機械学習・データマイニング分野の概要

神嶌 敏弘

www.kamishima.net

2023-07-27 更新

注意

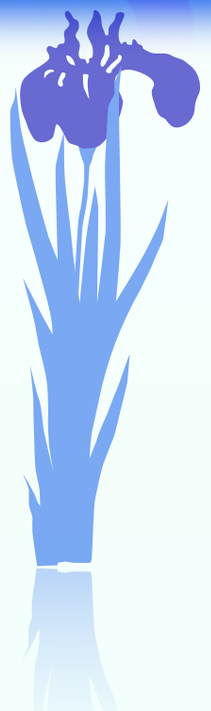
このマークの部分は私の私見に基づくものです
研究者間で同意がとれていた，客観的見地からの内容ではありません

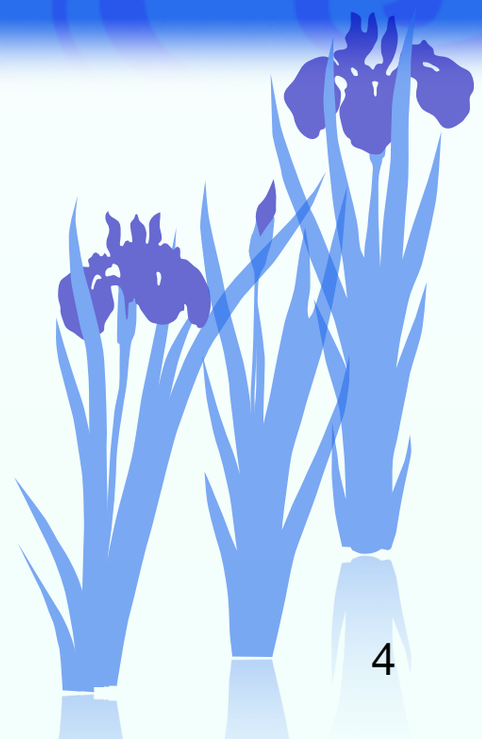
最新版は機械学習・データマイニング分野の概要」として
このページより配布しています
http://www.kamishima.net/jp/kaisetsu/

目次

第Ⅰ部

機械学習・データマイニングとは？





機械学習とは

機械学習の定義

The field of study that gives computers the ability to learn without being explicitly programmed.
— *A. L. Samuel* [1959]

明示的にプログラミングすることなく
コンピュータに学ぶ能力を与えようとする研究分野

※ Coursera の Andrew Ng による機械学習コース などでよく参照されているが，出典をたどることはできなかった．1959年の一般紙に対するインタビュー記事によるものと推察される

Programming computers to learn from experience should eventually eliminate the need for much of this detailed programming effort.
— *A. L. Samuel* [Samuel 59]

経験から学ぶように計算機をプログラミングすることで，細部をプログラミングするのに必要になる手間の多くは減らせる

The field of machine learning is concerned with the question of how to construct computer programs that automatically improve with experience.
— *T. M. Mitchell* [Mitchell 99]

機械学習分野では，経験から自動的に改善を図れるようなコンピュータプログラムを構築する方法について議論している

機械学習の枠組み

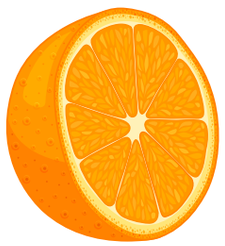
学習

データ

これはリンゴ？

はい

いいえ



いいえ

→

モデル

データ中の
入力と結果の
規則性を集約

推論

これはリンゴ？

↓

はい

学習したモデルを
使った予測

（学習済み）モデル

モデル

規則，ルール，パターン

入力と結果の間の規則性

$$\text{結果} = f(\ \text{特徴}\ )$$

出力 **はい**

入力に応じた予測結果

入力

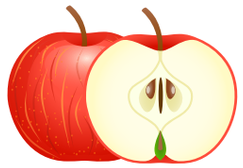
結果が分からないモノやコト

（学習済み）モデル：入力に応じて結果を予測する

入力の表現

計算機にモノやコトを分からせるために，入力は特徴を使って表現

特徴（属性，説明変数，独立変数，計画変数，共変量）
入力のある側面が，どのような状態にあるのかを表す

例：  形は丸い？ ➡ はい　色は？ ➡ 赤い　重さは？ ➡ 200g

特徴に基づいて場合分けし，それぞれの場合に応じて結果を予測

例：「色は赤か緑」「高さ + 幅 + 奥行き ≦ 160cm」

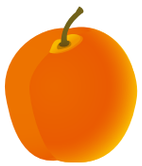
$$\text{結果} = f(\ \text{特徴}\ )$$

入力は特徴を使って表現

結果の予測

$$\text{結果} = f(\ \text{特徴}\ )$$

目的変数，被説明変数，従属変数，応答変数，基準変数


いいえ


いいえ


いいえ


はい


いいえ


はい

特徴に基づく場合分け

色＝赤

はい

データ中の結果を集約することで予測

集約 ＝ 多数決，平均，確率，…

機械学習は道具

予測する結果は利用者が決める

何を予測するかは利用者が指定する

＋

利用者が与えたデータを集約したものが予測結果

機械学習は「結果を集約することで予測」する

➡ 予測がうまくいくかどうかはデータ次第

⬇

利用者が，目的に応じて予測対象を選び，適切なデータを準備する必要

機械学習は道具

機械学習をうまく使いこなせるかは利用者次第

機械学習は馬：競うものではなく，乗るもの [Domingos 15, Domingos 21]

視点を広げる機械学習

機械学習は道具
何をするための？

利用者個人が扱える情報は限られている

- **限定合理性**：合理的に振る舞うがそれは限られた知識と能力に基づく
- 「およそ人は自分の望みを勝手に信じてしまう」 [ガリア戦記]

人間は自身の能力を拡張するために道具を使う

機械学習
結果を集約することで予測

多くの人の視点や
多くの情報を考慮できる

機械学習は利用者の視点を広げるための道具

愚者は経験に学び，賢者はデータに学ぶ

予測することの限界

データに表れている結果を集約することで予測

同じ特徴で表された入力について，同じ結果が得られると仮定

特徴が同じでも同じ結果になるとは限らない

原因の例：特徴にはない情報への依存，実は無作為に決まっている

データは，ありとあらゆる状況を網羅している訳ではない

未知の状況について予測しなくてはならない

汎化：未知の状況と似た状況では似た結果になるなどの仮定を導入して
より一般的な状況に対処できるようにする

不良設定問題：数学的に解が定まらない問題

予測は確率的になるので，不確実な部分が必ず残る

ヒュームの「帰納の問題」

あらゆる状況は尽くせないので，未知の状況は必ず生じる

⬇

18世紀の哲学者デビッド・ヒュームの「帰納の問題」

過去の経験から学んだことを，
確信をもって将来のことにも適用できる方法はあるのだろうか？

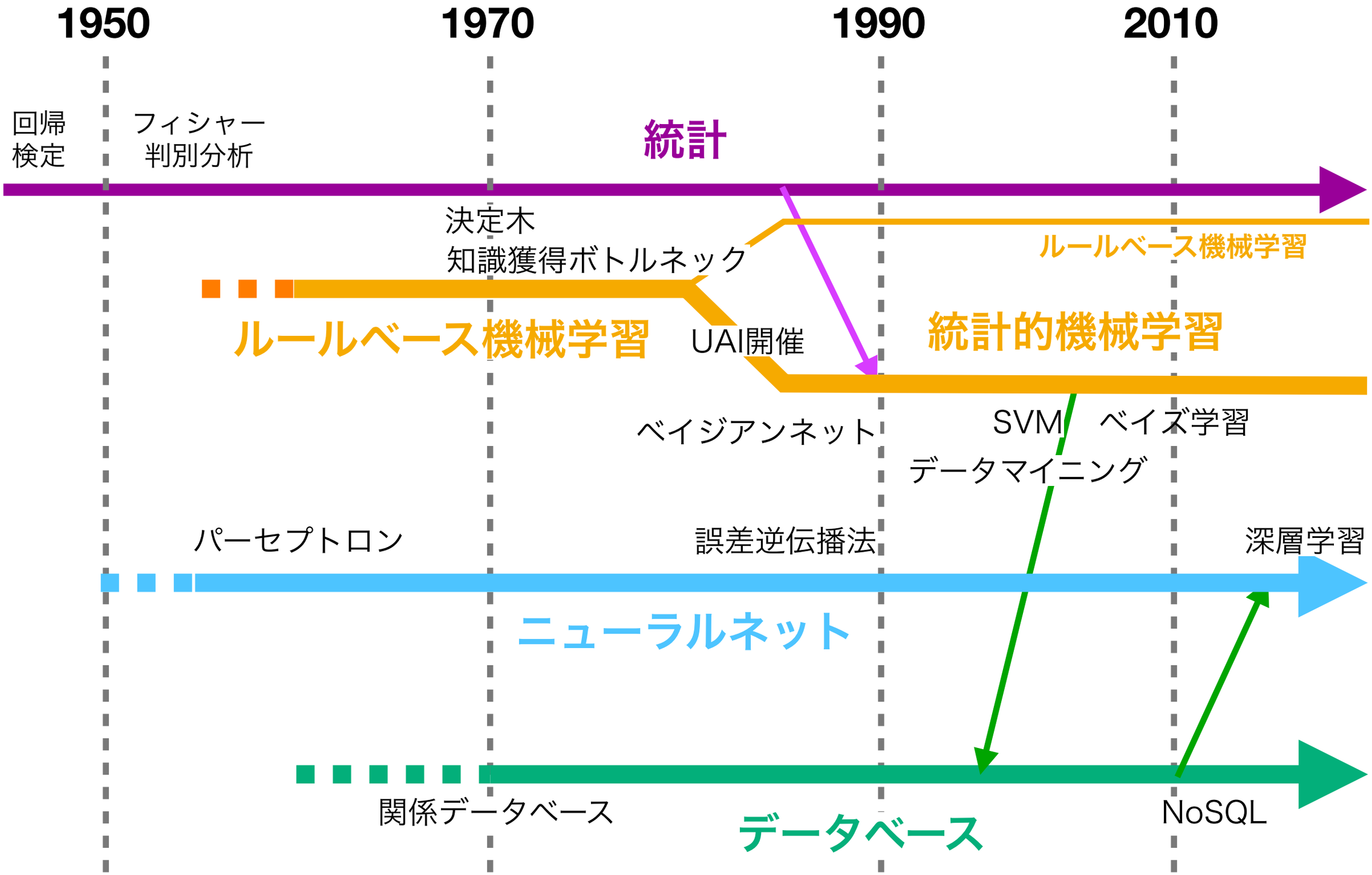
バートランド・ラッセルの帰納主義者の七面鳥

- 最初の日に，9時に餌をもらえたが，最初は疑っていた
- その後もずっと9時に餌をもらえたので，9時に餌をもらえると確信
- クリスマスの朝，餌をもらいにゆくと，丸焼きにされてしまった

人間の場合でも予測は不確実になる

データ分析分野の研究動向

データ分析に関わる分野の変遷



日本の機械学習分野の変遷

- 1995年あたりのDMブームの前後，**機械学習系のグループは，企業では NEC／NTT研究所／日本IBM を除いて解散**
  - 間接的な貢献しかない「要素技術」は，この時期の「選択と集中」のかけ声の中撤退が相次いだ
- 研究系のコミュニティは日本中で約100名ぐらいに縮小
  - 人工知能系の「発見科学」と，学習理論系の1998年からの「情報論的学習理論ワークショップ」
- 00年代には，数学・物理・画像・自然言語処理などの周辺分野から，**現在の機械学習分野を牽引する人達が参入**
  - 新規参入してくれた人の力添えで，2007年にモダンな機械学習の教科書であるPRML本を翻訳
- 山西研が東大計数に2009年にできたのは転機
  - 00年代に参入してくれた人が独立して研究室を立ち上げはじめた

下馬評的予測

私の提唱する機械学習の大原理

手作業でやっていた規則の生成が，どんどん複雑化して手に負えなくなったら，機械学習が適用されるようになる

- 形態素解析，音声認識の音韻モデル，機械翻訳などはこの道をたどりブレークスルーをもたらした．情報抽出・ソフトウェア工学などが進行中．次はデータの前処理とか(?)
- Igor Perisic@RecSys2015：各サービスごとに多種のDBを参照して複雑に ➡ データパイプラインのアイデア

歴史は繰り返す

- 80年代のニューラルネットは10年代に深層学習で復活，60年代のパーセプトロンは00年代のオンライン学習で復活，80年代の決定木は00年代にブースティング・RFで復活
- この循環するなら90年代のカーネル法が20年代に？

00年代の機械学習ブーム

00年代以降のデータ分析技術の進展でどう変わったのか？

[Panel on Big Data @ KDD2012]

Signal + Noise ➡ **Signal + Weaker Signal + Noise**

信号　外乱　　信号　弱い信号　外乱

Christos Faloutsos

The issue is not just size, the issue is granularity

単に大規模なことが問題なのではない，分析の詳細さこそが重要

Michael I. Jordan

⬇

今まで不明瞭だった細部の情報も取り出せるようになった

弱い信号をとらえるために

データ分析の過程で

- **問題点の認識**：強い信号を捉える手法では，不都合な問題点があることに気づく
- **手がかりの探索**：不都合を解消するために，分析に役立つ先験的な知識を探し出す
- **解決法の開発**：手がかりを活用できる分析手法を開発し，利用できる

↓

問題に応じたテーラーメイドの分析

分析結果を見るときに

分析の前提を把握した上で，結果を読み解く

例：検索データからの経済指標の予測

[Varian 13]

Google の検索語の傾向から経済指標を予測する

- **問題点の認識**：単純な関連性の指標で調べると，検索語も経済指標も多種多様なので，**本当は無関係だが偶然に似てしまう場合がある**
  - 例：検索語 “インド料理店”とUSの自動車販売台数
- **手がかりの探索**：全体のトレンドや季節的な影響などの**要素に分解してみて，それでも関連があれば**，本当に関連があるのではないか？
- **解決法の開発**：**要素ごとに分解する手法を考案**して，ミシガン大消費者信頼感指数などの予測を行った
- **使うときの注意**：要素に分解しても偶然に関連することは完全になくなるのではなく，**その可能性が減少するだけという前提**

