっている可能性が高いということだ。

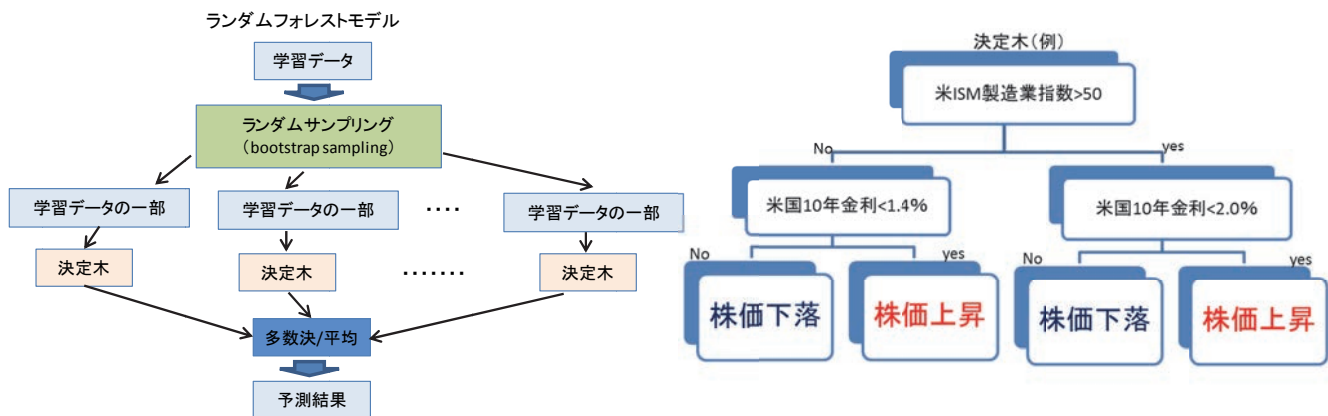
(2019年10月15日 記)

【付録 機械学習アルゴリズム詳細】

付録1：ランダムフォレスト

数百～数千回ランダムサンプリングしたデータを決定木に当てはめ、それら算出結果を多数決/平均するモデル。欠損値などのノイズに強く、パフォーマンスのよいモデルとされている。チューニング項目が少なく、数ある機械学習アルゴリズムの中でも使い勝手がよい。

付録図表1：決定木とランダムフォレストの概念図

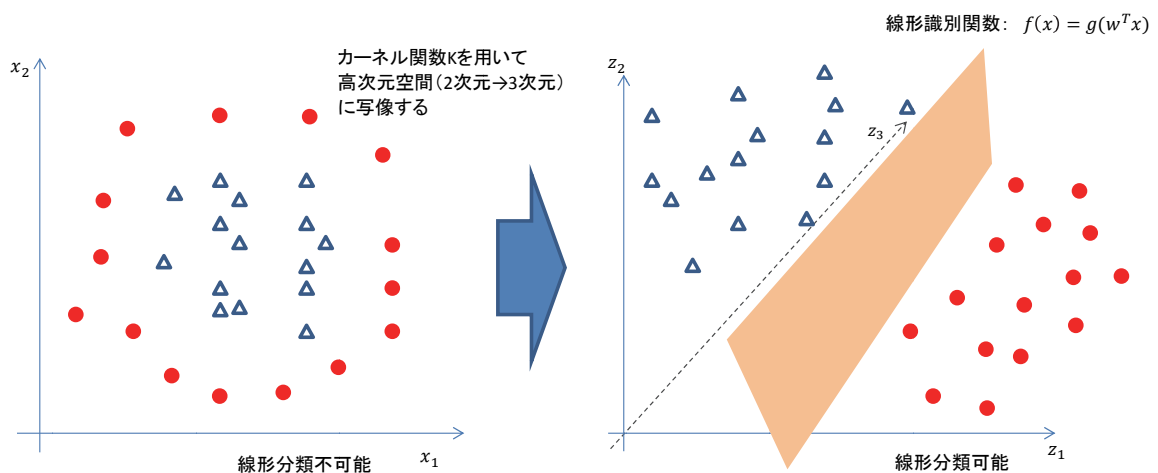


出所：三菱 UFJ 信託銀行作成

付録2：サポートベクターマシン(SVM)

学習データを元々のデータ空間から高次元空間に写像し、その高次元空間上でデータ解析を行うモデル。例えば、付録図表2のとおり、2次元では判別しにくい境界でも3次元に写像することで分離が明確になるようなケースがわかりやすいであろう。写像を高次元化するほど線形分離しやすくなるが、一方でオーバーフィット(過学習)の可能性も高まる。比較的少ない学習データでも良い結果が出やすい、機械学習の中で最も人気のあるモデルの一つ。

