

日本銀行金融機構局金融高度化センター

AIを活用した金融の高度化に関するワークショップ

第1回「AIを活用した金融の高度化『総論』」

AIと機械学習の直感的理解と金融への応用

統計数理研究所

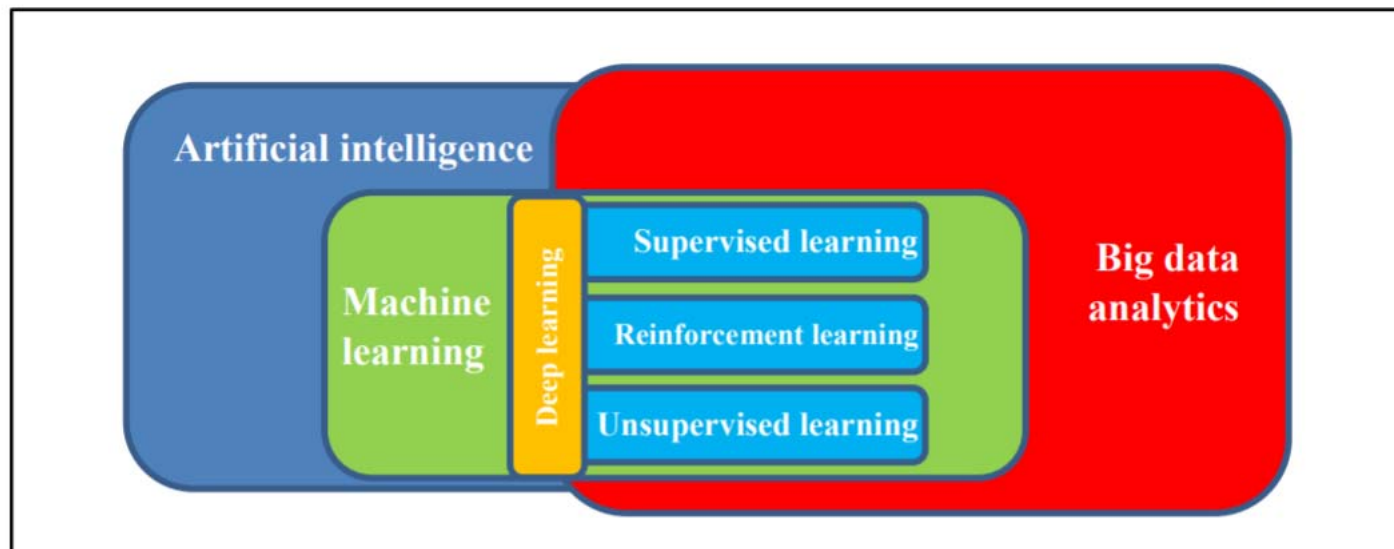
山下智志

Contents

1. 人工知能、機械学習と統計学
 2. 信用リスクモデルにおける機械学習
 3. 人工知能、機械学習の発展を阻むもの
 4. 金融におけるこれまでの人工知能ブームの顛末
 5. ビッグデータとデータ構造化
- 参考資料1. データリレーション(名寄せ)の技術
- 参考資料2. フィンテック、ビットコイン

ビッグデータ、人工知能、機械学習

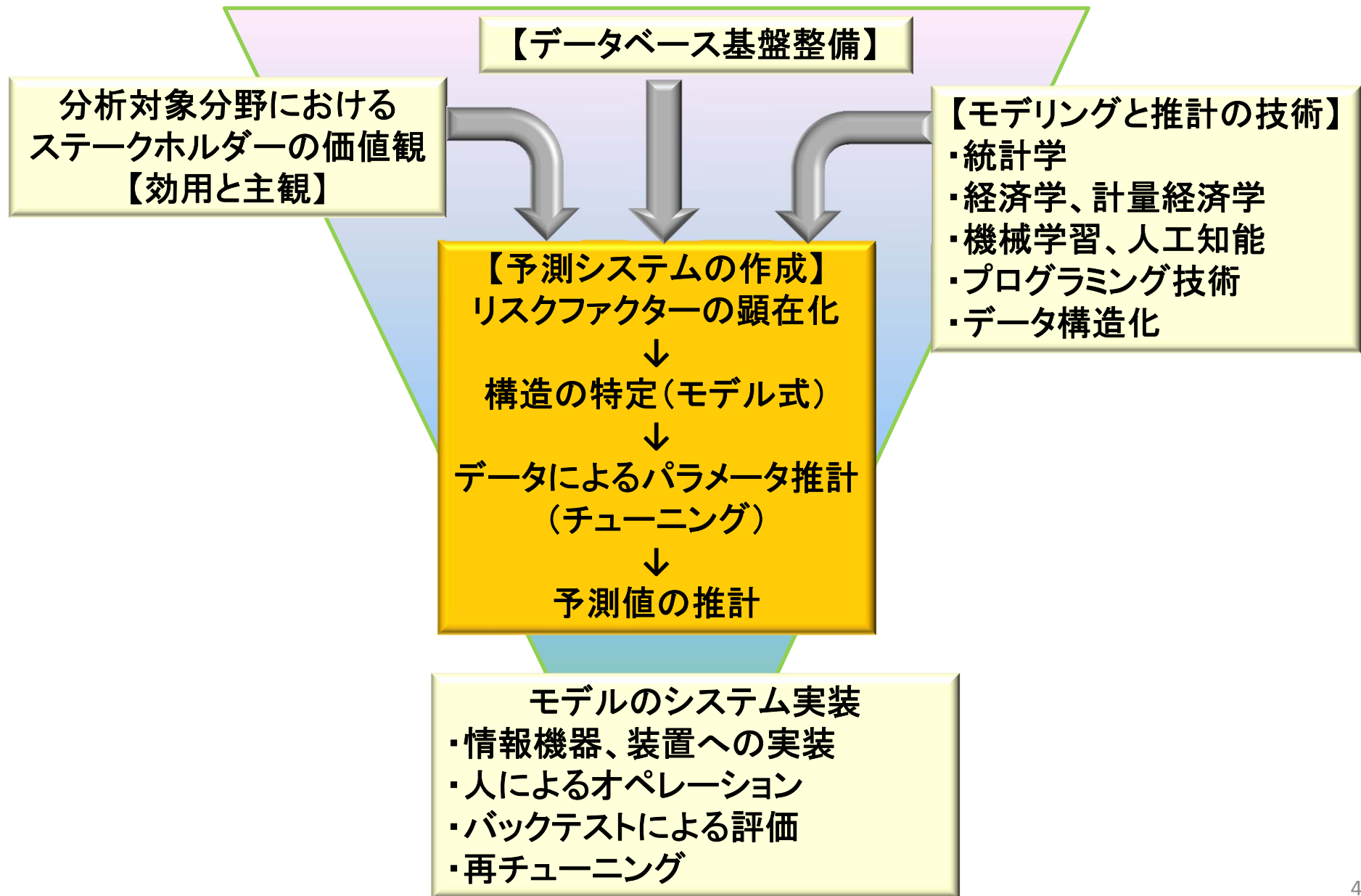
Figure 1: A schematic view of AI, machine learning and big data analytics



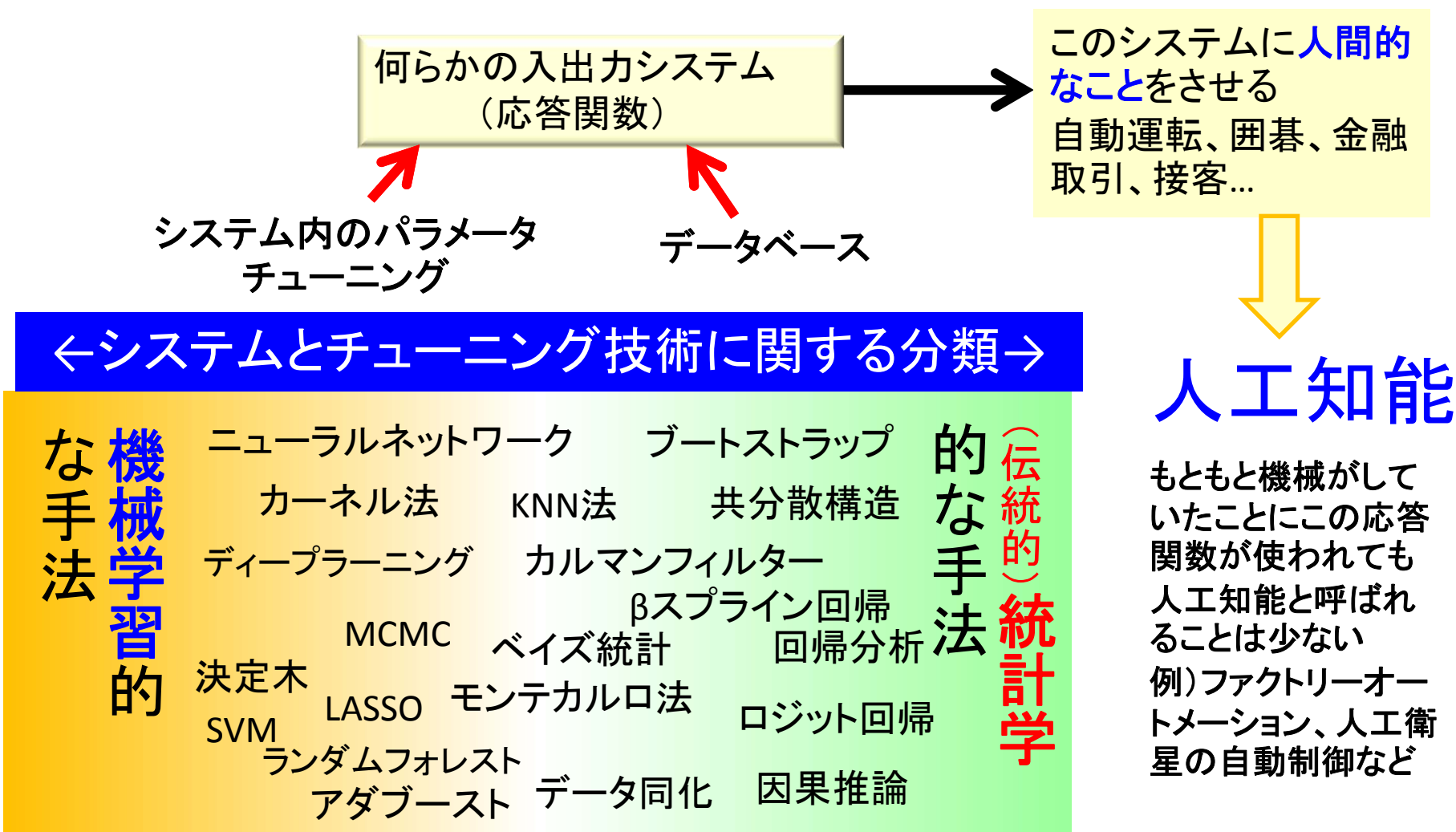
Artificial intelligence and machine learning in financial services
Market developments and financial stability implications