例：Webカムで車の台数を数える

[Idé+ 17]

交通インフラが整備されていない地域で交通量を把握するため，安価なWebカムで車の台数を数える

- **問題点の認識**：ぼけていたり，車が重なって写っていて**既存の画像処理技術では数えることができない**
- **手がかりの探索**：**車の台数は自然数で，それら大きさはほぼ同じ**
- **解決法の開発**：予測台数が自然数になるという**情報を生かした予測手法を開発**
- **使うときの注意**：車の大きさにばらつきがある場合などには**数え間違いをすることもある**

深層学習

深層学習

深層学習 (Deep Learning)

第3次ニューラルネットワーク黄金期

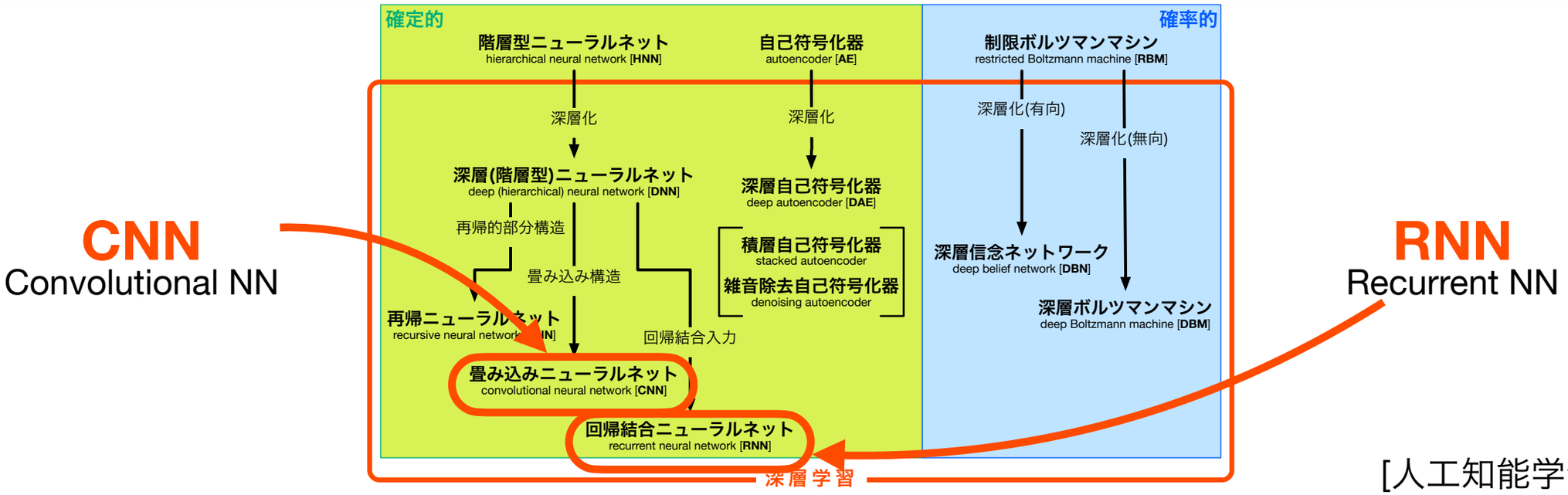
層の数が多いニューラルネットワークと大量データによる性能向上

- ニューラルネットワークは20年ほど氷河期にあり，有望視されていなかったが，その間も Hinton らは地道に研究を続けていた
- ReLU，DropOut，Adamなどの要素技術による改良

深層学習の成果

- 2011年に音声認識分野では注目されかけていたが，決定的だったのは一般画像認識のコンテスト ILSVRC2012 での突出した成果
- 音声認識・画像認識の分野では，従来のマルコフモデルやSIFT特徴量に基づく方法を駆逐した
- 自然言語処理でも，単純な方法にもかかわらず従来手法と同等の性能を達成

2015年時点での深層学習



[人工知能学会 15]

深層学習には多くのタイプがあるが，現在は CNN と RNN が活躍

- モデル自体は新しくはないが，要素技術による改良が効いた

第2次ニューラルネットワーク黄金期から加わった要素技術

- **DropOut**：経験にとらわれすぎる「過学習」問題への対処
- **GPGPU・並列計算**：大規模計算をできるハード・基盤ソフト
- **オンライン学習**：00年代に発展した大量データ用のアルゴリズム
- **活性化関数**：ReLU や MaxOut などの深層NNの勾配消失問題対策

深層学習が活躍している分野

画像認識

- 一般物体認識：画像中に写っている対象物が何であるかを識別
- 画像の変換：入力-出力の対から，変換規則を学習
- 画像と音声・文の対応付け：適切な説明文や音声を生成

音声認識・音声合成

- 音声認識：音声を自然言語文に変換する
- 音声合成：自然言語文から音声を生成する

自然言語処理

- 文の生成を伴う処理で非常に有効
- 品詞タグ付け（品詞を推定）チャンキング（句にまとめる）固有表現（日時や固有名詞などの抽出）構文解析（係り受け関係）
- 機械翻訳，会話生成

ニューラルネットワークの歴史

1943 現在のニューラルネットの基本単位であるMcCulloch-Pittsモデル
1958 パーセプトロンと誤り訂正学習則により第1次黄金期に
1969 Minsky らのパーセプトロンの限界の指摘で第1次氷河期に
1980 現在のCNNの源流である福島のネオコグニトロン
1986 バックプロパゲーション (BP) という多層NNの学習手法で第2次黄金期
1989 ネオコグニトロンとBPを組み合わせた現在のCNNをLeCunらが開発
1989 Waibelによる時系列データを扱う時間遅れNNの提案
1990 現在のRNNの源流であるElmanネットワークの提案
1995 ALVINN：ニューラルネットによる公道走行実験の成功
1995 Vapnikらによるサポートベクトルマシンの開発でNNは第2次氷河期
1997 現在のRNNの主流であるLSTM法の提案
2006 現在は使われなくなったが事前学習による多層NNの学習の提案．深層学習の始まりとされる
2012 ILSVRCで劇的な結果を収めたことで注目を集め第3次黄金期に突入
2014 過学習に対処するための手法 DropOut の提案

http://jsai-deeplearning.github.io/support/nnhistory.pdf

脳科学との関係

人工ニューラルネットワークと脳科学の関係点

McCullough-Pitts モデル

- 脳内のニューロンのシナプスの信号処理を参考にした
- 入力信号の線形結合 + 活性化関数 という人工ニューロンの多層化

モデルネオコグニトロン・畳み込みニューラルネット

- 人間の視覚信号の処理を参考にし，ある範囲の情報をまとめていく畳み込み構造を特徴とする
- $V_1 \rightarrow V_2 \rightarrow V_4 \rightarrow$ITなどの各視覚野と似た信号が獲得できるとの報告

↓

- 少なくとも機械学習の分野では，これらの脳との関連は重視されず，純粋な数理モデルとして扱われている
- ILSVRC2012のAlexNetは8層 ➡ 2015年のMSRAでは150層 ともはやだいぶかけ離れている

深層学習は本物か？誇張か？

Deep Learning: does reality match the hype?

Yes, DL advances are real and likely to lead to true Artificial Intelligence	20%
Partially, DL advances are real but the reality does not match the hype	58%
No, DL is mostly hype	14%
Not sure (52)	8%

投票総数：634

データサイエンティストの多くにとっての認識では，深層学習による進展は確かにあるが，それだけでは不十分である（2016年2月）

http://www.kdnuggets.com/2016/02/deep-learning-not-enough.html

深層学習推進派の意見

ICML2015の肯定派5人によるパネル

まとめブログ（邦訳）

音声認識・画像認識の次に深層学習が活躍するのは？

- 自然言語処理，ヘルスケア，ルールベースAIの領域

インダストリの計算資源にアカデミアが追いつかない問題

- インフラのオープン化が進むだろう

バブルと3度目の氷河期の有無

- 資金獲得，投資先，ジャーナリストの暴走で，評価が過剰 (overhype) になっていると
- もう社会で使われるステージになっているから冬は来ないだろう
- 研究者が提案アプローチの限界を論文で述べるなどして，これらを抑制していけば，期待と実際の差は小さくしていけるだろう

深層学習懐疑派の意見

- 前回の冬の原因である**調整の難しさは解消していない**
- サポートベクトルマシンのような**解の最適性保証はできない**
- 深層学習は特徴抽出が可能だといっても，全体でみれば非線形関数だから**深層学習と同等のことは他のモデルでもいずれはできるのでは？**
- **データに依存した「統計量」と，普遍性のある「知識」とは異なる**

[Bottou 2015, p.55あたりから]

  - 同じ一般画像認識でも違うデータ集合で訓練すると性能がでない
  - 深層学習した結果をだます絵が簡単に作れる
  - 実際の運用でアフリカ系の人にゴリラと分類してしまった問題

機械学習アルゴリズムと特許

機械学習での特許：アルゴリズム単体での特許の影響力はあまりない

- 特許でのアルゴリズムは抽象的ではなく，具体的な実装に限られるので，部分的に変更して回避しやすい
- コミュニティとして理論部分への特許には疑念があり，パテントトロール対策としての特許の傾向

話題になった特許

- 推薦システム：US Patent 4870579 1987年などは，実用化されたころには無効に
- 画像認識のSIFT特徴量：公開ソフトウェア OpenCV では非商用のみの利用となり，その後，若干違う特許フリーの手法も開発も導入
- 頻出パターンマイニング（IBM），自然言語処理の word2vec，深層学習のDropOut や バッチ正規化 （Google）などがあるが，どちらかといえばトロール対策なのでは？

人工知能

人工知能とは

What is Artificial intelligence

John McCarthy

http://www-formal.stanford.edu/jmc/whatisai/whatisai.html

日本語訳： http://www.ai-gakkai.or.jp/whatsai/Alfaq.html

It is the science and engineering of making intelligent machines, especially intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI does not have to confine itself to methods that are biologically observable.

知的な機械，特に，知的なコンピュータプログラムを作る科学と技術である．人の知能を理解するためにコンピュータを使うことと関係があるが，自然界の生物が行っている知的手段だけに研究対象を限定してはいない．

人工知能についての誤解

「人工知能」というモノがあるわけではない

自動車：人間の足の走るという機能の代わりをするモノ

人工知能：人間の脳の考えるという機能の代わりをするモノ

↓

実際の人工知能

モノではなく，何らかの知的処理を行う**人工知能技術**

＋

多くの**関連技術の集合体**

人工知能活用の利点

人工知能・機械学習技術の活用

- 今までのものごとのやり方を，人工知能を使ったものに置き換える

人工知能・機械学習技術活用の利点

多くのデータや多様な要素を考慮できる

- **Alpha GO**：盤面上の多様な要素と盤面の形勢との間の対応関係を，人類が今までに行った以上の数の対戦履歴から見いだした

非常に膨大な情報の中から目的の情報を素早く発見できる

- **Alpha GO**：どの手筋が良さそうかの判断についても対戦履歴を利用
- **材料科学**：材料の試験には時間も費用もかかるが，より有望な材料を戦略的に探索できる