付録図表2：サポートベクターマシン概念図

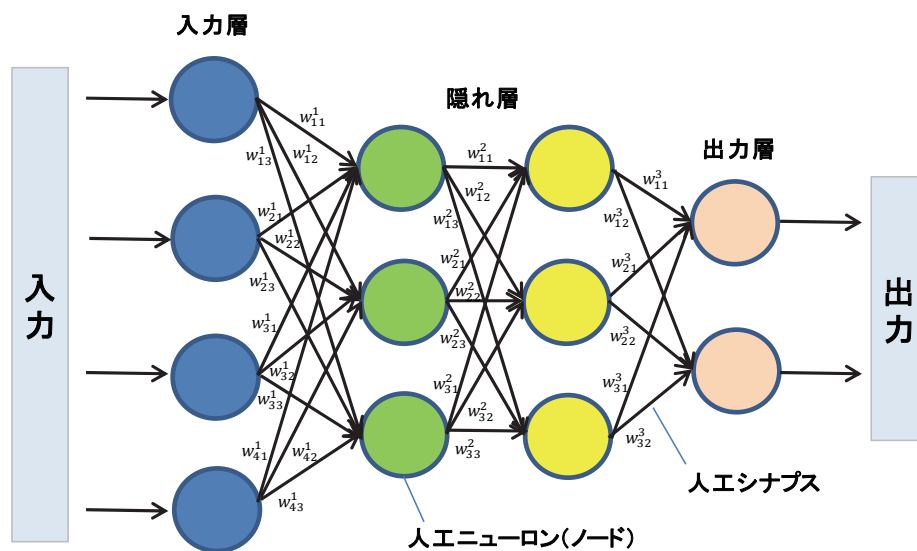


出所：三菱 UFJ 信託銀行作成

## 付録3：ディープラーニング

人間の脳神経(視神経)を模したモデル(ニューラルネットワーク)。一般にニューロンが4層以上のものをディープラーニングと呼ぶ。生物学のニューロンはシナプスを經由して信号と呼ばれる電気刺激を受け、他のニューロンから十分な数の信号を受け取ると、自分自身で信号を生成する。その構造を模したものが付録図表3である。各人工ニューロン(ノード)には活性化関数と呼ばれる数式が組み込まれており、ある一定以上の数値が入ると、“1”のような数字が出力される。各ノードから出力された数字に人工シナプスのウェイト( $w_{ij}^k$ )を掛け合わせたものが次のノードに伝達される。ウェイトは学習データの解答から最適化される。このようにして過去の入力データ(学習データ)のパターンを記憶する。この構造からもお分かりになると思うが、ある時点のデータが翌時点に影響を与えるような系列関連の情報を学習する能力は同モデルには備わっていない。そのため、系列関連の情報を伝える場合は、同じニューラルネットワークの中でも畳み込みニューラルネットワーク(CNN)、再起型ニューラルネットワーク(RNN)、Long short-term memory(LSTM)などのモデルを使用するか、学習データに時系列の情報(一階差データや移動平均など)を入れる必要がある。同モデルは柔軟かつ強力な機械学習アルゴリズムである一方、大量のデータを必要とする。

付録図表3：ディープラーニング概念図



出所：三菱UFJ信託銀行作成

※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】

- ・「AI ファンドマネージャーの力量は？」2019/ 2 /27 日本経済新聞
- ・「株式暴落の正体」武者陵司 2018/12/27 株探ニュース
- ・「Information in the Yield Curve about Future Recessions」Michael D. Bauer and Thomas M.Mertens Federal Reserve Bank of San Francisco [2018]
- ・「Tests for unit roots and initial condition」Ulrich K.Muller and Graham Elliott[2003]
- ・「ヘッジファンド、18年成績好調はブリッジウォーターとルネッサンス」2019/ 3 /28 Bloomberg
- ・「現代時系列分析」田中勝人 [2006/11]
- ・「Rによる計量経済分析」福地純一郎 伊藤有希 [2011/ 7]
- ・「Python 3ではじめるシステムトレード ——環境構築と売買戦略」森谷博之 [2016/11]
- ・「シミュレーターでまなぶニューラルネットワーク」森谷博之 [2019/ 1]
- ・バフェット氏、4つの名言に学ぶ「真の長期投資」谷山歩 [2016/12]
- ・「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/ 9]
- ・「Hands-On Machine Learning with Scikit-Learn and TensorFlow」Aurélien Géron [2017/ 4]
- ・「はじめてのパターン認識」平井有三 [2012/ 7]
- ・「深層学習 (DEEP LEARNING)」Ian Goodfellow, Yoshua Bengio, Aaron Courville [2018/ 3]



## 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）