the Financial Stability Board (2017)

モデリングの一般的構造と実装(統計モデル、AI共通)

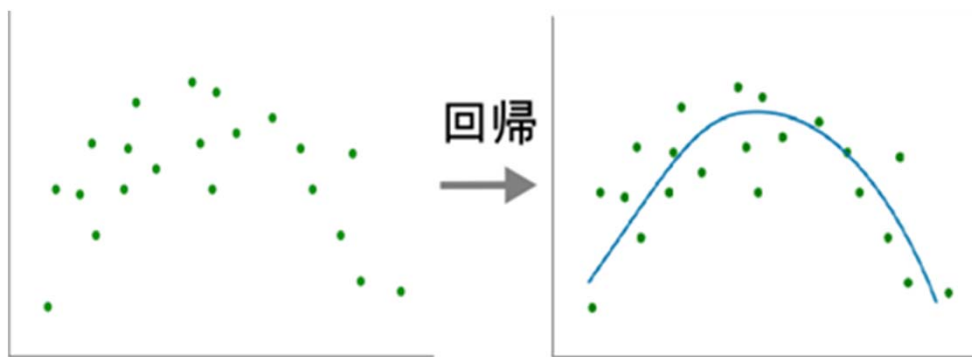


人工知能、機械学習、統計学の関係

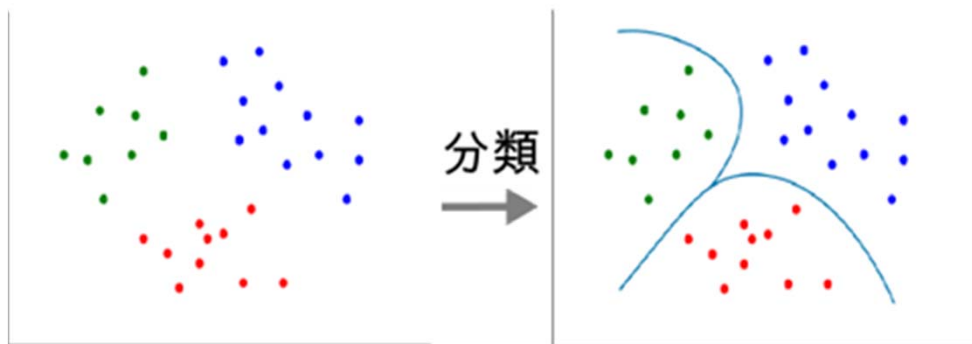


主に回帰と判別のための変数**選択**とパラメータ**推計**の同時同定
(まれにデータ構造化)

機械学習が活躍する課題を単純化したら...