⬇

- 勘で偶然にうまくいくことはあっても，長期的には人工知能技術を活用した方が効率的

機械学習技術の運用上の制限

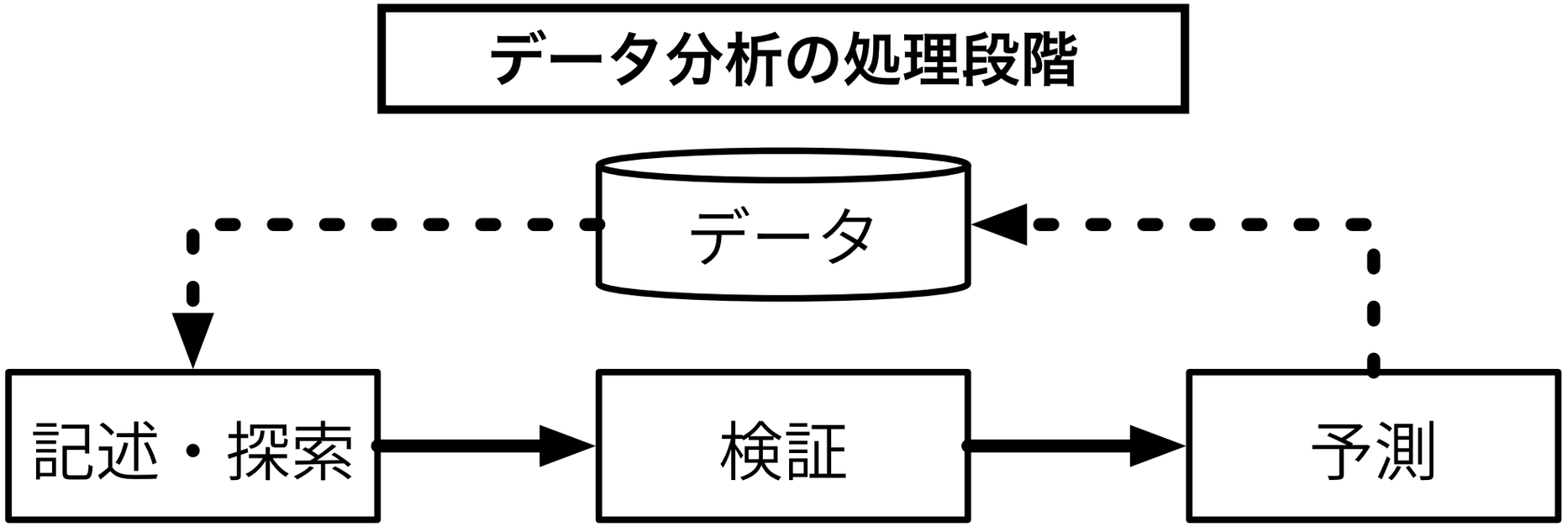
データ分析の役割

アルゴリズム分野と並んで知的処理の基盤になる**要素技術**

↓

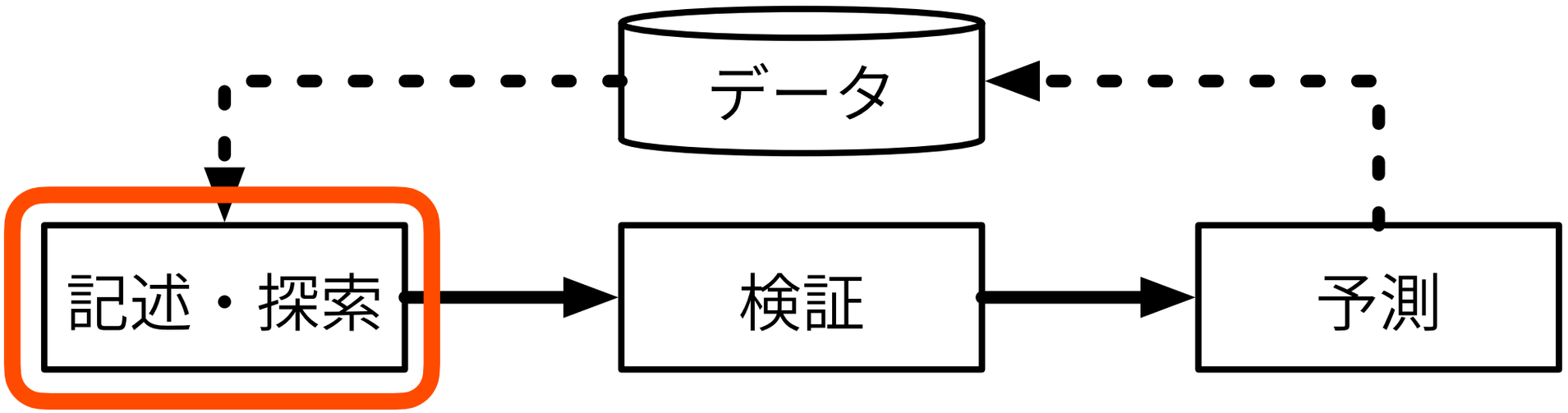
必ず他の技術との**連携して利用される**

連携する分野の知識（**ドメイン知識**）が必要



各段階の目的に応じて適切な処理手法を選ぶ必要

データ分析の処理段階：記述・探索

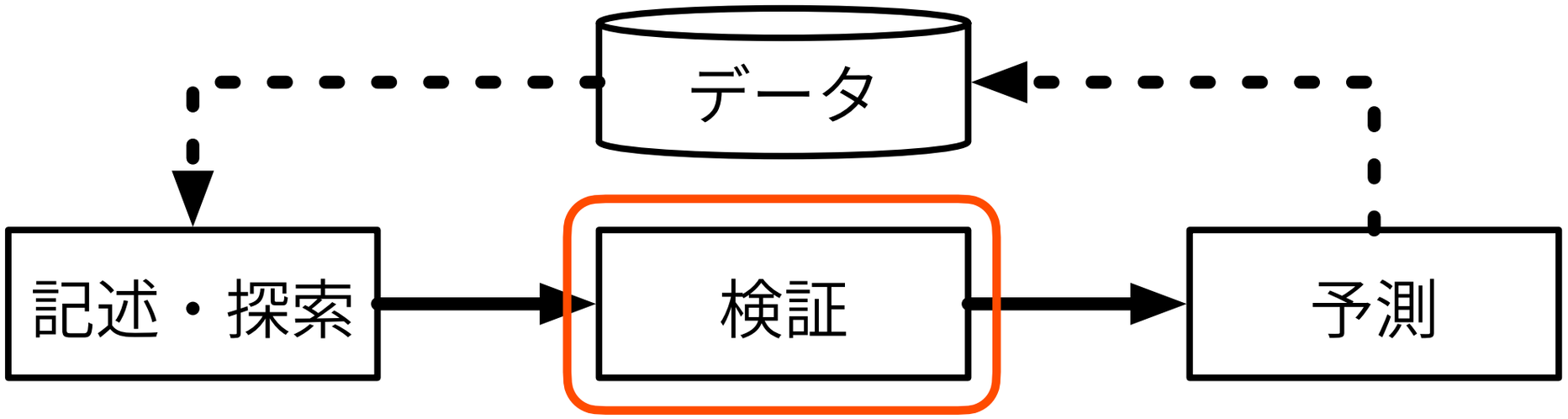


記述・探索：取得してきたデータを俯瞰して把握する

後の検証や予測の段階での処理を容易にしたり，データの表す事象についての仮説を立てたりする目的

- **記述統計**：データの平均を計算するなどの単純なもの
- **可視化手法**：直感的に把握できるようにグラフなどで図示
- **クラスタリング，次元削減，頻出パターンマイニングなど**：各種の探索的な分析手法

データ分析の処理段階：検証

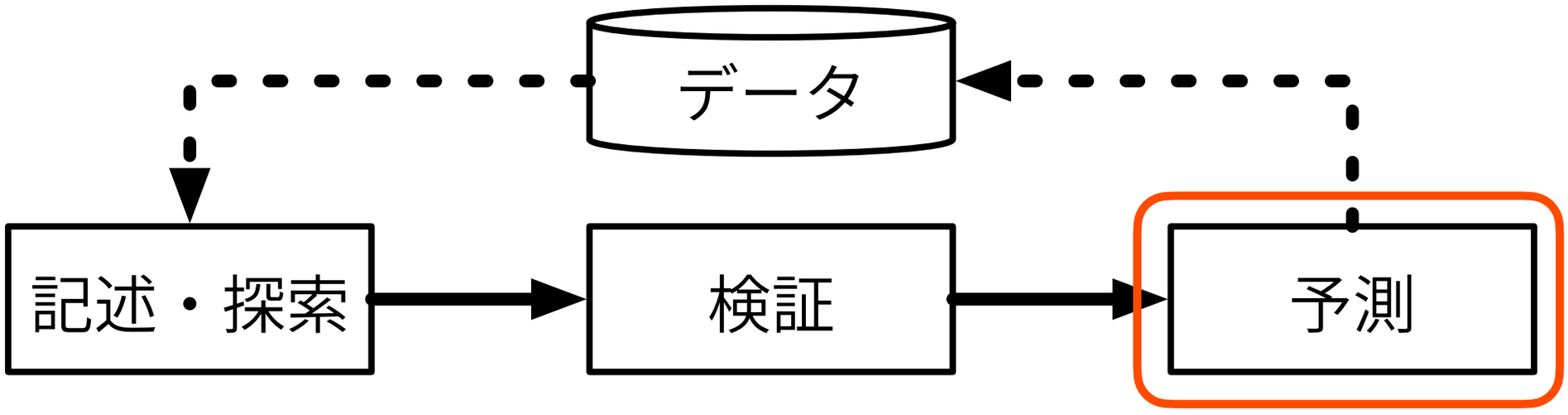


検証：仮説の妥当性を検証する

平均の差や独立性など特定の性質がデータにあるかどうかといった仮説を検証する

- **統計的仮説検定**：仮説検証を目的とした手法
- **因果推論**：因子間の因果関係の強さなどを詳細に検証

データ分析の処理段階：予測



予測：未観測の事象の状態を推定する

明日の気温やモノの種類など未観測の事象がどのような状態になるかを予測する

- **回帰分析など**：基本的な統計的予測手法は予測の他にも，部分的に検証にも用いられる
- **機械学習**：統計的予測手法より予測に特化した手法

運用の難しさ：目標の定式化

オズの魔法使い問題：**実世界の目標の達成度を示す指標**を最適化

⬇

機械学習の利用者が，**実世界での目標を把握**していなければならない

目的が達成されたときに，良くなるような**指標を定式化する必要**

目標：Web広告の収入増 ➡ **指標**：クリック率

- 推薦システム：利用者の嗜好の予測精度が上がっても，本当に改善したかった利用者の満足度は必ずしも向上しない [McNee 06]
- 数週間のデータに基づいてクリック率を最適化しても，それが長期にわたるクリック率の向上に必ずしも繋がらない [Kohavi 15]

⬇

運用しながら実世界の目標を日々明確にしてゆき，それが達成されるように指標を調整する

運用の難しさ：不良設定問題

機械学習は逆問題で，解が解けたかどうかが不明確な**不良設定問題**

⬇

学習したルールの**挙動は確定的ではなく確率的**

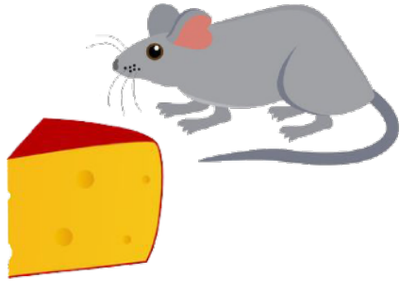
[Bottou 15（連結成分の図の原典は Minsky & Papert, Perceptorons, 1968）]

連結成分

形式的に解ける良設定問題

アルゴリズム論

⬌



「ねずみ」っぽさ

「チーズ」っぽさ

例示しかできない不良設定問題

機械学習・データマイニング

- 実世界の不良設定問題を形式世界の問題として解く機械学習の宿命
- アフリカ系の人の写真をゴリラと識別した社会問題 [Barr 15]

⬇

制御できない部分が残ることを前提とした運用上の工夫が必要

運用の難しさ：適切な訓練データ

現状の機械学習では，背景知識からの演繹ではなく，

データからの帰納にほぼ完全に依存している

⬇

実世界での**目標に過不足のない情報**を含んだ訓練データ

リーク (leakage)：運用時には利用できない情報を学習に使ってしまい，実際には学習結果を運用できない [Perlich+ 11]

- 販売の成功予測に販売員名の情報を使う
  - ➡ 販売員を割当済みな時点で顧客はすでに購入意思があり，無意味

標本選択バイアス：運用時の対象と，訓練データ用の対象が不一致

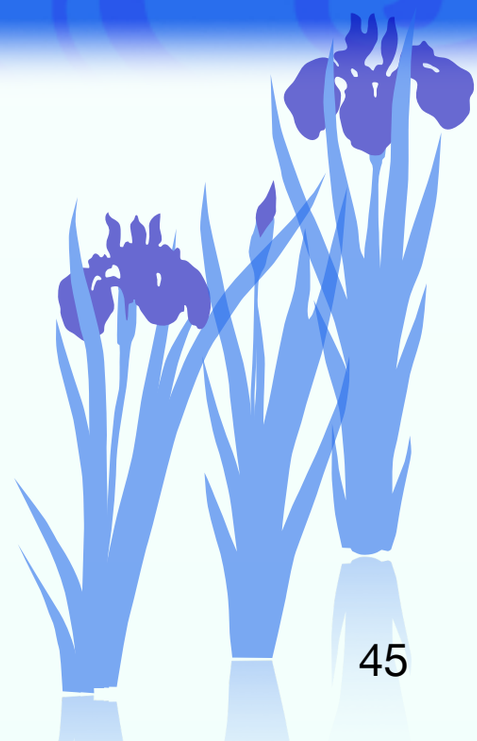
- 疾患の予測について，ある病院のデータから学習したが，入力する検査結果の手順や薬品の違いで，別の病院の患者では精度が低下

⬇

訓練時と運用時の性能指標をみながら，機械学習の利用者が，データや情報を実世界の目標に合わせて逐次的に取捨選択する

第II部

機械学習・データマイニングの基本原理



人工知能技術・知的システム

論理的推論

少なくとも，次のいずれかの論理的推論の一つを計算機上で行うのが，人工知能技術（弱い人工知能）や知的システムといえる

論理学における論理的推論は3種類

演繹	帰納	アブダクション
deduction	induction	abduction
ソクラテスは人間 全ての人間は死ぬ ⬇ ソクラテスは死ぬ	ソクラテスは死ぬ ソクラテスは人間 ⬇ 全ての人間は死ぬ	ソクラテスは死ぬ 全ての人間は死ぬ ⬇ ソクラテスは人間
特殊な結論	**一般的な規則**	**仮説・説明**

※ 帰納とアブダクションの違い：帰納ではソクラテスのことを人間全体に一般化しているが，アブダクションではソクラテスについての言及のままで，参照している対象は変わっていない

論理的推論

[Michalski 93]

演繹 (deduction)

機械学習での推論段階

ソクラテスは人間　$a \in X$

全ての人間は死ぬ　$\forall x \in X,\ q(x)$

ソクラテスは死ぬ　$q(a)$

帰納 (induction)

機械学習での学習段階

ソクラテスは死ぬ　$q(a)$

ソクラテスは人間　$a \in X$

全ての人間は死ぬ　$\forall x \in X,\ q(x)$

演繹では順方向にたどる　　帰納では逆方向にたどる

前提 + 背景知識 ➡ 結論

帰納的な学習でも，どう一般化するかは，前提に依存している

⬇

人間・計算機に関わらず，帰納的推論でも純粋には客観的ではなく，
どのような前提に基づいた結果なのかは知っておく必要

機械学習の基本原理

不可能性

不可能：現状で支持されている物理法則に反しているか，形式的な証明によって否定的に解決されている

- **永久機関**：エネルギー保存則に反している
- **Arrowの不可能性定理**：四つの民主的な基準を満たす意思統合手段は存在しえない

非常に困難：実現を否定する理論や法則はないが，技術・経済面に課題

- **同様の意味**：理論上は可能，実現は困難，技術/経済的に困難
- **核融合炉，軌道エレベータ，錬金術**

未解決：悪魔の証明になったりして証明が不可能だったり，証明や問題が未解決だったりする場合

- **汎用人工知能**：唯物論的には可能だが，心身二元論的には不可能
- **タイムマシーン**：未解決問題

機械学習分野での不可能性に関わる三つの基本原理

機械学習の基本原理

機械学習分野での不可能性に関わる三つの基本原理

汎化誤差 (generalization error)

- この汎化誤差を小さくすることが機械学習の目標だが，それには観測できない情報が必要なので不良設定問題に

ノーフリーランチ定理 (no free lunch theorem)

- ありとあらゆる状況において，他のアルゴリズムの性能を必ず凌駕できるアルゴリズムは存在しえない

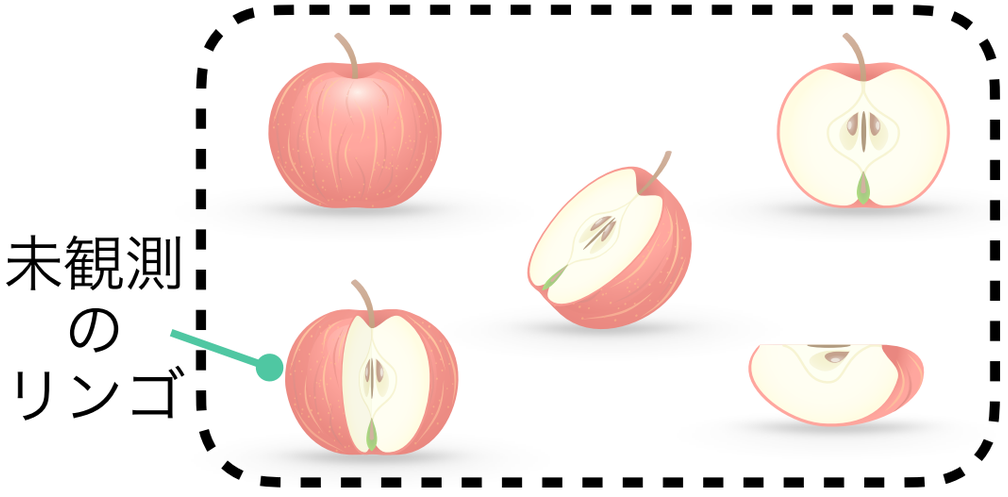
醜いアヒルの子の定理 (ugly duckling theorem)

- 対象を分類するときには，その対象のある側面を重視し，他の側面を軽視するということを伴う

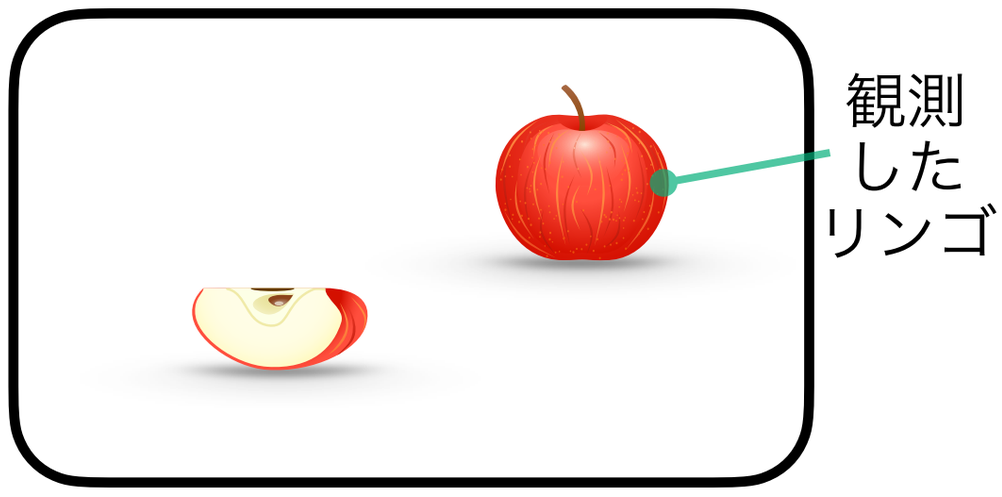
機械学習の基本原理として紹介したが，形式的証明に基づく不可能なので人間でも不可能

汎化誤差と経験誤差

見たこともないものも含めたリンゴ



実際に見たリンゴ



真のモデル
観測は不可能

訓練データ・標本
すでに観測済み

汎化誤差：未観測のものも含めた真のモデルから得た対象に対する予測の誤り

経験誤差・**標本誤差**：訓練データ中で観測済みの対象に対する予測の誤り

機械学習の目標：経験誤差ではなく，汎化誤差を最小にしたい

⬇

真のモデルの観測は不可能なので検証できない（不良設定問題）

不良設定問題に仮定を導入

不良設定問題：仮定を導入して解く

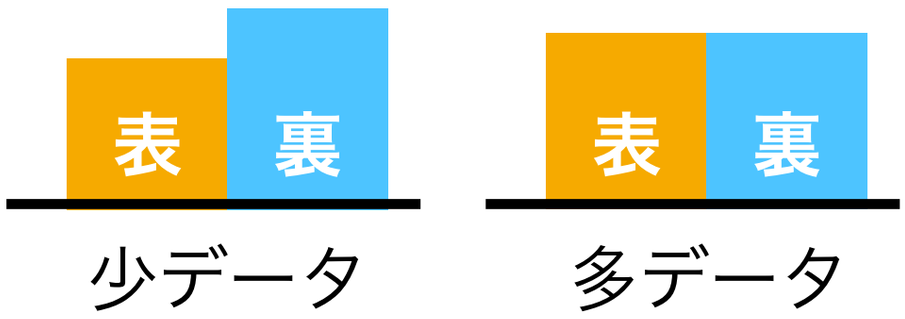
↓

もし仮定が現実と大きく異なる場合には機械学習は失敗する

漸近論：無限個の経験データでの経験誤差は，汎化誤差に一致

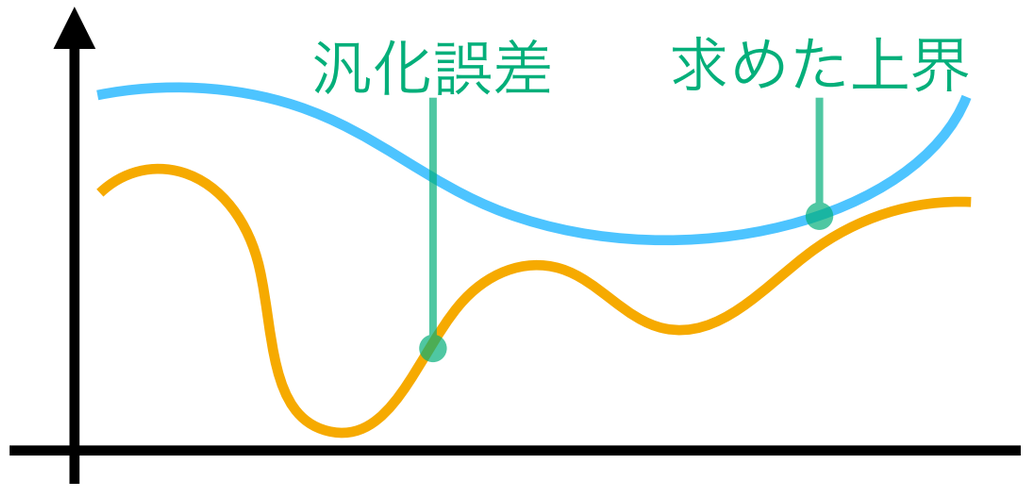
同一同分布から得た訓練データからの予測値は，データ数が増えるに従って真のモデルのそれに一致

コイン投げ



経験リスク最小化：リスクの上界と汎化誤差の最小値は一致

経験リスク：訓練誤差から，汎化誤差そのものは見積もれないが，その上界は計算できる



ノーフリーランチ定理

[Wolpert 96]

ノーフリーランチ定理：全ての分類問題を考えたとき，どのようなアルゴリズムも平均的には，その汎化誤差に関して事前の差はない

⬇

あるアルゴリズム A がある予測問題で，アルゴリズム B より汎化誤差に関して性能が良かったとしても，アルゴリズム B が A より良くなるような別の予測問題が存在する

- どのアルゴリズムも他より常によいということはありえない
  - ➡ **多くの機械学習アルゴリズムが考案されている理由**
- 事前には差がないということは，解こうとする問題についての情報が少しでもあれば，それを活用したアルゴリズムが有利になる
  - ➡ **いろいろな状況に合わせたアルゴリズムの構築や選択が重要**

ノーフリーランチ定理：詳細

http://no-free-lunch.org/

教師あり学習でのノーフリーランチ定理 [Wolpert 1996]

- **前提**：ノイズなし，損失は誤分類誤差
  - d：訓練データ集合，m：訓練データ数
  - f：目標である真のモデル，h：学習アルゴリズムが出力する仮説モデル
  - C：訓練データにない事例に対する誤差（汎化誤差）
- $\mathrm{E}[C \mid d]$，$\mathrm{E}[C \mid m]$，$\mathrm{E}[C \mid f, d]$，$\mathrm{E}[C \mid f, m]$ のいずれを誤分類尺度として採用しても，いかなるアルゴリズムの性能も平均的には同等である

探索と最適化でのノーフリーランチ定理 [Wolpert+ 1997]

- **前提**：有限空間，データの再サンプリングなし
- コスト関数を最大・最小化する最適化アルゴリズムは，全ての可能なコスト関数を考えたとき，全てのアルゴリズムの性能は同等

醜いアヒルの子の定理

[Watanabe 69]

醜いアヒルの子の定理：対象を表現している全ての特徴を同等に重要とみなす限り，純粋に形式的な観点では，他より類似している対象の集まりというものは存在しえない

⬇

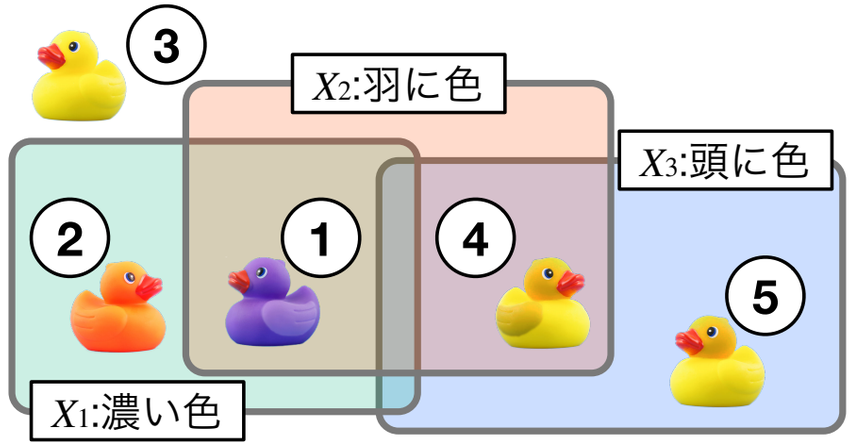
類似した対象が集まったクラスというものを実世界で見いだしているならば，対象のある特徴を重視したり，逆に軽視したりしているということである．そして，どの特徴を重視したり軽視したりするかは形式的な判断の範疇の外で決めている

- 予測問題にとって重要な特徴は限られているという仮説を支持
  - ➡ 次元削減，特徴選択，正則化の技法などが有効である理由

醜いアヒルの子の定理：詳細

醜いアヒルの子の定理：n 個のブール特徴を使って対象を表現し，これらの特徴を用いた全ての可能な命題論理式の集合を考える．
このとき，一対の異なる対象 $\mathbf{x}_a$ と $\mathbf{x}_b$ が同時に満たす論理式の数は，対象の対の選び方によらず一定である．

醜いアヒルの子 ① は，特徴 X_1 と X_2 は真だが，
特徴 X_3 は偽．論理式では $X_1 \wedge X_2 \wedge \bar{X}_3$



①と②のアヒルとそれ以外のアヒルとを異なるクラスに分類する命題論理式は，例えば

$$(X_1 \wedge X_2 \wedge \bar{X}_3) \vee (x_1 \wedge \bar{X}_2 \wedge \bar{X}_3)$$

を満たすかどうかで分類できる．3個の特徴で8種類のアヒルがいるとき，一対のアヒルを識別する可能な命題論理式の数は 2^{8-2} 個で一定．

ここで，一対のアヒルがどれだけ似ているかを，これらのアヒルを同じクラスに分類する命題論理式の数で定義すると，この数は対の選び方によらず一定になる．

この結果，全ての特徴を同等に扱うと醜いアヒルの子を他のアヒルと区別できない．

モデル

特徴

特徴 (feature)・属性 (attribute)：予測対象をある側面で見たときに，どのような状態にあるのかを表すもの

例：　予測対象　　カテゴリ・名義　　数値　　順序

色 → 赤　　重量 → 200g

色 → 黄色　　重量 → 300g

私の嗜好
↓
バナナが好き

特徴・属性　　**特徴値・属性値・値**

特徴ベクトル・属性ベクトル：予測対象を記述する特徴（属性）をベクトルの形式にまとめたもの

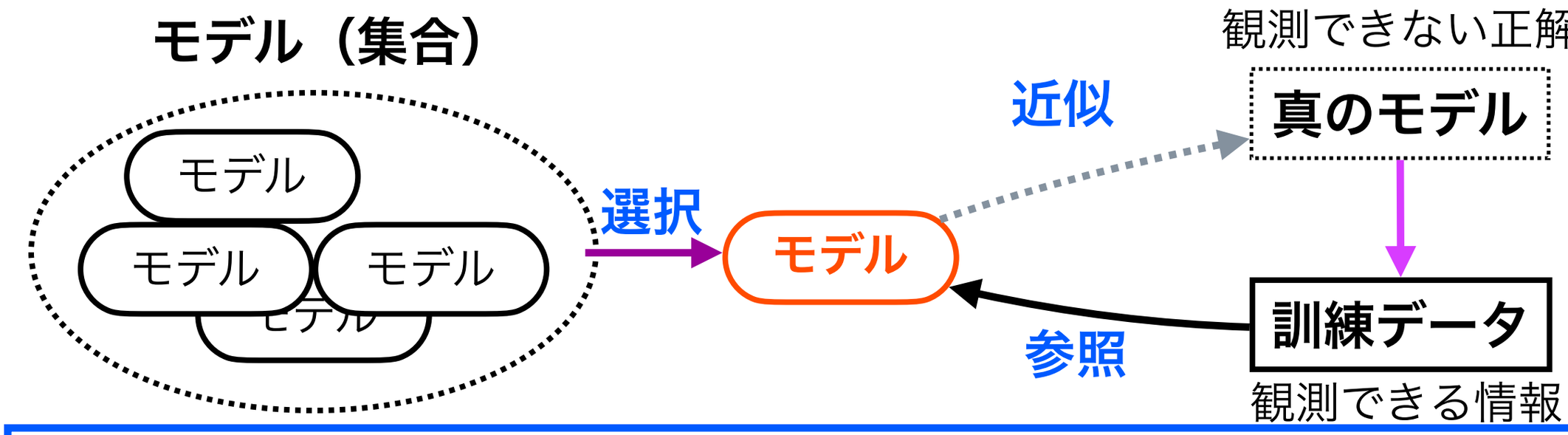
リンゴ ➡ (色→赤，香り→あり，重量→300g，…，高さ→7cm)

$$\mathbf{x}_i = (x_{i1} \quad x_{i2} \quad x_{i3} \quad x_{im})$$

統計的な手法ではカテゴリ値も数値に変換して，空間的中の点と見なす

モデル

モデル (model)・仮説 (hypothesis)：入力される予測対象から，予測結果への対象対応を記述する写像，もしくはそれらの候補集合で，数学的な関数か論理式を用いて表現する