データより変数間の関係を記述する(モデル)。
構築されたモデルにより、予測を行い最適戦略を作成する。
→株価予測、マーケティング、

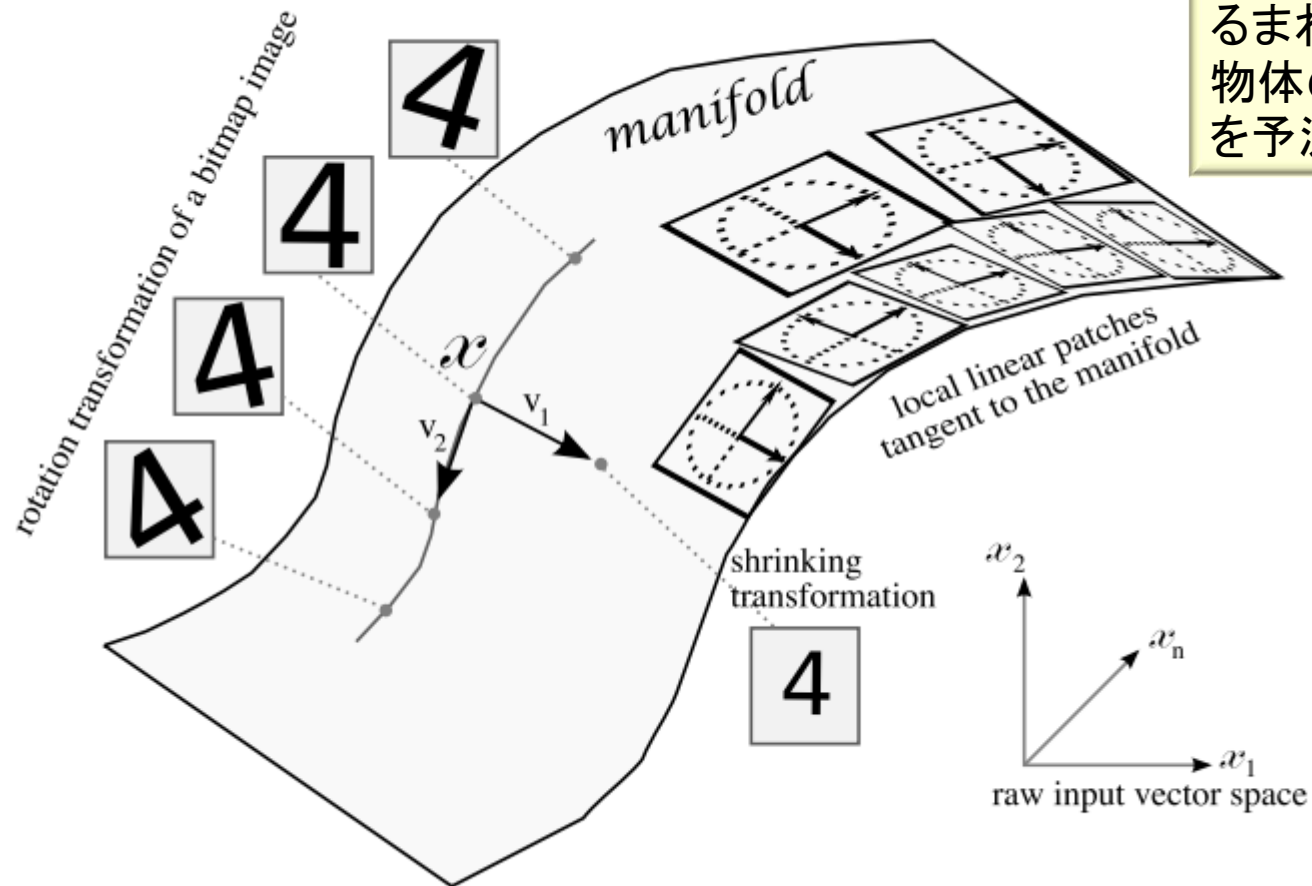


データより変数と判別との関係を記述する(分類モデル)
構築されたモデルにより、予測を行い最適戦略を作成する。
→顧客対応、デフォルト予測

回帰と分類の両方をつくり、組み合わせる
→自動運転、囲碁、

極端な形状を示すデータに対する回帰

Real Data Are on Highly Curved Manifolds

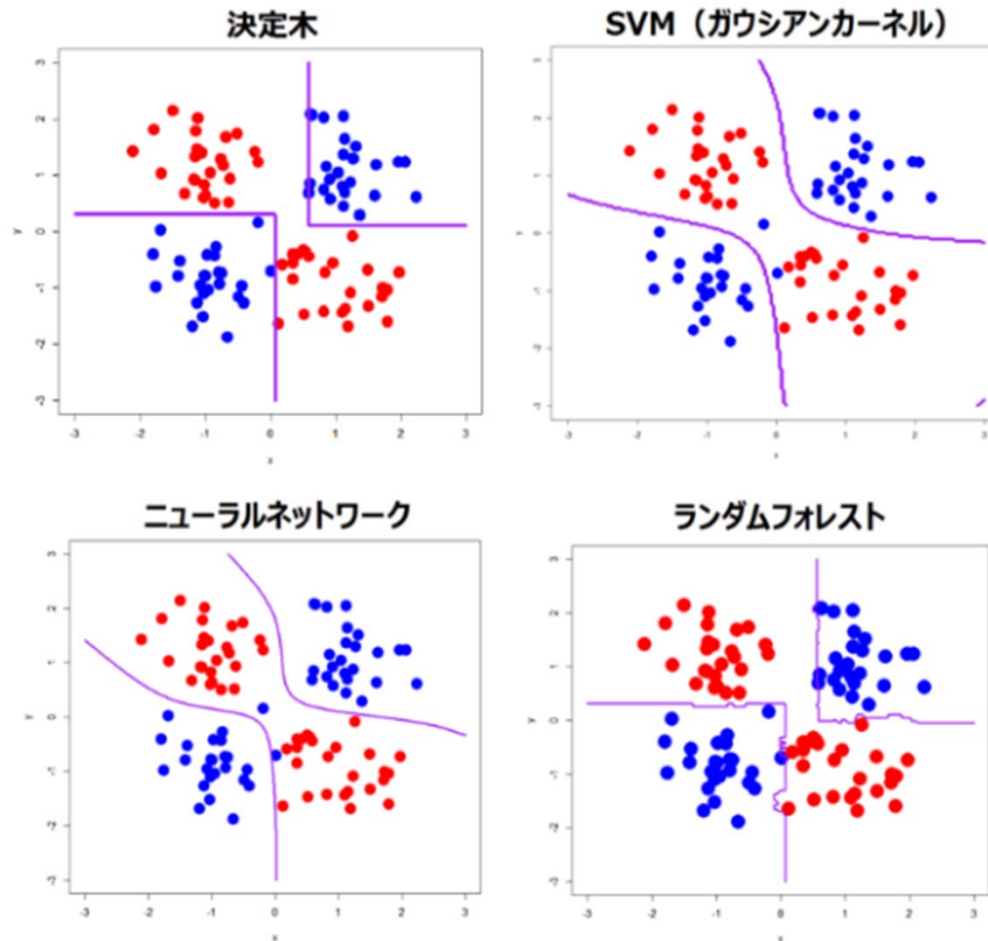


変な平面上をぐるぐるまわりながら動く物体の将来の形状を予測する

出典 Yoshua Bengio: ICML2012Tutorial, June, 26th, 2012, Edinburgh, Scotland

分類に関する機械学習の直感的理解(再掲)

2つの財務指標と**デフォルト企業(青)**、**非デフォルト企業(赤)**の分布
→デフォルト領域と非デフォルト領域に分割する
【通常の統計モデルでは判別しにくいような問題をわざと作ってみた】

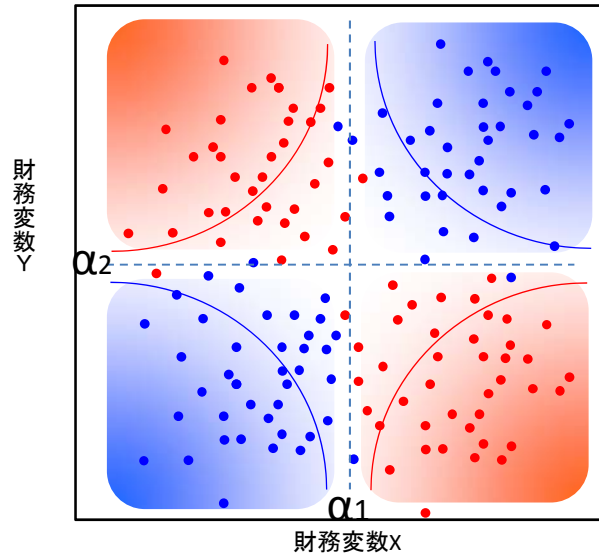


どれも信用リスクモデリングに有効

統数研: 松井教授の資料、樋口知之、人工知能は見よう見まねの究極形、情報管理、Vol. 59 (2016) 8月号

同じことが統計モデルで記述可能か？

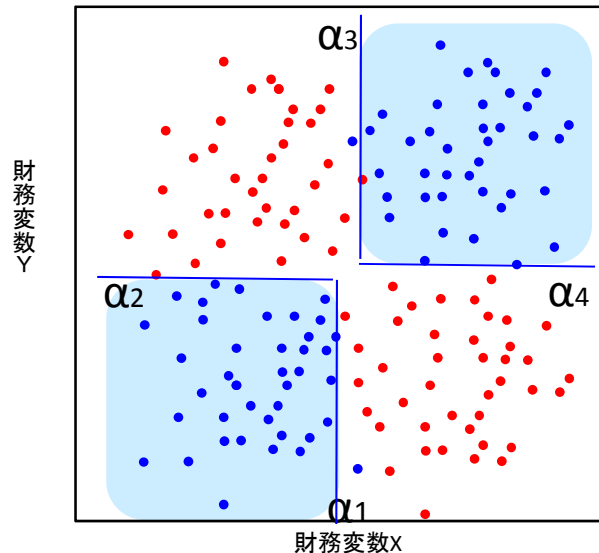
2つの指標とデフォルト企業(青)、非デフォルト企業(赤)の分布



統計モデルでも変数を工夫することによって複雑な配置を表現することが可能である

$Z = \beta(X - \alpha_1)(Y - \alpha_2)$ のイメージ

二つ以上のデータを用いて一つの項(合成変数)を作ることをクロス、もしくは**クロス変数**と呼ぶ。

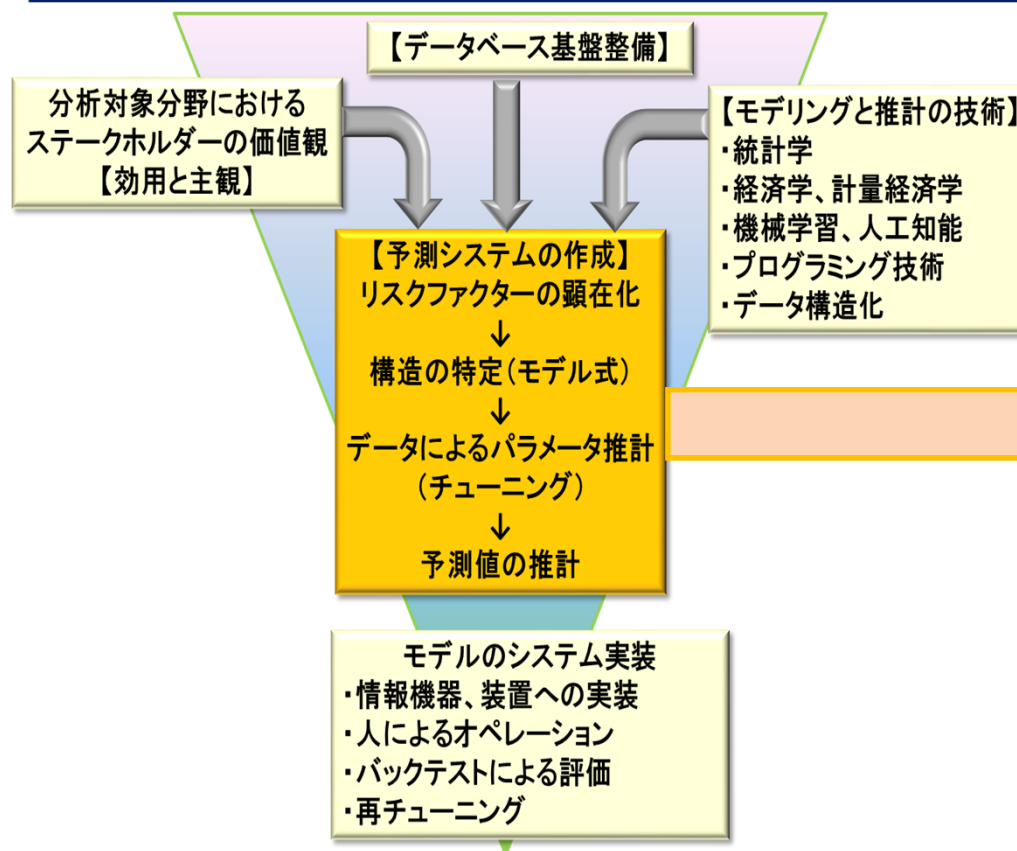


$Z = \delta(X < \alpha_1 \text{ and } Y < \alpha_2, X > \alpha_3 \text{ and } Y > \alpha_4)$ のイメージ

$\delta(*)$ は*の条件に合う場合は1、
合わない場合は0となる変数
デルタ関数もしくは**フラグ変数**と呼ぶ

クロス変数やフラグ変数を用いることによって統計モデルの表現に自由度が増す。

機械学習は統計学的みれば何をしていることになるのか？



① 予測システムの作成のうち、
①a 説明変数の選択
 (モデル式の作成)
①b パラメータの推計
 (チューニング)
を**自動**で求めている。

② 厳密解ではなく試行実験による
近似解を求めている

③ 試行実験は大きな計算負荷
→ 高性能なコンピュータ

Contents

1. 人工知能、機械学習と統計学
 2. 信用リスクモデルにおける機械学習
 3. 人工知能、機械学習の発展を阻むもの
 4. 金融におけるこれまでの人工知能ブームの顛末
 5. ビッグデータとデータ構造化
- 参考資料1. データリレーション(名寄せ)の技術
- 参考資料2. フィンテック、ビットコイン

クレジットスコアリングの具体的な方法論のバリエーション

- 統計学（及び機械学習）の手法を用いたアプローチ

- ▶ 過去のデータから「デフォルト」「非デフォルト」の2群に最も効率よく分類するモデルを作成し、新たな分析対象企業のデータに適用してデフォルト確率を予測する。

- ▶ データの分類アルゴリズム

- ✓ 判別分析

統計学の手法を用いたアプローチ

- ✓ ロジック・モデル etc...

- ✓ ニューラル・ネットワーク

機械学習の手法を用いたアプローチ

- ✓ 決定木分析

- ✓ サポート・ベクター・マシン (SVM) etc...