機械学習での学習：訓練データを参照して，モデル集合から真のモデルを最もよく近似すると思われるモデルを選択する

モデル集合は，真のモデルと一致するものを，多くの場合含んでいない

↓

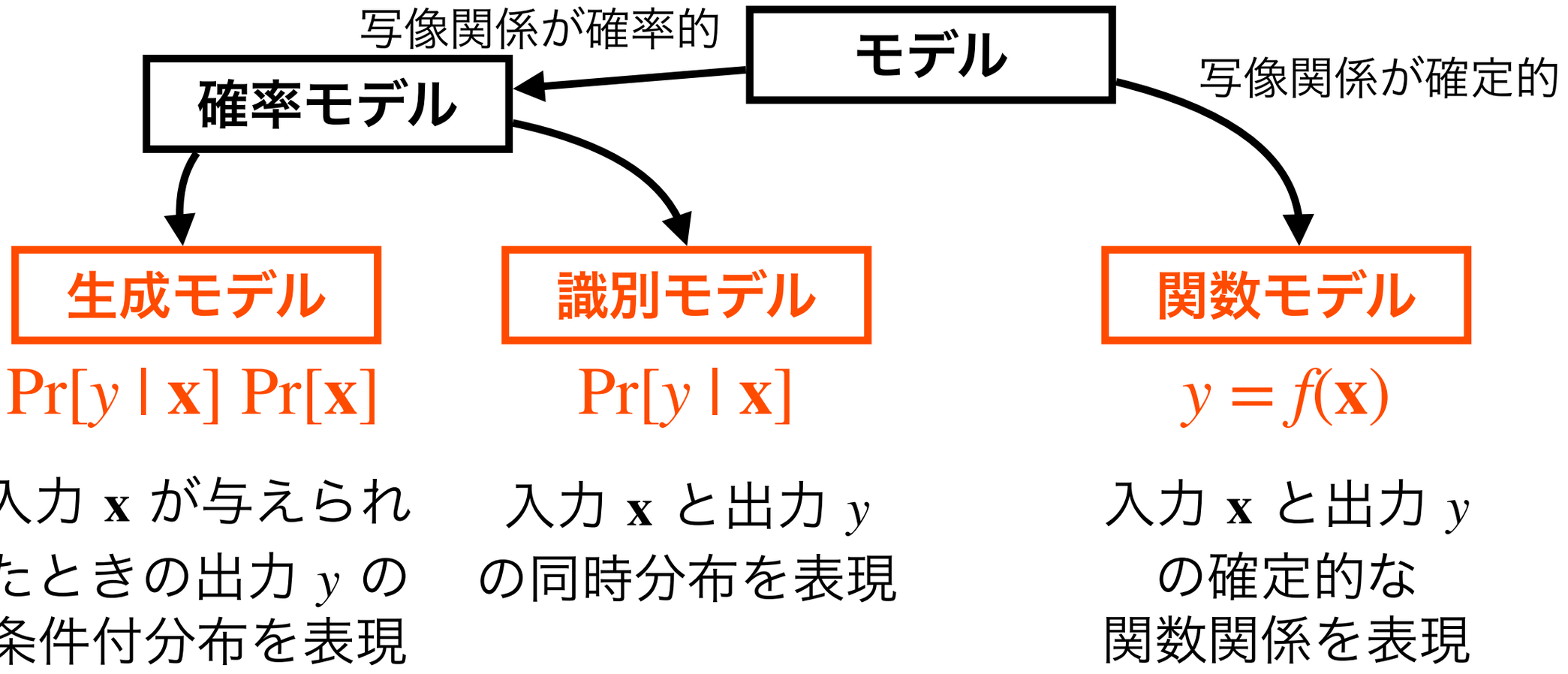
どのモデルもある意味「にせもの」ではあるが，なにかしらの役に立つ

"Essentially, all models are wrong, but some are useful" — *George E. P. Box*

生成モデル・識別モデル・関数モデル

入力（予測対象）$\mathbf{x}$ と出力（予測結果）y との写像関係の分類

モデル表現する方針の違い



識別モデルと生成モデルでは識別の方がやや高精度といわれることもあるが，ノーフリーランチ定理により基本的には優劣はない

モデルの複雑さ

モデルの複雑さ：予測対象と予測結果の写像関係をより詳細で複雑に記述できるかどうかの度合い
モデル集合が複雑であるほど，一般により多くの訓練データが必要

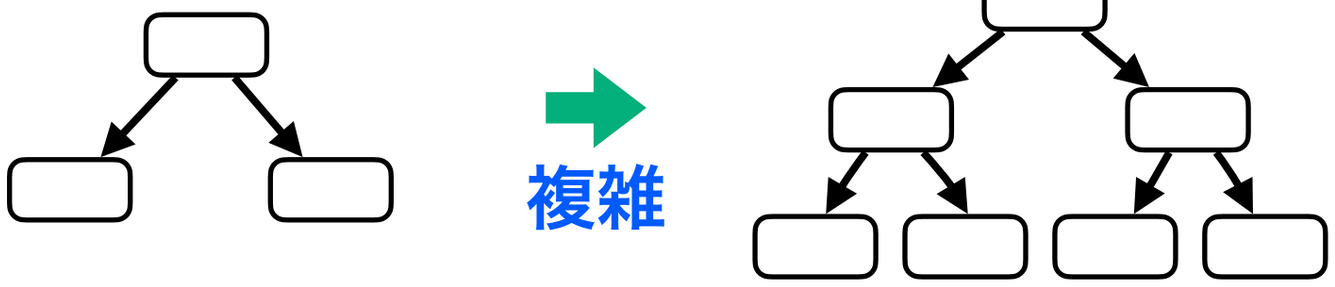
写像関係を多項式で表した例：

$$y = w_1 x_1 + w_2 x_2 + b \quad \Rightarrow \text{複雑} \quad y = v_1 x_1^2 + v_2 x_2^2 + w_1 x_1 + w_2 x_2 + b$$

高次の多項式の方がより複雑な写像関係を表現できるだろう

➡ 高次の多項式の方が，1次式より複雑なモデル

写像関係を決定木で表した例：



段数の大きな深い決定木の方がより複雑な写像関係を表現にできる

➡ 段数の多い定木の方が，少ない決定木より複雑なモデル

パラメトリック・ノンパラメトリック

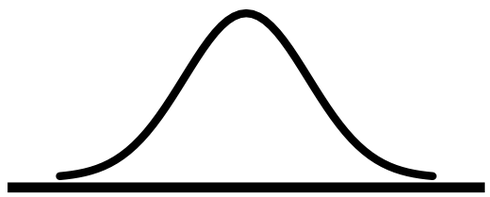
入力と出力の対応関係を表現する方針による分類

パラメトリックモデルでは，パラメータによって完全に分布や関数の形状が決定されるが，**ノンパラメトリックモデル**では基本的にデータからその形状が決まり，パラメータが決めるのは滑らかさに限定

※　定義や見解に幅のある用語だが，ここでは [Bishop 06, Bishop 07, 2.5節] に従った

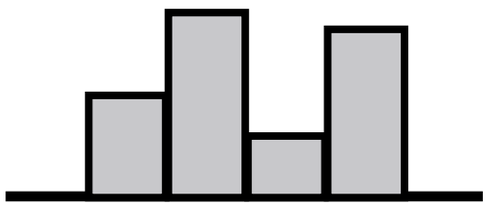
パラメータ：モデルの分布や関数の集合のうちの一つを指定するための入力で，他の入力と区別して $\Pr[y \mid \mathbf{x}; \boldsymbol{\theta}]$ や $y = f(\mathbf{x}; \boldsymbol{\theta})$ などとも表記

パラメトリック



訓練データ数と無関係に，ガウス分布の形状は平均・分散パラメータで決定

ノンパラメトリック



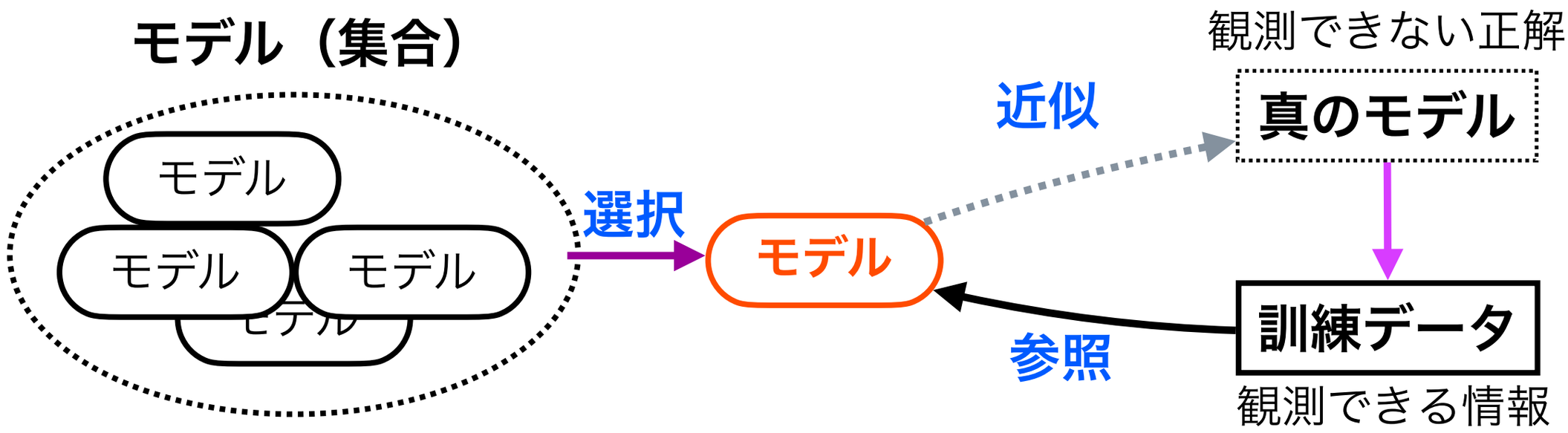
ヒストグラムではビン数パラメータで滑らかさのみが決まる

一般にパラメトリックよりノンパラメトリックの方が複雑なモデル

データからの学習

最尤推定

機械学習での学習：訓練データを参照して，モデル集合から真のモデルを最もよく近似すると思われるモデルを選択する



どのような「近似」をするかで，いろいろな学習方法がある

最尤推定：確率的なモデル集合の中から，訓練データが発生する確率が最も高いモデルを選択する学習の基準

最も基本的な推定方法でよく使われている

KLダイバージェンス

最尤推定で，真のモデルを最も近似するモデルが選択できる

- 真のモデルの分布を $\Pr^*[X]$ と，あるモデルの分布を $\Pr_\theta[X]$ と表記
- 真のモデルから得た訓練データ $X_i \sim \Pr^*[X]$ の集合 $\mathbf{X} = \{X_1, X_2, \ldots, X_n\}$

訓練データでの平均は，真のモデル上の期待値に漸近的に一致

尤度の対数

$$\frac{1}{n}\sum_{i=1}^{n} \log \Pr_\theta[X_i] \xrightarrow{n} \mathrm{E}_{\Pr^*[X]}[\log \Pr_\theta[X]]$$

大数の法則

真のモデル上の期待値

真の分布の対数の期待値と尤度の対数の平均と差で近似精度を測る

$$\mathrm{E}_{\Pr^*[X]}[\log \Pr^*[X]] - \frac{1}{n}\sum_{i=1}^{n} \log \Pr_\theta[X_i]$$

$$\xrightarrow{n} \mathrm{E}_{\Pr^*[X]}[\log \Pr^*[X] - \log \Pr_\theta[X]]$$

$$= D_{\mathrm{KL}}\left(\Pr^*[X] \,\|\, \Pr_\theta[X]\right) \geq 0$$

KLダイバージェンス

尤度が大きくなると真の分布へのKLダイバージェンスは小さくなる

推定の種類

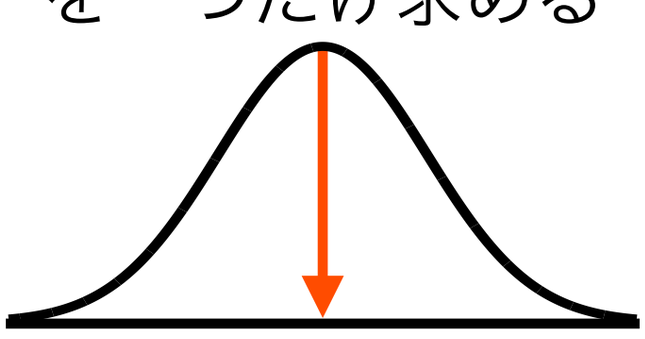
推定した予測値やパラメータを求める形式の種類

※　予測値：新規の入力に対する出力

※　パラメータ：モデル集合から特定のモデルを指定するもの

点推定

最も確実性の高い値を一つだけ求める

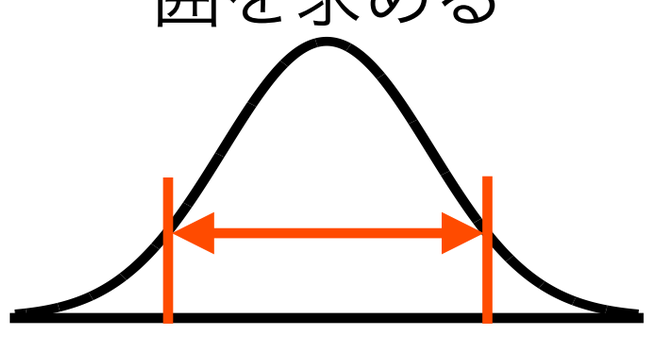


$$\hat{\theta} = \theta^*$$

最尤推定やMAP推定など，分布や関数の最頻値を使うもので，最もよく利用される

区間推定

推定値が存在する範囲を求める

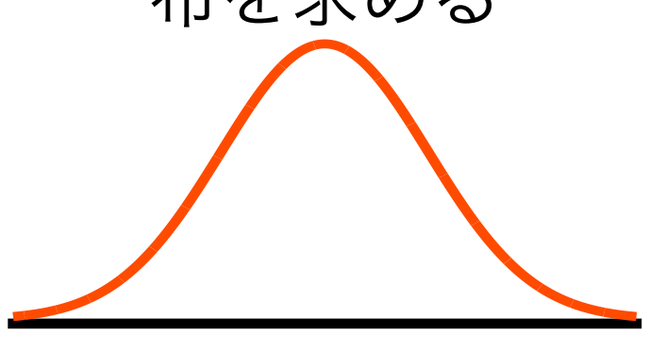


$$\hat{\theta} \in [\alpha, \beta]$$

この範囲に推定値が存在する確率が95%といった形で統計分野でよく利用される

分布推定

推定値が存在する分布を求める



$$\hat{\theta} \sim \Pr[\theta]$$

事前分布を導入した生成モデルと組み合わせてベイズ推定などで利用される

過学習（過剰適合，過適合）

過学習（過剰適合，over-fitting）：訓練データに合わせ過ぎたモデルを選択したために，経験誤差は小さいが，汎化誤差は大きくなり，本来の学習の目的を達成できていない状態

- **経験誤差**：訓練データに対する予測誤差
- **汎化誤差**：真のモデルからのデータに対する予測誤差

モデルA

例外的な訓練データにも細かく合わせた規則

経験誤差 ➡ 小　汎化誤差 ➡ 大

モデルB

例外的なデータは無視した簡潔な規則

経験誤差 ➡ 小　汎化誤差 ➡ 大

複雑なモデルを選択することで，経験誤差だけを小さくした過学習の状態にすると，汎化誤差の小さな望ましいモデルは得られない

バイアス・バリアンス

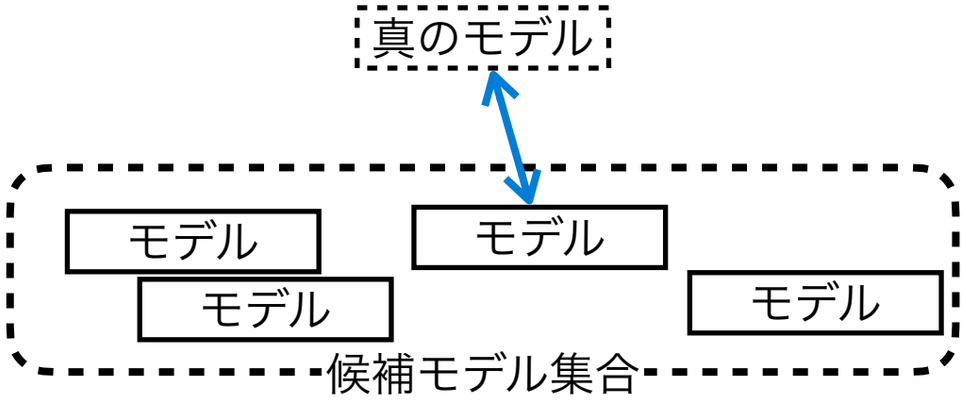
汎化誤差は，バイアス，バリアンス，そしてノイズの三つに分割できる

※ ノイズ：モデル集合の選択に依存せず，本質的に減らせない真のモデルのばらつき

汎化誤差 ＝ バイアス ＋ バリアンス ＋ ノイズ

バイアス

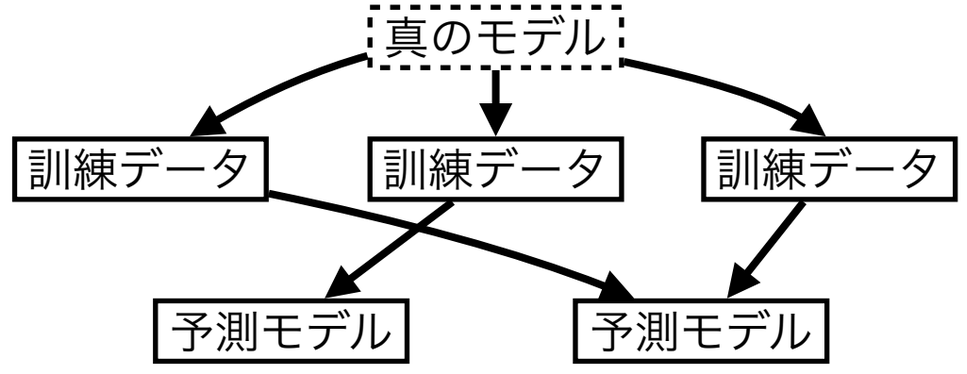
候補モデル集合に真のモデルは含まれないことで生じる誤差



単純なモデル集合ほど大きくなる

バリアンス

訓練データが異なると，異なる予測モデルが選択されることで生じる誤差



複雑なモデル集合ほど大きくなる

バイアスとバリアンスは同時には小さくできない

↓

バイアスとバリアンスのバランスをとって全体の誤差を小さくする

第III部
機械学習・データマイニング研究
の諸問題

モデルのグループ

モデルのグループ

[Domingos 15, Domingos 21]

Domingos が著書「The Master Algorithm」で論じた，モデルの違いに基づく機械学習の研究グループ分類

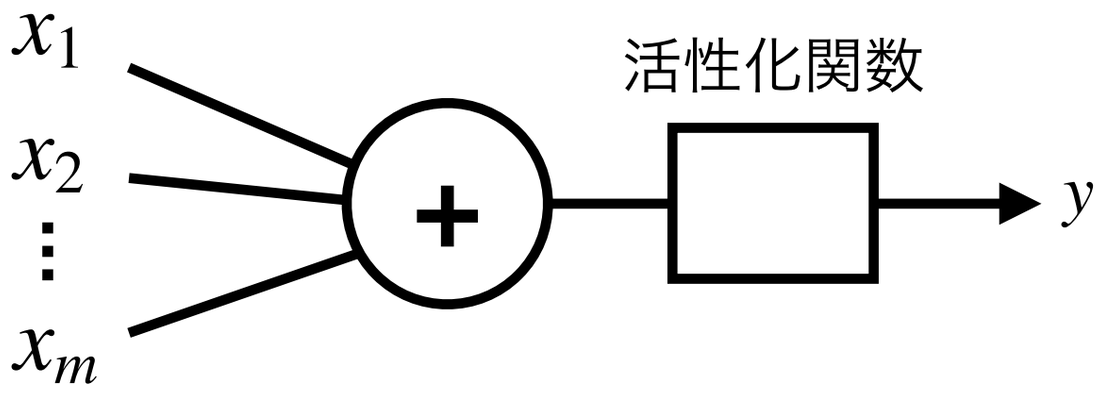
- **Symbolists（記号主義者）**論理式をベースにしたモデルを利用
- **Connectionists（コネクショニスト）**脳の神経細胞の仕組みを参考にしたニューラルネットを使ったモデル
- **Evolutionaries（進化主義者）**生物の遺伝の仕組みを参考にして，適切なモデルを学習する
  - ➡ モデルより最適化の方針なのでここでは除外
- **Bayesian（ベイズ主義者）**生成モデルと事後確率推定を利用する
- **Analogizers（類推主義者）**対象の類似性に基づくモデル
  - ➡ ここでは基本の統計系のモデルとあわせて統計・カーネルとする

※ これらの方針は明確に分かれるものではなく，重複する部分も多い

ニューラルネット

ニューラルネット (neural network)：脳の神経細胞の仕組みを参考にした基本単位を組み合わせたモデルを用いる方法

McCulloch-Pitts の神経細胞のモデル



- McCulloch-Pittsモデルをはじめとして基本単位，活性化関数，およびこれらの接続方法などに様々な選択肢がある
- 黄金期と氷河期を繰り返し，2012年以降は「深層学習」により3度目の黄金期を迎えている

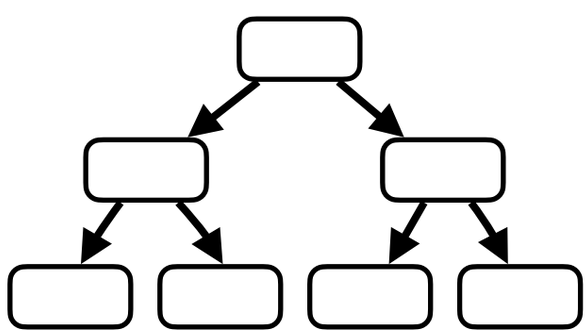
各種のニューラルネットモデル

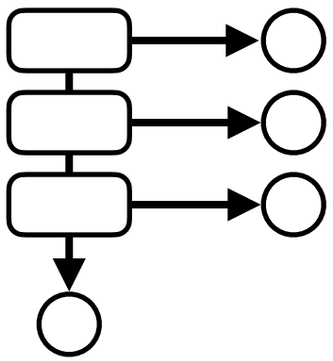
モデルの複雑さ	生成モデル	識別モデル	関数モデル
簡潔			パーセプトロン
	制限ボルツマンマシン		
複雑	ボルツマンマシン		階層ニューラルネット
			回帰結合ニューラルネット
超複雑	深層ボルツマンマシン 変分自己符号化器 拡散モデル	深層学習	畳み込みニューラルネット 深層ニューラルネット transformer

ルールベースモデル

ルールベース：論理式をつかった条件判断を使ったモデル

決定木・決定リスト：条件判断と分岐を繰り返すモデル





命題論理：単体の分類対象についての記述

述語論理：対象間の関連についても記述できる

- 決定木はアンサンブル学習と結びつき，勾配ブースティング木の著名な実装 xgboost で広く使われている
- 確定的な命題論理・述語論理は統計的機械学習以前の主流だった
- 命題論理・述語論理に確率的な要素を導入したMarkov論理ネットなどが開発され利用されている

各種のルールベースモデル

モデルの複雑さ	生成モデル	識別モデル	関数モデル
簡潔		決定木 決定リスト	命題論理
複雑	確率論理		述語論理
超複雑		ランダムフォレスト 勾配ブースティング木	

ベイズモデル

ベイズモデル：パラメータの事前分布を導入した生成モデルと，ベイズ則と周辺化を用いて事後確率を計算するベイズ推定との組み合わせ

ベイズ則：パラメータの事前分布と尤度関数を事後分布に変換できる

事後分布

$$\Pr[\theta|x] = \frac{\Pr[x|\theta]\Pr[\theta]}{\sum \Pr[x|\theta]\Pr[\theta]}$$

事前分布

尤度関数

- 考え方としては古くからあったが，分布を推定するのは困難だったが，マルコフ連鎖モンテカルロ法と計算機によって計算が容易になって普及した
- 分布を推定するため予測の信頼度を得ることができる利点
- 自然言語処理で単語の情報を集約するトピックモデルや，確率過程を導入したノンパラメトリックベイズなどの進展があった

各種のベイズモデル

モデルの複雑さ	生成モデル	識別モデル	関数モデル
簡潔	単純ベイズ		
	階層ベイズ		
複雑	ベイジアンネット		
	一般のグラフィカルモデル		
	マルコフ確率場		
超複雑	ノンパラメトリックベイズ		
	ガウス過程		
	拡散モデル		

統計・カーネル

統計・カーネル：統計分野で長い間，厳密な分析が行われてきた線形モデルを，カーネルトリックによって複雑なモデルに変換可能に

カーネルトリック：二つの特徴ベクトルの内積を表すカーネル関数を導入することで，複雑な高次元モデルの明示的な計算を回避できる

$$y = \sum_i \alpha_i k(\mathbf{x}, \mathbf{x}_i)$$

カーネル関数

- 経験リスク最小化原理に基づくサポートベクトルマシンは，大域的な最適解を計算できる利点などから普及した
- カーネル関数の設計には自由度があり，目的に合わせて非常に多くのカーネルが提案された
- ベイズモデルにもカーネルは導入可能で，ガウス過程などの例

各種の統計・カーネルモデル

モデルの複雑さ	生成モデル	識別モデル	関数モデル
簡潔		ロジスティック回帰	線形回帰
複雑		サポートベクトルマシン その他のカーネル手法	
超複雑	ガウス過程 関連ベクトルマシン		

教師情報に基づく

形式的問題の分類

教師情報の提示方法

教師情報の訓練データへの与え方に基づく形式的問題設定の分類

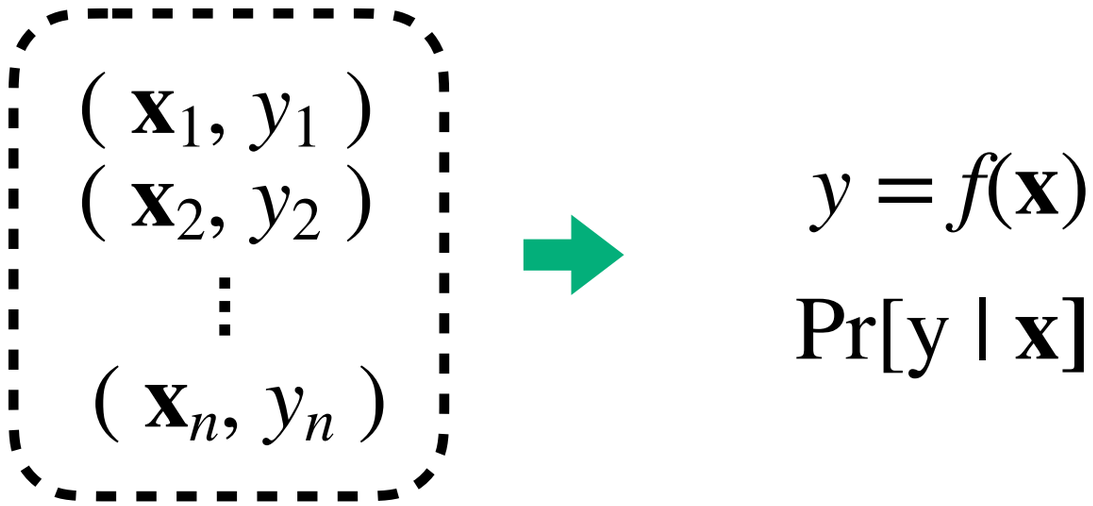
- **教師情報**：予測結果の情報を具体例の形で示したもの

- **教師あり学習 (supervised learning)**：各訓練データごとに教師情報を付加している
- **教師なし学習 (unsupervised learning)**：各訓練データに教師情報を付加しない
- **強化学習 (reinforcement learning)**：教師情報は報酬という形式で，個別ではなく，一連の行動の結果与えられる