- 確率論の手法を用いたアプローチ

- ▶ マートンモデル

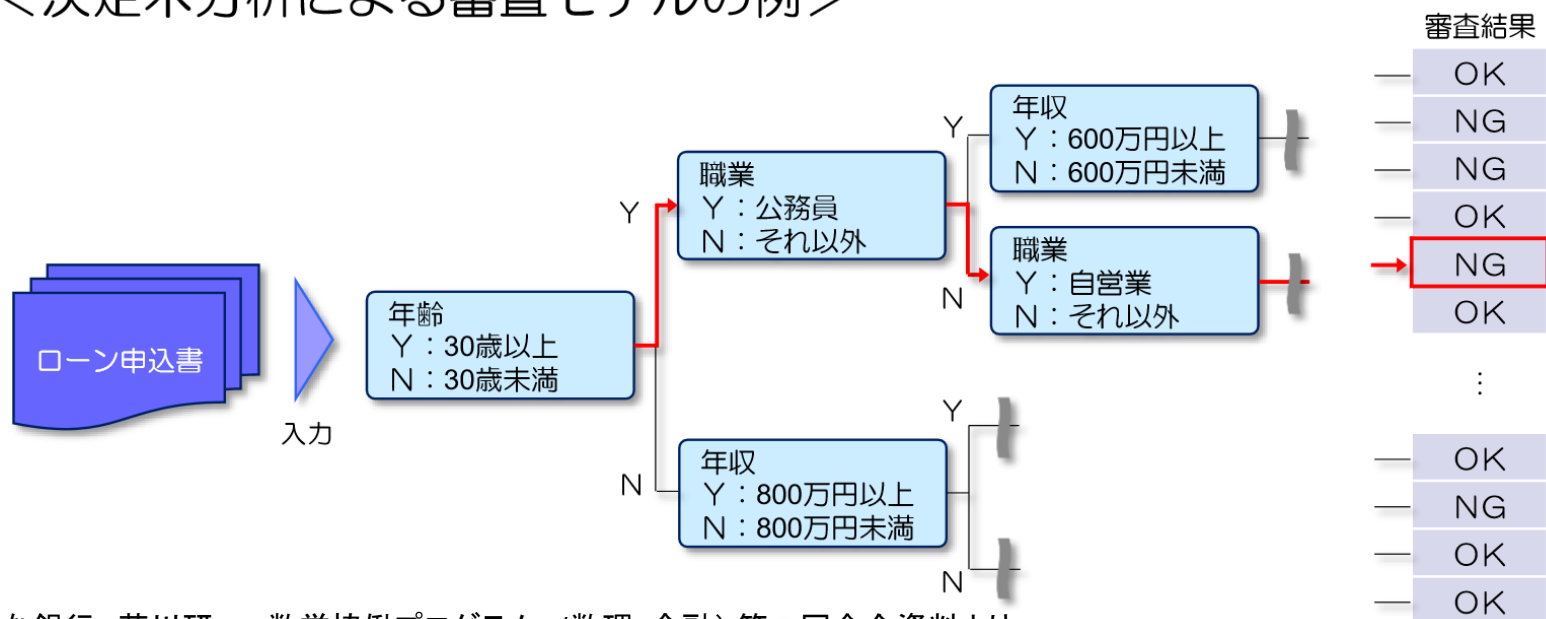
機械学習的モデルによるクレジットスコアリング(1)

● 決定木モデル

- ▶ カードローンなどの審査モデルに用いられている

ローンの申込情報、個人信用情報、銀行取引状況などをもとにしたスコアリングモデル

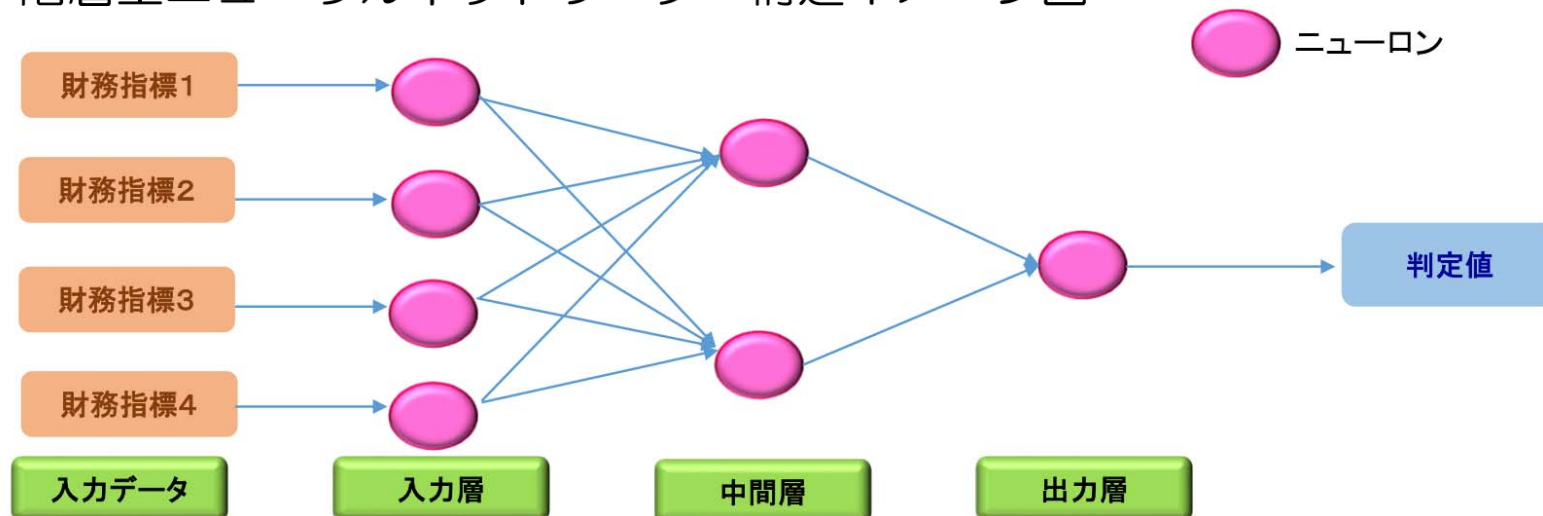
＜決定木分析による審査モデルの例＞



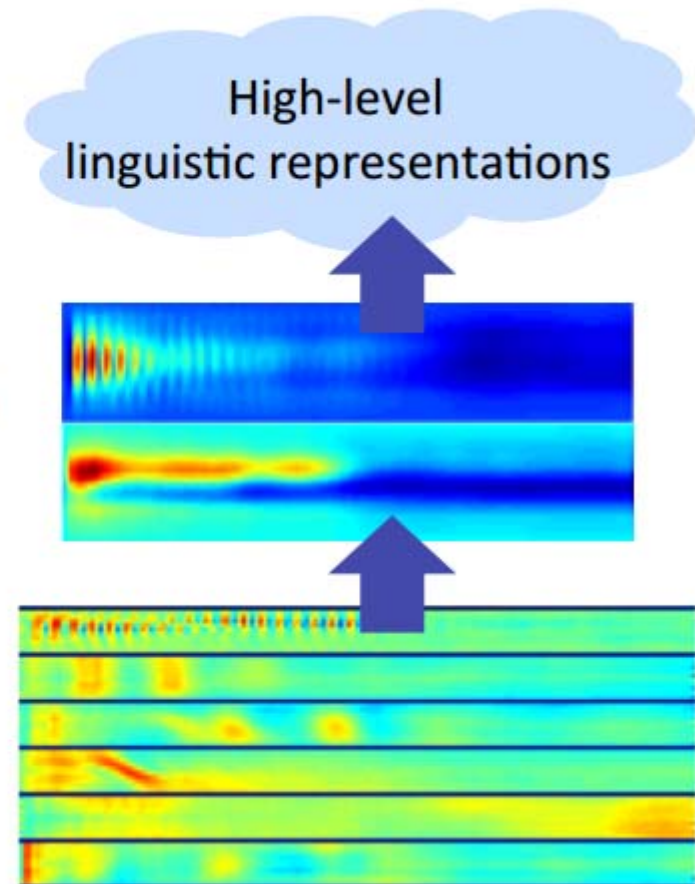
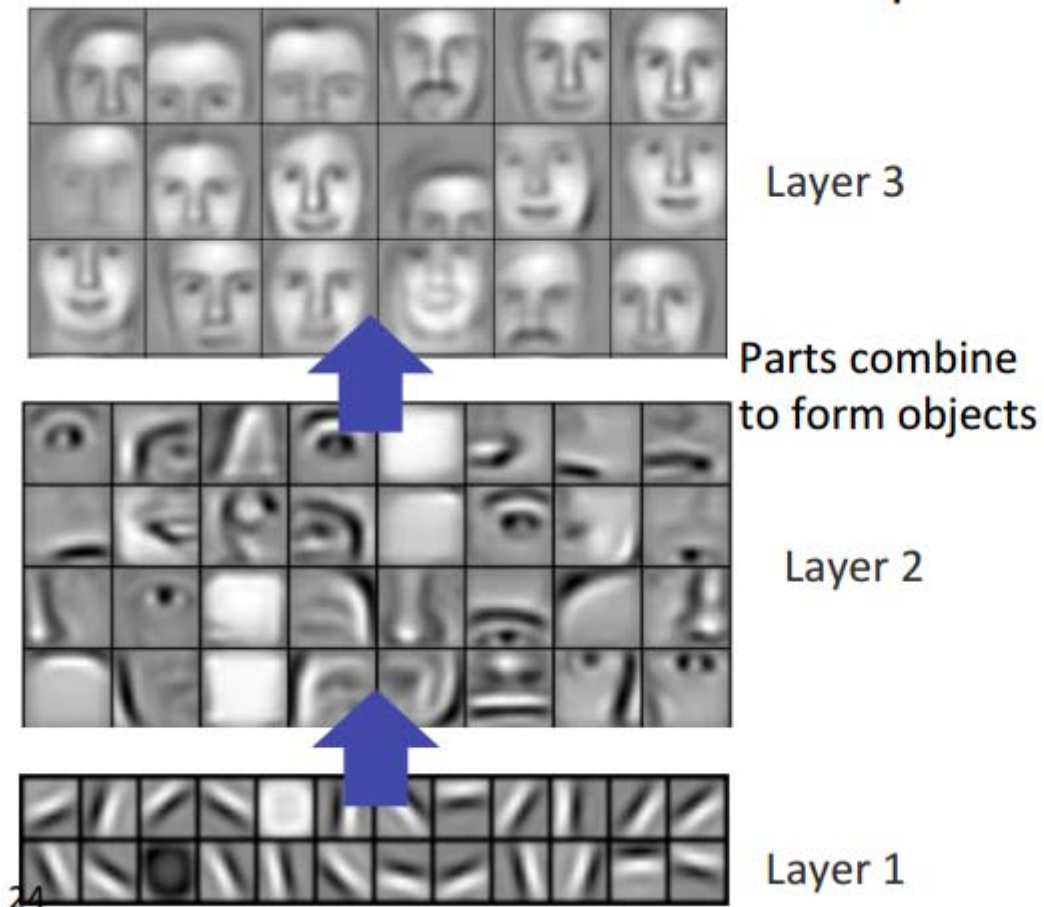
機械学習的モデルによるクレジットスコアリング(2)

●ニューラルネットワーク

- ニューラルネットとは、脳のニューロンの相互作用をコンピュータ上で単純化したモデルで、分類や予測のためのサンプルデータから、学習によりパターンを一般化することで、グループ分けする手法である。
- 連結構造
 - ・階層型ニューラルネットワーク
 - ・相互結合型ニューラルネットワーク
- 階層型ニューラルネットワーク 構造イメージ図



階層型ニューラルネットワークの図的イメージ

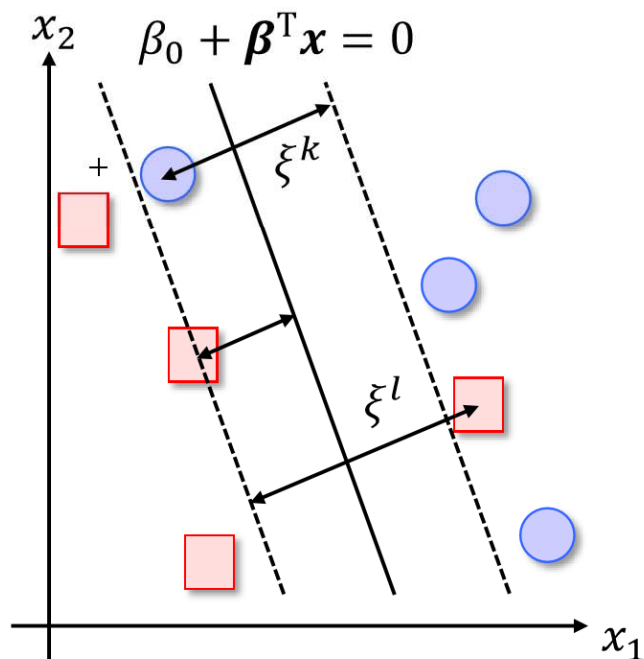


機械学習的モデルによるクレジットスコアリング(3)

● サポート・ベクター・マシン (SVM)

➤ 線形判別不可能な場合

○と□を完全に判別する超平面が存在しない場合



✓ スラック変数 ξ^k ($k = 1, \dots, m$)、パラメータ $C > 0$

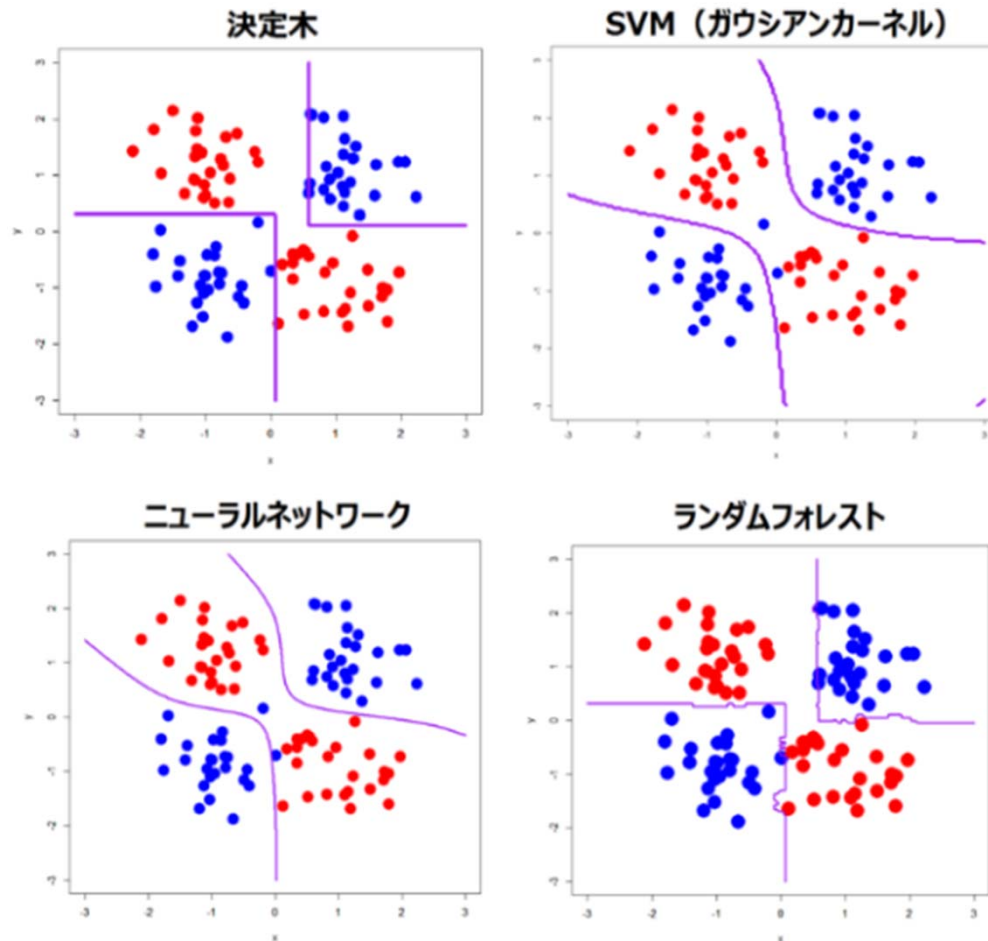
$$\begin{aligned} & \underset{\beta_0, \beta, \xi}{\text{maximize}} && \|\beta\|^2 + C \left(\sum_{k=1}^m \xi^k \right) \\ & \text{subject to} && y^k (\beta_0 + \beta^T x^k) \geq 1 - \xi^k, \quad k = 1, \dots, m \end{aligned}$$

✓ 双対問題

$$\begin{aligned} & \underset{\alpha}{\text{maximize}} && \sum_{k=1}^m \alpha^k - \frac{1}{2} \sum_{k=1}^m \sum_{l=1}^m y^k y^l \alpha^k \alpha^l (x^k)^T x^l \\ & \text{subject to} && \sum_{k=1}^m y^k \alpha^k = 0 \\ & && 0 \leq \alpha \leq C \mathbf{1} \end{aligned}$$

分類に関する機械学習の直感的理解(再掲)

2つの財務指標とデフォルト企業(青)、非デフォルト企業(赤)の分布
→デフォルト領域と非デフォルト領域に分割する
【通常の統計モデルでは判別しにくいような問題をわざと作ってみた】



どれも信用リスクモデリングに有効

統数研: 松井教授の資料、樋口知之、人工知能は見よう見まねの究極形、情報管理、Vol. 59 (2016) 8月号

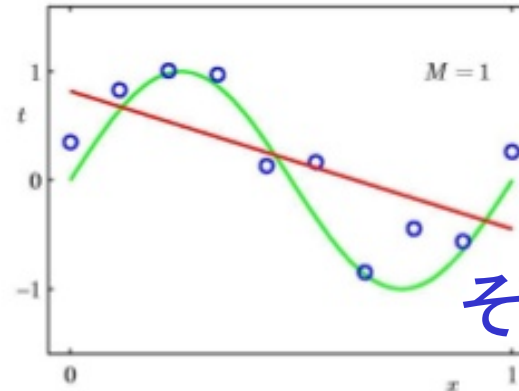
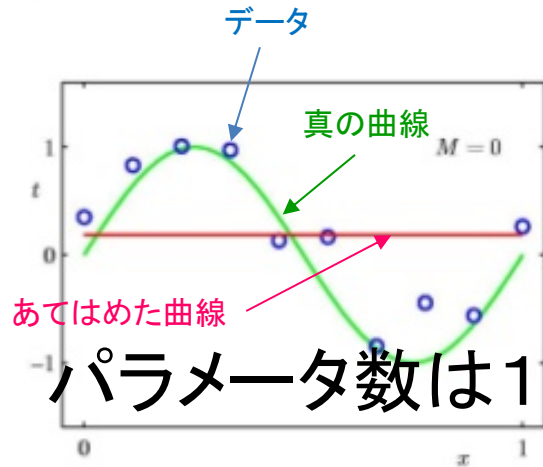
Contents