教師あり学習

教師あり学習：個々の訓練データに，その予測結果である教師情報を付加している形式的問題設定

予測対象と予測結果の写像関係を獲得することが目標になる



- 最もよく研究された問題設定
- 教師情報があるため，学習した規則の性能評価が他の場合より容易
- 教師情報を人間が与える場合などは，訓練データの確保が困難に

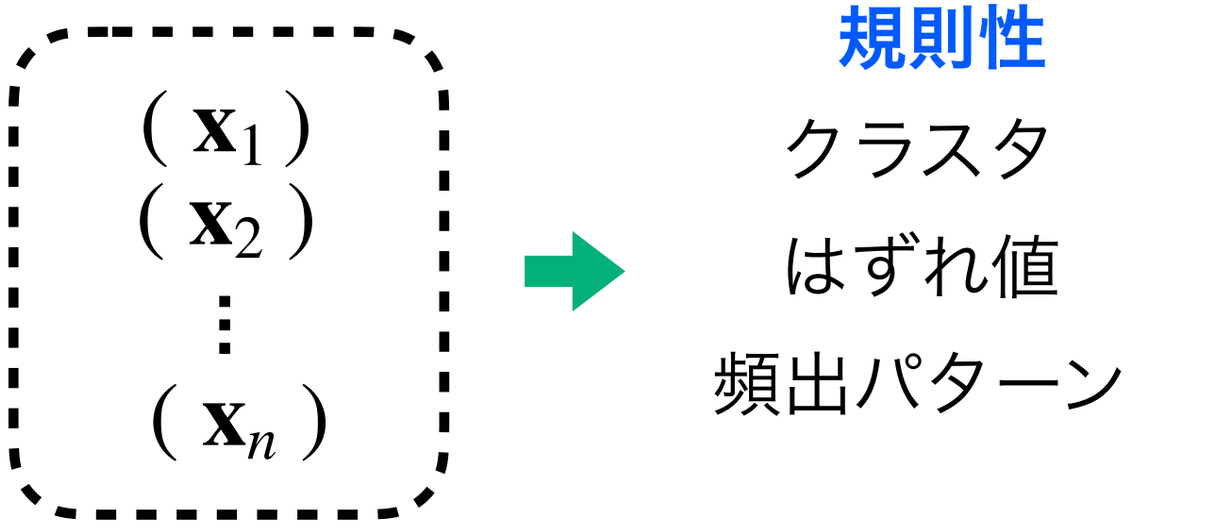
教師あり学習とその派生問題

- **クラス分類 (classification)**：予測対象が，事前に定めた有限離散集合であるクラスである場合
- **回帰 (regression)**：予測対象が実数である場合
- **ランキング学習 (learning to rank)** ／ **順序回帰 (ordinal regression)**：予測変数が，上中下といった順序関係のある離散値（順序変量）である場合．情報検索で適合する文書を順位付けする応用など
- **集約の学習 (learning to aggregate)**：入力が集合で出力が離散・実数などのスカラー量である関数関係の学習

- **半教師あり学習 (semi-supervised learning)**：訓練データの全てに教師情報があるのではなく，教師情報のない訓練データも存在する場合

教師なし学習

教師なし学習：個々の訓練データに教師情報がない場合
訓練データ中で一定の条件を満たす構造・事例・パターンを発見する



- 教師情報がないので，獲得した構造やパターンの性能評価が難しい
- 統計的機械学習としては，訓練データの分布推定問題に帰着できる
- 教師情報を準備する必要がないことは運用上の大きな利点

教師なし学習とその派生問題

- **クラスタリング (clustering)**：訓練データ集合を，その内部では似ていて，外部では似ていないようなクラスタとよぶ部分集合に分割
- **はずれ値検出 (anomaly detection)**：まれにしか生じない，主要な規則性に従わないデータを特定
- **頻出パターンマイニング (frequent pattern mining)**：非常によく生じる訓練データの規則性を発見する

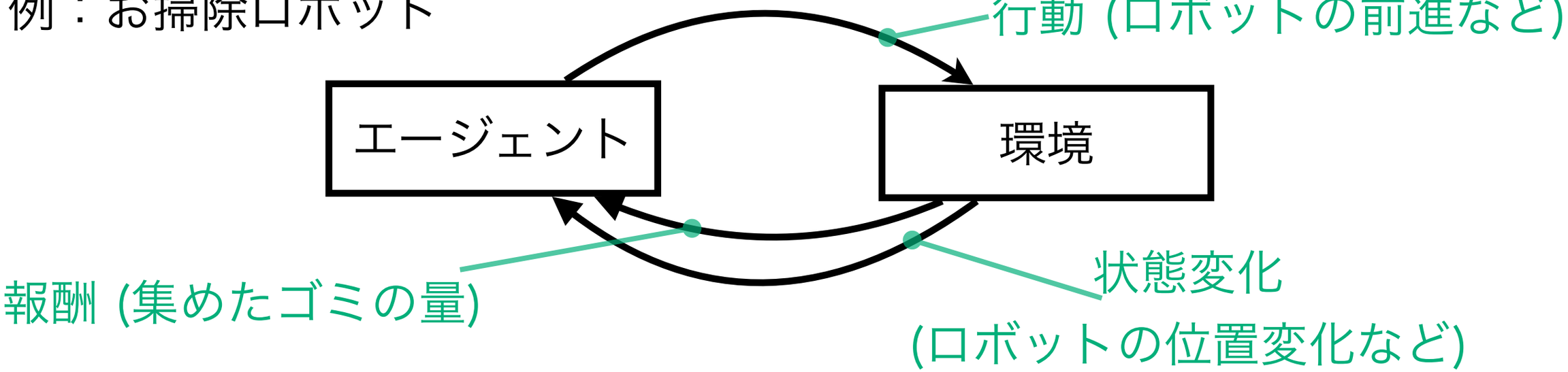
- **半教師ありクラスタリング (semi-supervised clustering)**：一対の訓練データが同じクラスタの要素になるべき (must link) や，違うクラスタに分かれるべき (cannot link) という教師情報を一部与える

強化学習

強化学習：環境に対する行動の結果，報酬という形式の教師情報と自身への状態変化が生じる

一連の行動の結果，累積報酬を最大化するような行動の決定方針（方策）を学習する

例：お掃除ロボット



- ロボットの制御や囲碁・チェスなどの対戦ゲームで利用される
- 環境の状態を把握する探索 (exploration) と，より多くの報酬を得ようとする利用 (exploitation) のバランスが重要になる

強化学習の派生問題

- **逆強化学習 (inverse reinforcement learning)**：最適な行動と環境が分かってるときに，その行動を導く報酬を設計する問題
- **見習い学習・徒弟学習 (apprenticeship learning)**：逆強化学習で報酬に加えて，最適な方策をも獲得する．最適な行動をまるで師匠のようにして，それをまねる方策を獲得する

その他の観点からの分類

その他の形式的問題設定

- **統計的学習理論 (statistical learning thory)**：汎化誤差の上界や収束性・一致性の理論
- **正則化 (regularization)**：汎化誤差を小さくするためにパラメータに制限を加える．特徴を疎にするといった背景知識を加える役割も
- **モデル選択 (model selection)**：汎化誤差を小さくするようなモデルを選ぶためのアルゴリズムや規準
- **特徴の操作**：必要な特徴を選ぶ特徴選択や，特徴を組み合わせて新たな特徴を作る特徴生成・特徴拡張
- **次元削減 (dimension reduction)**：高次元の特徴空間の中から，目的に必要な部分空間を選び出す
- **因果推論 (causal inference)**：いろいろな要因間の因果関係や，その関係の影響の度合いを調べる
- **データ同化 (data assimilation)**：数値シミュレーションの初期条件や境界条件の不備を，観測データからの帰納で補う

学習の枠組みに関する課題

- **転移学習 (transfer learning)**：目的のタスク用の訓練データだけでなく，類似した学習問題の訓練データを活用して，よりすぐれた予測モデルを得る
- **オンライン学習 (online learning)** ／ 逐次学習 (sequential learning)：訓練データが一度に与えられるのではなく，一つずつ逐次的に与えられる場合
- **アンサンブル学習 (ensemble learning)**：複数の予測モデルを組み合わせて，より高性能な予測モデルを作る
- **能動学習 (active learning)**：少ない訓練データでよい予測モデルを選択できるように，能動的に教師情報を得る学習の枠組み
- **メタ学習 (meta learning)**：適切な学習手法の選択手法を学習

学習手法の課題

- **分布推定**：ベイズ推定のための事後確率推定法
  - **マルコフ連鎖モンテカルロ**：サンプリングによる事後分布．系列データで状態をサンプルで表現する粒子フィルタなども
  - **変分ベイズ**：事後分布の近似計算手法
- **最適化**：目的関数を最適化する方法についての議論
  - **非線形最適化**：非線形関数，特にある種の凸性のあるものの最適化
  - **離散最適化**：パラメータが離散の最適化問題
  - **遺伝アルゴリズム**：遺伝での情報の伝播をモデル化した最適化
- **分散計算**：複数の計算機を用いた計算パラダイム．MapReduceやRDDなど

データの性質に関する課題

- **大規模データ**：扱えるデータの規模を大きくするための，圧縮データ構造，ハッシュ，サンプリングなどの技術
- **マルチラベル (multi-label) ／ マルチインスタンス (multi-instance)**：一つの対象に複数のラベルをつけたり，画像に複数の物体があって個々の物体のラベルをつけたりする
- **ラベルランキング (label ranking)**：複数のラベルを適切なものから順に整列
- **ロバスト推定 (robust estimation)**：はずれ値を含むデータに対して安定的な推定をする
- **不均衡データ (imbalanced data)**：クラス間で，訓練データ数に大きな偏りがある
- **不確実データ (uncertain data)**：特徴量が点ではなく，範囲や分布の形で与えられる

データの形式

- **構造データ**，**非構造データ**，**半構造データ**：表形式の構造データ，テキストなどの整備されいない非構造データ，これらの混合が半構造
- **関係データ (relational data)**：対象の間の関係の情報．利用者とアイテムの関係を扱う推薦システムなどで利用
- **時系列 (time series)**：時間経過に伴う変化を示したデータ
  - **データストリーム (data stream)**：一つのデータの処理時間に制限があり，データ数は無限でありうるような時系列データ
- **グラフ (graph)**：一般のグラフ構造．ソーシャルネットワークの友人関係や化合物を表現する．
- **空間データ (spatial data)**：地理情報を扱う．行政区画ごとの統計量や，空間中の離散的な点での観測値（地質調査や犯罪発生地点）などの形式

運用上の課題

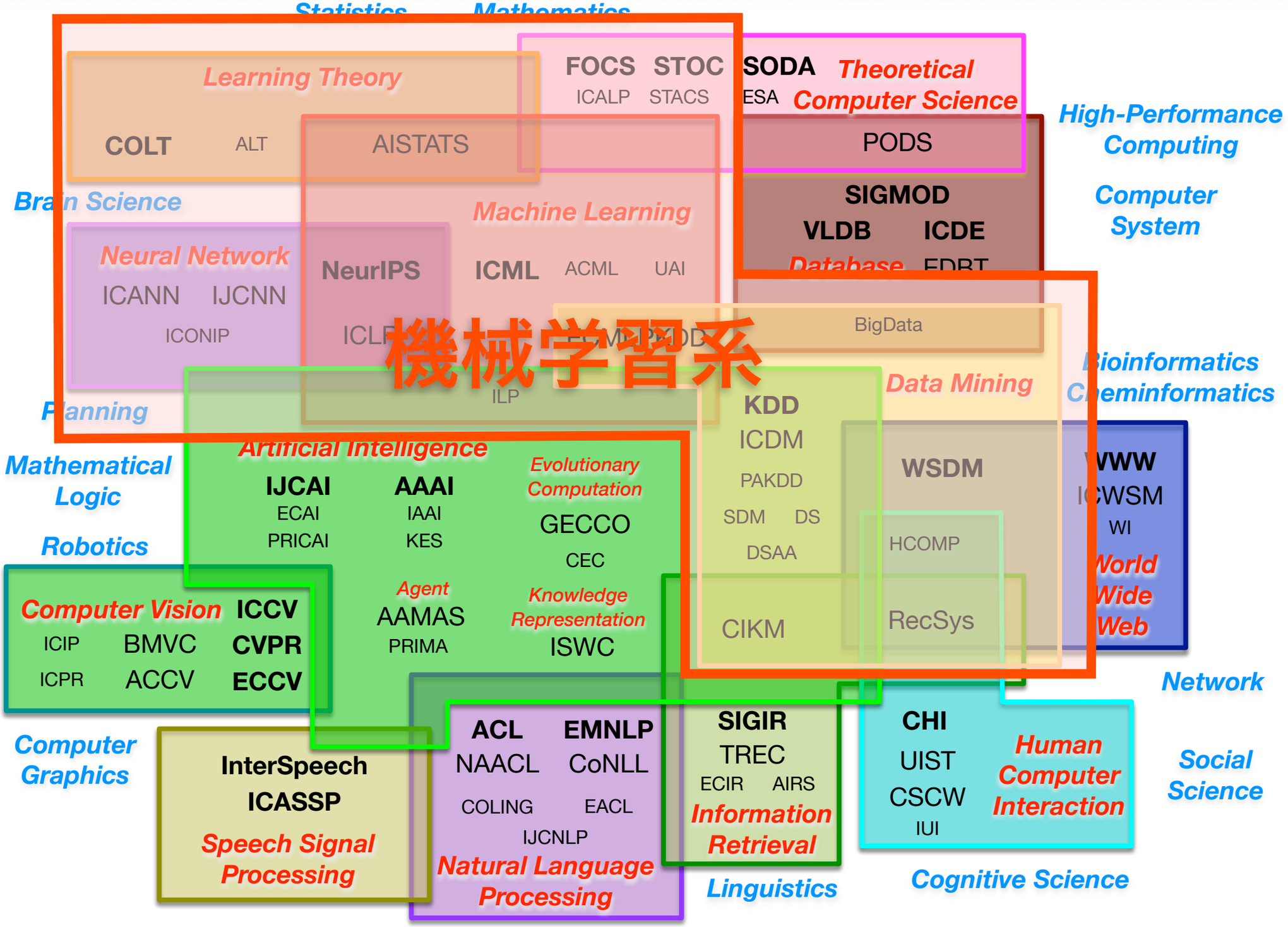
- **プライバシ (privacy)**：個人情報を秘匿と，データ分析とを両立する
- **公平性 (fairness)**：予測モデルが，社会的な公平性を保つようにする
- **説明 (explanation)・解釈可能性 (interpretablity)**：予測結果の根拠や，予測モデルの予測過程を人間に説明する
- **安全性 (security)**：訓練データに対する悪意のある介入や改竄などで，予測モデルが改変されないようにする
- **人間計算 (human computation)**：人間の主観的な判断が必要な場合などに，一部の処理を人間に任せる
- **ハードウェア**：クラスタやGPUを機械学習に利用