1. 人工知能、機械学習と統計学
 2. 信用リスクモデルにおける機械学習
 3. 人工知能、機械学習の発展を阻むもの
 4. 金融におけるこれまでの人工知能ブームの顛末
 5. ビッグデータとデータ構造化
- 参考資料1. データリレーション(名寄せ)の技術
- 参考資料2. フィンテック、ビットコイン

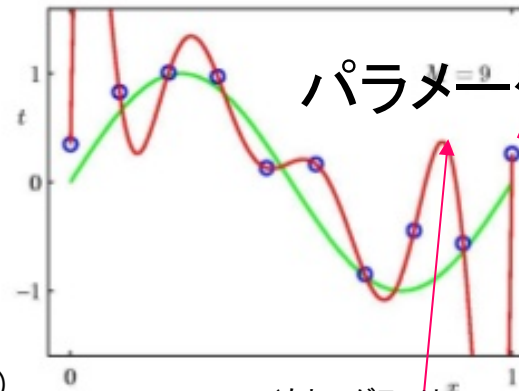
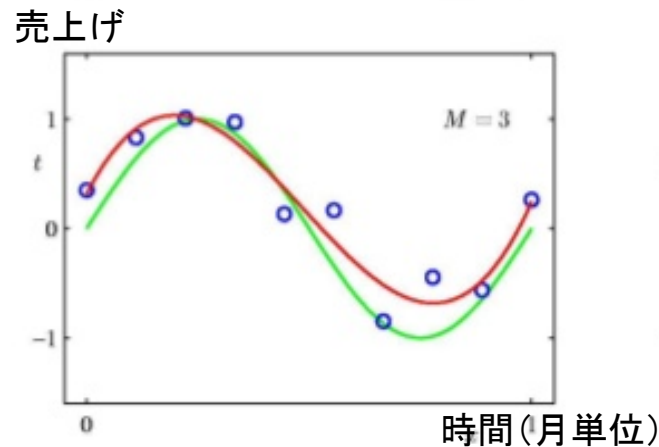
表現能力の高いモデルには過学習 (over fitting) が避けられない

過学習

- 与えられたデータにモデル(回帰曲線)が完全に一致すること
- 過去データ(既知の現象)の説明力は最高だが、未来データ(未知の現象)の予測力は最悪!



そもそも外挿は無理



パラメータ数 = データ数
誤差ゼロ

(もとのグラフは<http://www.slideshare.net/alembert2000/prml-at-1>より)

内挿の失敗: こんな予測を信じますか?

過学習を避けるテクニック

過学習(オーバーフィット)は統計学、機械学習共通の問題だが...

モデルのオーバーフィットを避ける(評価する)統計量および方法

統計モデル

AIC(赤池情報量規準)、BIC、Cp統計量

p値基準

罰則付き最尤法

制約付き最尤法

多重共線性回避

→重相関

→リッジ

バックテスト

k-ホールドクロスバリデーション

ブートストラップ

機械学習的アプローチ

統計モデルは多くの
研究成果がある

機械学習の計算量について

- ① 予測システムの作成のうち、
 - ①a 説明変数の選択
(モデル式の作成)
 - ①b パラメータの推計
(チューニング)を**自動**で求めている。

② 厳密解ではなく試行実験による
近似解を求めている

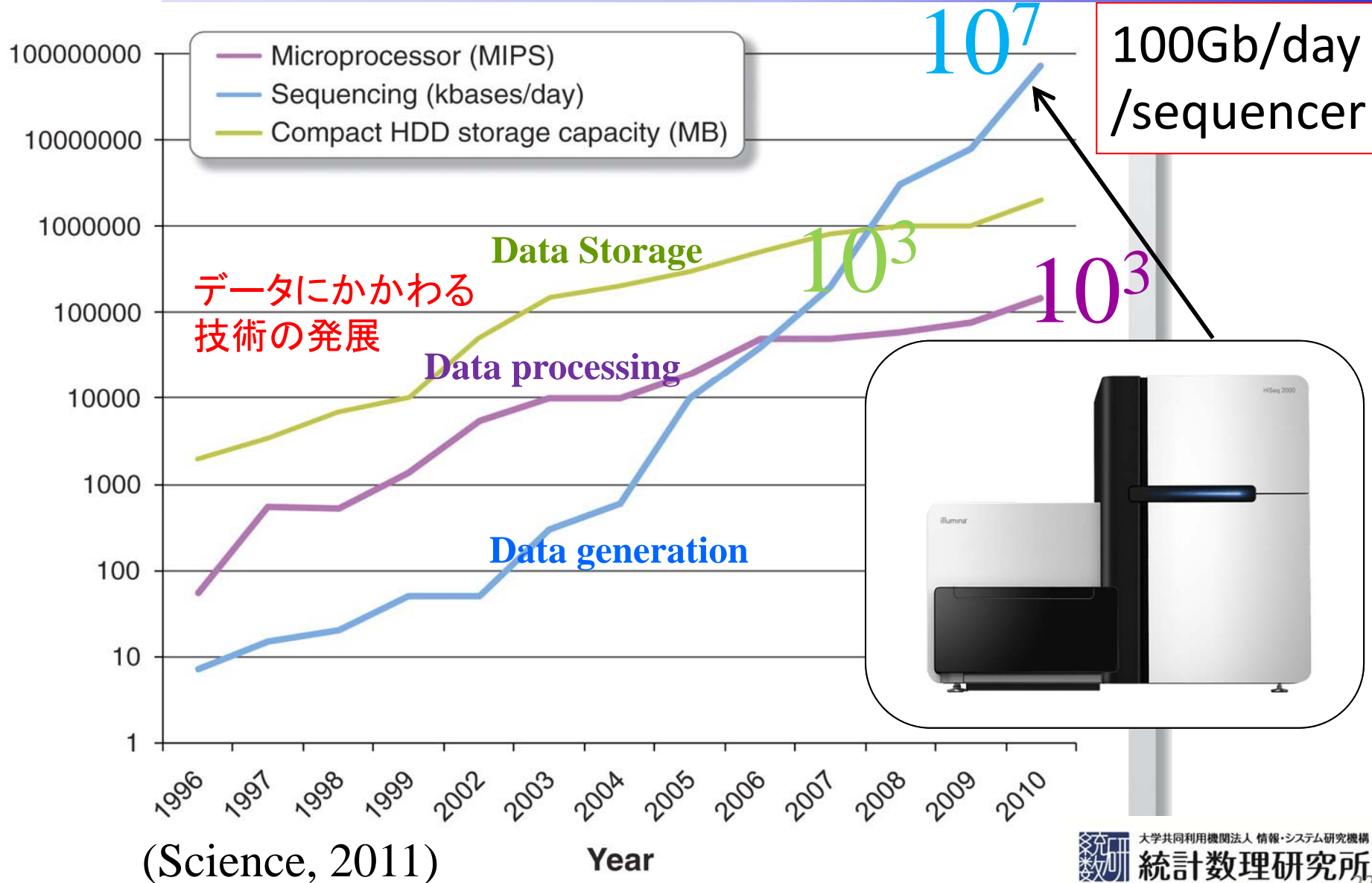
③ **試行実験は大きな計算負荷**
→高性能なコンピュータ

ビッグデータ時代における
データベースの巨大化

コンピュータ性能の進歩
が解決してくれるイメー
ジだが...

ビッグデータ+AIの前に立ちはだかるハードル ハードの性能向上だけではとうてい無理

Moore's and Kryder's Laws fall far behind



データ構造化と機械学習的アプローチの関係

機械学習・人工知能が強い分野とは？

データ構造を特定困難なデータ

- タテ(フィールド)、ヨコ(レコード)の2次元構造化しにくいデータ
- ・将棋データベース
- ・画像情報、音声データ
- ・プロファイリングされていないマーケティングデータ
- ・インターネット文字情報などのキーワードが特定できないテキスト



非線形はそれほど重要ではない

計算量や検定量のため、できれば統計モデルを使いたい



データ構造化技術

統計モデル対機械学習的アプローチの例 (財務・担保・保証によるLGD推計)

- 10-foldのクロスバリデーションでの推定結果
- (I) 【二項ロジット】 or 【AdaBoost】
- (II) 【ロジスティック回帰】 or 【SVR】
- 1段階SVRは先行研究では最も推定精度が高いと言われている。

数値は一部秘匿性のため結果を歪めない範囲でランダム化している

(I) + (II)	R2	Spearman	MAE	RMSE	RAE	連続AR値	サンプル数
(1) 二項ロジット + ロジスティック回帰	0.383	0.539	0.144	0.219	0.668	0.736	570
(2) 二項ロジット + SVR	0.386	0.527	0.145	0.21	0.665	0.738	570
(3) AdaBoost + ロジスティック回帰	0.395	0.576	0.135	0.229	0.604	0.749	570
(4) AdaBoost + SVR	0.402	0.578	0.132	0.227	0.600	0.746	570
(5) SVR	0.342	0.497	0.140	0.230	0.653	0.678	570

十分に構造化されたデータベースにおいては
大きな差異はない

1段階で直接LGD推定。先行研究では最も精度がいいと言われている。

- (4) 【AdaBoost + SVR】 がR2, MAE, RMSE, RAE, 連続AR値に関しては推定精度が高い。
- 先行研究で最も精度が高いと言われている1段階のSVRより推定精度が高い。

Contents

1. 人工知能、機械学習と統計学
 2. 信用リスクモデルにおける機械学習
 3. 人工知能、機械学習の発展を阻むもの
 4. 金融におけるこれまでの人工知能ブームの顛末
 5. ビッグデータとデータ構造化
- 参考資料1. データリレーション(名寄せ)の技術
- 参考資料2. フィンテック、ビットコイン

人工知能研究の歴史

知的活動を再現

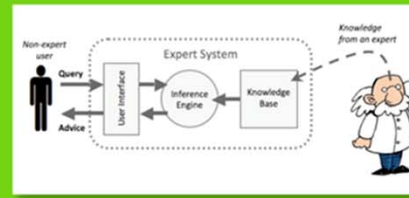
1960年代 記号処理方式

記号処理のためのルールや数式をプログラム化し思考や推論など人間が行う情報処理を行わせる



1980年代 エキスパートシステム

専門家の知識やノウハウをルール化し、コンピューターに処理を行わせる



2000年代 統計的アプローチ

膨大なデータをベイズ理論に基づく統計的手法で計算し自らルール生成し情報処理する

$$P(A|H) = \frac{P(H|A)P(A)}{P(H)}$$

インターネット登場

ビッグデータ・IoT

1950

1960

1970

1980

1990

2000

2010

メインフレーム登場

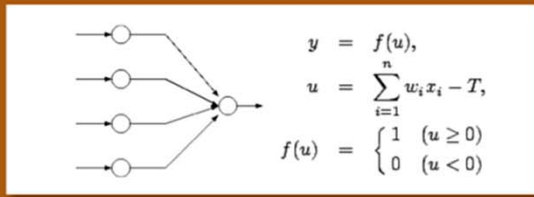
PC登場

ハードウェア性能向上

脳の活動を再現

1958年 パーセプトロン

脳の神経活動を数式モデル化しコンピューターに処理させる初歩的なニューラル・ネットワーク

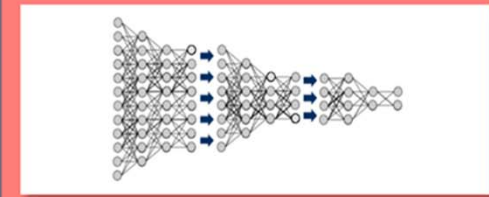


人工知能の冬

人工知能の冬

2006年 ディープラーニング

脳科学の研究成果を取り入れより忠実に脳の神経活動を再現



第1次ブーム 1990年代前半

- ・使われた技術: ニューラルネットワーク、ファジィ、(カオス・フラクタル)、
- ・適用分野: 市場運用、特に株式市場と為替市場
- ・金融商品: 投資信託
- ・効果: ほとんど成果無し、
→モデル・手法に問題があったか、導入時期が不運であったかは不明
- ・結果: 1990年代半ばには消散

第2次ブーム 2000年前後

- ・使われた技術: ニューラルネットワーク、サポートベクターマシン
- ・適用分野: 信用リスク分析、(オペレーショナルリスク分析)
- ・金融商品: 銀行の法人貸し出し、消費者金融の審査
- ・効果: 予測精度については一定の成果あり
- ・結果: 銀行法人貸し出しについては数年で終了、
消費者金融の審査については継続

第3次ブーム 2014年くらいから現在

人工知能第1次ブームの顛末



動機: 1990年に投資信託や年金運用、特金・ファントラなどの運用商品の行き詰まり
ポートフォリオ理論やオプション理論などの経験から
当時先端的な数学理論に飛びついた。

トライアル1: ファジイ理論

運用のプロフェッショナルの判断をモデル化(教師あり学習)
プロのノウハウをメンバーシップ関数に落とし込み、運用成績の向上を狙った。
実際にファジイ運用を売りにした投資信託商品があった。

結果

運用成績が上がらず敗退

原因

そもそも恒常的に好成績を上げている「運用のプロ」がない
→ いればその人を高額で雇って、巨大ファンドを運用させればよい
→ 一時期運用成績のよいファンドマネージャーがいたとしても
確立したノウハウがあるわけではない(cf.野村の日本戦略株ファンドなど)

トライアル2:ニューラルネットワーク(運用)

株式のリターンをニューラルネットワークで予測(教師あり学習)
中間層が1層のニューロモデルに落とし込み、運用成績の向上を狙った。
ファジィ・ニューロ運用を売りにした投資信託商品があった。

結果

運用成績が上がらず敗退

原因

- ・同時の計算能力から中間層を重層化することができず、非線形成分を十分に反映できなかった。
- ・中間層が1層であれば、Bスプラインなどの統計モデルと大差ない予測になるし、統計モデルのような検定量によるモデル評価ができないため不利。
- ・ニューラルネットワークによるモデリングは過学習との戦いであるが、それを認識している人が当時はいなかった。

人工知能第2次ブームの顛末

トライアル3: サポートベクターマシン(信用リスク)

- ・バーゼルⅡ(新BIS規制)による信用リスクモデル(内部格付モデル)の適用
- ・貸出先のデフォルト予測をサポートベクターマシンやニューラルネットワークで
- ・業界標準の2項ロジットモデルを超えることが目標

結果

- ・推計精度は高質な2項ロジットモデルと同等だが実務的に普及せず、数年で消滅

原因

- ・デファクトスタンダードの2項ロジットモデルと比較して顕著に有効とはいえなかった
- ・時系列的に短いデータベースだったので、ロバストネス(頑健性、過学習)に自信がもてなかった
- ・統計モデル(2項ロジットモデル)と違い、リスク要因がわからないため、稟議書との対応関係が不明であった。
 - 2項ロジットモデルでは推計パラメータにより点数表が作れる。また「自己資本比率の影響が何%」や「利益率が10%と改善されたらデフォルト確率は何%下がる」が明確であるのに対してサポートベクターマシンでは要因が明確でない。

人工知能第3次ブームの特徴

- ・ワトソン、囲碁・将棋、自動運転、マーケティングなど金融以外からの刺激
- ・フィンテック、仮想通貨など新興情報産業に対する漠然とした期待

・技術者からのボトムアップより、
トップダウンの意思決定が強力

・ビッグデータ活用の解
決策としてのAI

シーズプッシュではなくニーズプルの局面
迅速にニーズに応える
実用エビデンスを作る

JスコアにおけるIA活用法:みずほ&ソフトバンク

AIを活用して、お客さまのさまざまな情報から、お客さまの信用力と可能性をスコア化したものです。AIスコアは、信用力のスコア化において従来重視された情報に加え、必ずしも重視されなかったお客さまの情報を活用することで、スコアの精緻化を目指しています

低金利・高極度額の魅力的なお借入条件 貸付利率(年率)0.8% - 12.0%契約極度額10万円 - 1,000万円(※)

開発メンバーの紹介

当社の「AIスコア」は国内外のデータ解析コンペティションにおいて入賞実績、海外の政府系機関でデータ解析に関する研究の経験を持つ、AI/機械学習や分析技術に関してトップレベルの水準にあるメンバーのもと、試行錯誤を繰り返してJ.Score独自のAIスコアとして創り上げています。日々高度化するデータ分析・解析技術を追いかけて高品質かつ高性能なAIスコア創りを目指しています。

技術的な知識と業務的な経験は車の両輪の様なものです。以下の様な多様なバックグラウンドを持つ他のメンバーとコラボレーションする事で、各メンバーそれぞれ今までの経験を糧にデータサイエンティストと有機的につながりながらAIスコア創りに貢献しています。

- ・さまざまな業界・業種で経験
- ・みずほ銀行の個人融資審査業務
- ・ソフトバンクの割賦販売における与信モデル構築

かなりのレベルで人間が介入しているように見える。

JスコアにおけるIA活用法:みずほ&ソフトバンク

AIを活用して、お客さまのさまざまな情報から、お客さまの信用力と可能性をスコア化したものです。

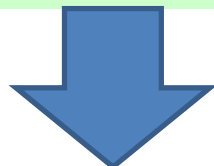
AIスコアは、信用力のスコア化において**従来重視された情報**に加え、必ずしも重視されなかったお客さまの情報を活用することで、スコアの精緻化を目指しています

低金利・高極度額の魅力的なお借入条件
貸付利率(年率)0.8% - 12.0%
契約極度額10万円 - 1,000万円(※)

消費者金融

小口、迅速審査、高金利 → 人的コストをかけられない

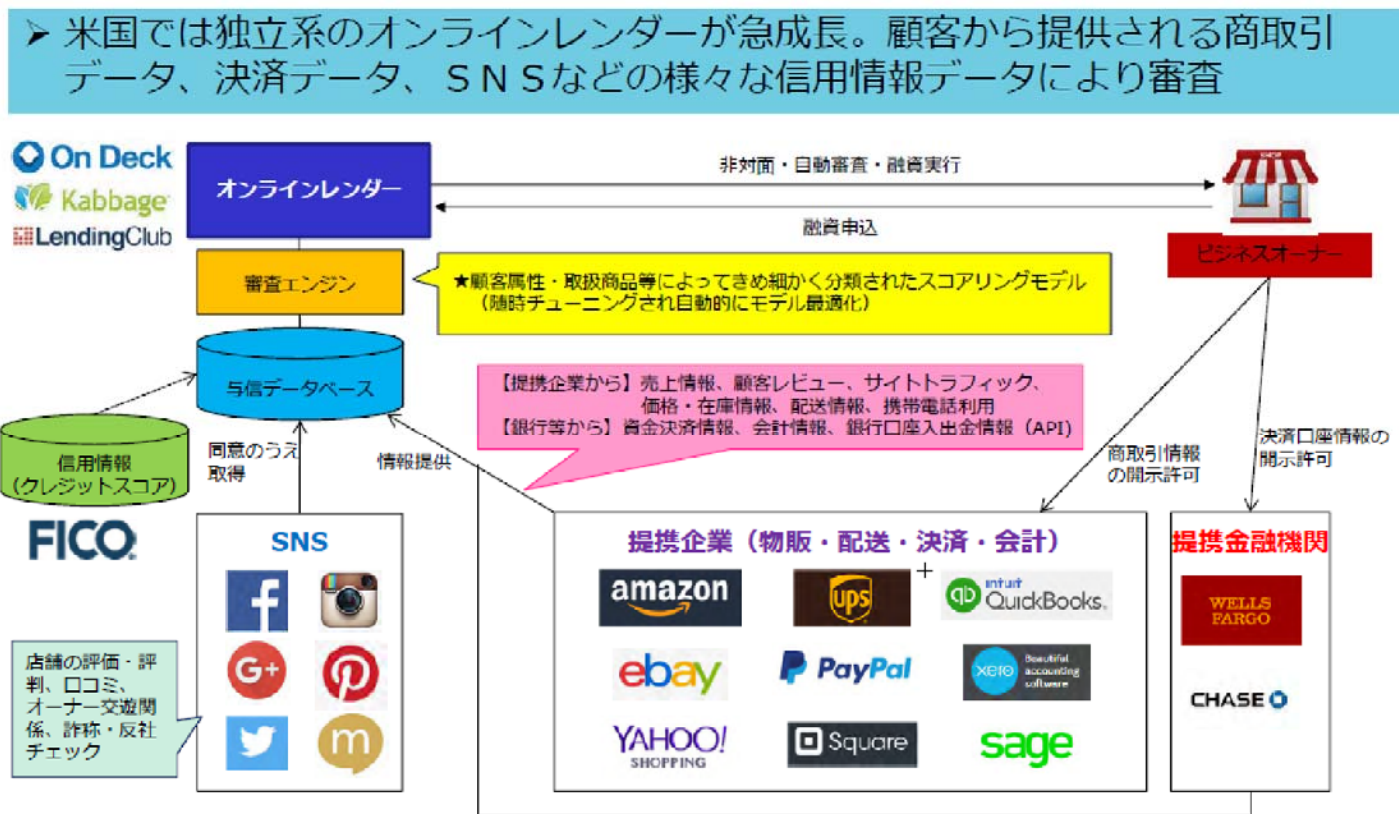
AIを必要とするニーズブルの代表的な例



モデル構築においても運用においても人的資源が必要、
AI=人的資源の排除ではない

近未来の信用リスク計量化モデル

主戦場は、機械学習などの変数選択&パラメータ推計から、
 (個人&企業の)データプロファイリングに移行しつつある。
 個人情報保護法などのコンプライアンスやデータ構造化技術が重要に



出典: 日本銀行金融高度化センター ITを活用した金融の高度化の推進に向けたワークショップ(第3回)
 杉本好正氏

Contents

1. 人工知能、機械学習と統計学
2. 信用リスクモデルにおける機械学習
3. 人工知能、機械学習の発展を阻むもの
4. 金融におけるこれまでの人工知能ブームの顛末
5. ビッグデータとデータ構造化

参考資料1. データリレーション(名寄せ)の技術

参考資料2. フィンテック、ビットコイン

機械学習はビッグデータと表裏一体
ビッグデータの構造記述において統計的方法と機械学習的方法が存在する

ビッグデータ(英語: Big data)[1]は、情報技術分野の用語としては、通常のデータベース管理ツールなどで取り扱う事が困難なほど巨大な大きさのデータの集まりのこと[2]。通常は、構造化データおよび非構造化データが含まれ、その多様性とサイズや要求される処理の速度と高度化のため、格納[3]と活用(検索/共有/分析[4]/可視化など)の実現を目指したもの。

- ① 巨大なデータ
- ② 複数のデータベース
- ③ 非構造化データ

↓

通常ツールで、取り扱うことが困難
(②は除外されることがある)

一般的な「ビッグデータ」の定義と性質

サイズからの定義と性質

→通常のデータベース管理ツールなどで取り扱う事が困難なほど大きい
(POSデータ)

質的定義と性質

→非構造化データ

(インターネットの複数サイトからの観測情報...数値情報を含む)

→高度テキスト情報

(新聞、雑誌などのテキスト情報、SNSの書き込みログ)

金融における「ビッグデータ」の定義と性質

1. 勘定系データベース(取引の日々のキャッシュフロー)
2. 与信系データベース(数値情報)
3. 与信系データベース(テキスト情報)
4. インターネット上の数値情報
5. インターネット上のテキスト情報
6. 国際マクロ情報

与信系データベース:統合データベース(数値情報)

1974 帝国データバンク COSMOS1サービス開始
東京商工リサーチ 企業情報データベースサービス開始

↓ 信用情報からデフォルト・財務情報へ

2000 RDB 事業法人データベースサービス提供開始
2001 CRD協会 デフォルトデータベースサービス、スコアリングの提供
2002 RDB 個人事業者データベースサービス 提供開始
2004 地方銀行協会 CRITSサービス開始

↓ デフォルト・財務情報から毀損・保全(担保保証)情報へ

RDB、CRITSの統合データベース(保全情報が不足?)
2007-2012 統計数理研究所 統合データベース、LGD・ELモデル化

↓

勘定系データの利用、政府統計マイクロデータの利用

データ構造化の分類と手法

・欠損値補間

経済統計、医療統計を中心に研究成果が多い。
経済 (Hot Deck, Cold Deck) と医療 (ICE, Knn, MICE) とは系統がやや異なる
単純は「削除」「平均値補間」は少なくなりつつある。

・異常値補正

研究成果は少ない。
3シグマ折り返し処理、数値情報の順位化、数値情報のカテゴリー化などの
経験的処理が一般的である。
一般的にモデルパラメータに与える影響が大きい

・リレーション

いわゆる名寄せ。
情報が複数のデータベースに分散している場合、必須の処理
特定フィールドの完全一致から、確率一致性へ進化

・テキストデータの数量化

リレーション(名寄せ)の技術について

金融機関におけるリレーションの例

与信データベース(行内はほぼ完成)

- ・案件情報を名寄せして債務者情報を作成
 - ・貸出要件(担保、保証、金額、ターム、毀損額)は案件データに紐付いている
 - ・財務情報、定性情報、債務者格付は債務者データに紐付けている
- 信用リスクモデリングを行う場合は案件・債務者リレーションが重要

勘定系データベース(進行中)

- ・勘定系データベースを与信判断に利用(与信系データベースとのリレーション)

外部データベース(これから)

- ・経済活動調査個票など公的センサスデータ
 - 無借金経営の企業に対する新規貸出、特定の業種に特殊情報がある、など。
- ・銀行間統合データベース
 - 回収方法などの要因分析

リスクアペタイトフレームワーク → データリレーションのスキルが必要

特定フィールドの完全一致から、統計的一致性評価へ

これまで

特定フィールド(名称、住所、設立年など)の完全一致

→ 名称・住所表記のぶれなどから、不完全一致の対応アルゴリズムはある
完全一致しないデータは基本的にリレーションしない

これから

類似度を指標化し、ある閾値を超えたものはリレーションを行う

→ 統計的マッチング

0. 統計的マッチングの際の前提条件

(1) 「条件付き独立性」

- ・ $[X, Z]$ と $[Y, Z]$ をマッチングする場合 (共変量は Z)、 X と Y に関する Z の条件付き独立性 (CIA: Conditional Independence Assumption) が成立していることが前提。

(2) 「強く無視できる割り当て」

- ・ どちらの群に割り当てられるかは共変量の値に依存し、従属変数による割り当ての影響はあくまで「共変量と従属変数の関係」を通じてのみ間接的に依存している。

高部勲(2017)より抜粋

新統計法と統計改革推進会議がもたらす未来

統計法の目的は、公的統計の作成及び提供に関し基本となる事項を定めることにより、公的統計の体系的かつ効率的な整備及びその有用性の確保を図り、国民経済の健全な発展及び国民生活の向上に寄与することとなっています(第1条)。

公的統計は行政利用だけではなく、社会全体で利用される情報基盤として位置付けられています。

基幹統計

国勢統計
住宅・土地統計
労働力統計
小売物価統計
家計統計
個人企業経済統計
科学技術研究統計
就業構造基本統計
全国消費実態統計
社会生活基本統計
経済構造統計

法人企業統計
民間給与実態統計
毎月勤労統計
薬事工業生産動態統計
医療施設統計
賃金構造基本統計
国民生活基礎統計
農業経営統計
工業統計
生産動態統計
商業動態統計
企業活動基本統計

これまでは集計データしか見れなかったが、今後は個別データ(個社ごと)を利用可能になった。
→営利目的は少し時間が必要

オンサイト拠点
統計数理研究所、神戸大学、一橋大学

金融機関内部の
顧客データとの名寄せ

まとめ

【これまでの機械学習・人工知能】

- ・ビッグデータから最適予想、最適モデルを自動的に導き出す技術
- ・実際は限られたデータを限られた目的に利用
- ・限定された世界の中では素晴らしい結果を得ている



【これからの機械学習・人工知能】

- ・完全自動の最適化は無理
- ・方法論の選択、目的関数の設定、データ構造化など人間の介在が不可欠
- ・適切な人間の介在によりより広範囲の分野、目的に適用
→これまでの統計モデルの作成過程に類似している。
- ・適切な介在ができる人材育成が鍵