第IV部
機械学習・データマイニング関連
の国際会議

機械学習・データマイニング・人工知能

関連国際会議の概要

関連国際会議の俯瞰図



http://www.kamishima.net/archive/MLDMAImap.pdf

国際会議俯瞰図の説明

※ **上の方が抽象的，下の方が具体的な問題を対象にする傾向**

機械学習・データマイニングの分野

- 学習理論, 機械学習, データマイニング, ニューラルネット

その他の分野

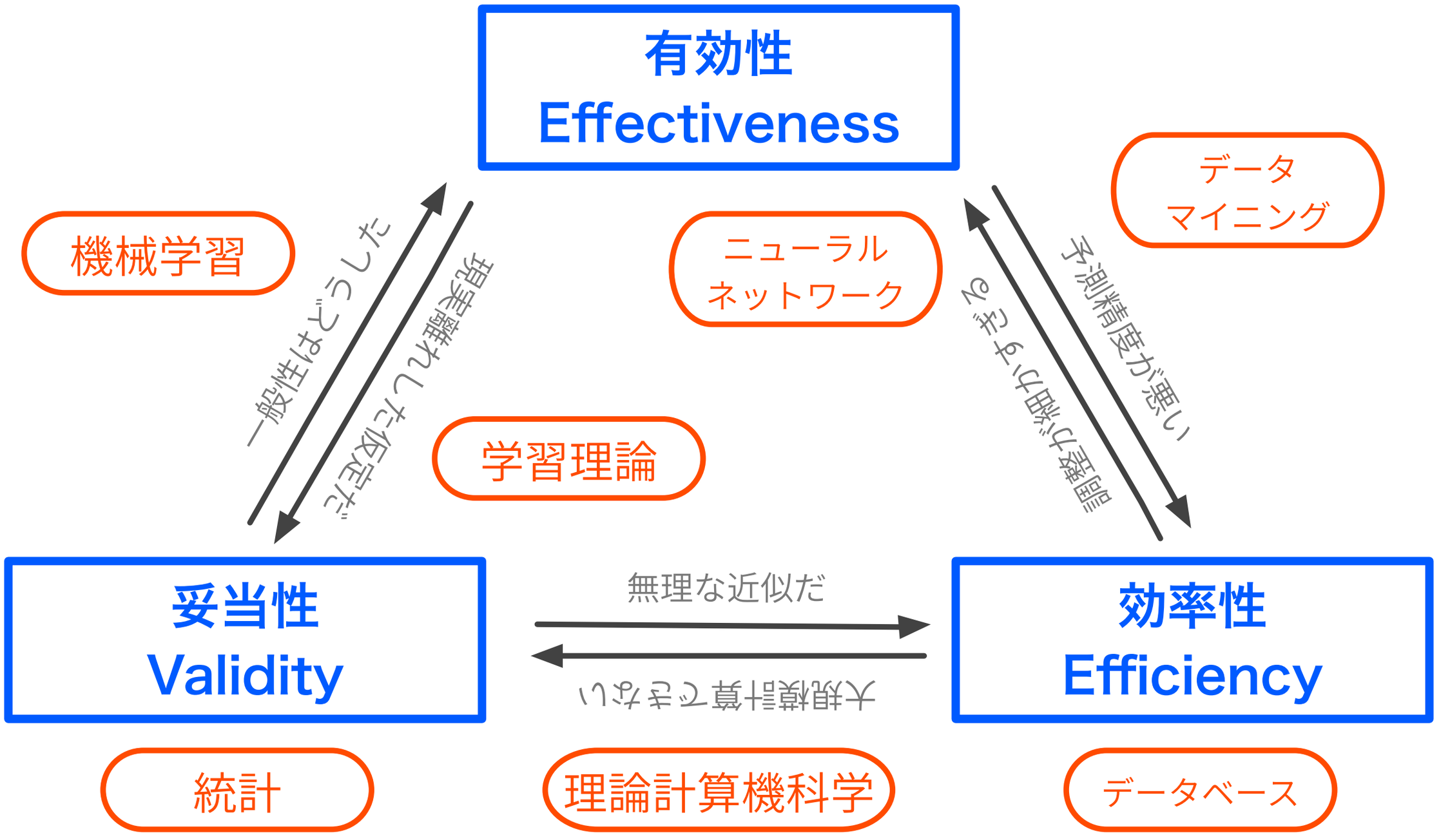
- **人工知能**：いろいろな分野を包括するような分野
- **CV**：画像の認識や合成
- **音声処理**：音声の認識・合成
- **自然言語処理**：自然言語の理解・合成
- **情報検索**：利用者に適切な情報を提供
- **HCI**：人間とコンピュータの連携
- **WWW**：Webの情報の分析とシステムの構築

人工知能

人工知能 (Artificial Intelligence)

- **目的**：知的な機械，特に，知的なコンピュータプログラムを作る科学と技術 [What is Artificial Intelligence, J. McCarthy: Basic Questions]
- **広義の分野定義**：ML, DM, AI国際会議一覧に示した全ての会議
- **狭義の分野定義**
  - AAAI / IJCAI などでは画像認識・音声認識の人はあまりいない
  - 雑な言い方をすれば，コミュニティが確立していない知的情報処理分野全般

各コミュニティの関心



有効性，妥当性，効率性の観点は一般にトレードオフの関係にある

学習理論

学習理論 (Learning Theory)

- **目的**：データから学習できるか？ できるとすればその条件は？ といったことを数理的に記述して厳密に議論する
- **他分野との関係**：機械学習アルゴリズムを実行すれば，確かに予測できたりしていることは，この分野の理論に基づいて保証される
- **コミュニティの指向**：公理や他の定理に基づいて，定理の形で証明できる結果を重視

機械学習

機械学習 (Machine Learning)

- **目的**：学習理論の保証に基づいて，データマイニングなどで使われる要素技術を提供する
  - 要素技術は，新しい情報を取り扱えるようにしたり，そもそも計算出来たり，それをより高速にしたりするもの
- **他分野との関係**：データマイニング分野より，より広範囲に適用できる，抽象的なレベルでの技術が対象
- **コミュニティの指向**：解いている問題や，提案している要素技術が自明でない新規の問題であることは重要

データマイニング

データマイニング (Data Mining)

- **目的**：機械学習分野で作られた要素技術を基本に，必要であれば追加の要素技術を開発し，それらを組み合わせて実世界の問題に対処
- **他分野との関係**：機械学習分野のように広範囲に適用できなくても，ある事例に対して有効な要素技術であれば研究対象となる
- **コミュニティの指向**：実世界の事例について，提案する手法が必然であるか，非常に有効であるということを重視する

機械学習・データマイニング・人工知能

関連国際会議の動向

国際会議の動向

国際会議の参加者数：急速に拡大

	2014	2015	2016	2017	2018	2019
KDD (DM)	2100	1100	2800	1700	3400	3200
ICML (ML)	1200	1600	3200	2400	5000	6000?
NeurIPS (ML)	2400	3800	5700	8000	8000	13000

企業スポンサーの動向

- 00年代前半はGoogle，IBM，Yahoo!，Microsoftなどの研究部門
- 00年代後半は米ネット企業 Amazon，Facebook，LinkedIn などに，中国の Tencent，Alibaba，Huawei やロシアのYandex など
- 10年代からは非ネット系の製造業や金融などに拡大
- NIPS2013は Facebook のザッカーバーグさんみずから乗り込んできてラボの設立を宣言し求人活動など加熱

日本の参加状況

データマイニング系

- 大学からの参加は減少傾向
- 日本とコアコミュニティとの繋がりは弱い
  - KDD2015 で標準化委員会創設のアナウンスがあったが何するのか日本に伝わってこない
- 2011ごろから企業の研究者だけでなくエンジニアの参加者が急速に増えたが，一方でコンサルなど企画系の職種の人は見かけない

機械学習系

- ICML/NeurIPSなど理論系の方はこれよりは若干状況はいい
- スポンサーに日本企業はほとんどなく，海外拠点の研究所がときどきブースを出しているのを目にする

会議の採録状況

投稿数・採択率の傾向

- 日本の投稿数は少なく，採択率はやや高め
- 投稿数が多いのは米中，高採択率はフィンランド・イスラエル

トップ会議採択率

- トップ会議の採択率は20%を切るような感じ
- これよりは低くならないように運営側も配慮している
- よく通ってる人でも何度か落ちてやっと通るようなもの

NIPS Experiment [Langford]

- 査読にはだいぶ運もある
- NIPS2014で査読精度を調査：同じ論文を別グループで査読
- 採録論文の57%で2グループ間に判定の相違がある
- 採録論文をもう一度査読すると40〜75%の確率で不採録に

機械学習・データマイニング
ニューラルネット・人工知能
の主要会議

ML/DM/NN/AI 関連会議

- **機械学習**：NeurIPS, ICML, ECMLPKDD, UAI, ACML
- **学習理論**：COLT, AISTATS, ALT
- **データマイニング**：KDD, ECMLPKDD, ICDM, WSDM, PAKDD, SDM, DS, ICMLA, BigData, DSAA
- **ニューラルネットワーク**：ICLR, IJCNN, ICANN, ICONIP
- **その他**：ILP, HCOMP, RecSys
- **人工知能全般**：IJCAI, AAAI
- **国内会議**：情報論的学習理論ワークショップ (IBIS)，人工知能学会全国大会

ICML

ICML (International Conference on Machine Learning)

- ホームページ，論文集，DBLP，1980年〜，6〜7月開催
- NeurIPS と双璧をなす機械学習のトップの国際会議
- 初期のころはワークショップと国際会議の形式で交互に開催されていたが，10回以降はconferenceとなった
- 80年代まではルールベースが中心だったが，90年代に統計的機械学習に中心が移った
- 90年代までは実験も必要だった ➡ 00年代に純粋に理論の論文が中心に移行 ➡ 10年代に深層学習系による実験の復活
- 00年代に International Machine Learning Society が主催団体として設立された.

NeurIPS

NeurIPS (Neural Information Processing Systems)

- ホームページ，論文集，DBLP，1987年開始，12月開催
- ICMLと双璧の機械学習のトップの会議
- ML/DM 分野で参加者数は最大
- 理論的な背景が明確なアルゴリズムなどが中心
- 90〜00年代は neural と名前についているにも関わらずニューラルネットはあまり扱われなかったが，10年代は深層学習の中心になり復活した
- 2008年の第21回までは，会議録の発行年が開催の翌年であった
- 2018年以降，略称をNIPS→NeurIPSに変更
- オーラルは選ばれた数10件ほどと，他は多数のポスター発表がある．ポスターの前で何時間にわたって楽しそうに議論している

KDD

KDD (International Conference on Knowledge Discovery and Data Mining)

- ホームページ，論文集，DBLP，1995年〜，8月開催
- データマイニング分野のトップ会議
- 最初の4回は AAAI の主催だったが，5回以降は ACM の主催に
- 理論的な背景もある程度明確にした上で，実験も必須，精度やスケーラビリティも重視される
- 理論よりの研究トラックと応用の応用データ科学トラック（2015年まではインダストリアル＆政府トラック）の二つのトラックがある
- インダストリアルの影響が強く，採択論文数の上位は企業の研究機関が占める
- 2008年までは北米のみの開催だったが，2009年のパリ以降は，3年に一度北米以外でも開催

ECMLPKDD

ECML (European Conference on Machine Learning) PKDD (European Conference on Principles and Practice of Knowledge Discovery in Databases)

- ホームページ, ECML@DBLP, PKDD@DBLP, 1987年～(ECML), 1997年～(PKDD), 9月開催
- ECMLは機械学習で ICML に次ぐレベルで, ヨーロッパで開催される会議. 会議録はSpringerのLecture Notes から出版される.
- PKDDはデータマイニング分野でKDDに次ぐレベルで, ICDM と同等
- 二つの会議は当所は別の会議であったが, 2000年以降共催されるようになり, 2008年以降は運営も統合されている
- ヨーロッパのコミュニティ内で, 知り合いを招待するような感じのアットホームさがある

ICDM

ICDM (IEEE International Conference on Data Mining)

- ホームページ，論文集，DBLP，2001年～，11～12月開催
- データマイニングで KDD に次ぐレベルの会議で，ECMLPKDDと同等
- IEEE Computer Society が主催
- KDD が2008年までは北米のみの開催であったのに対し，ICDM は米→亜太→米→欧 の4年周期の持ち回りで開催
- KDDはアメリカ中心だが，それよりはアジア系の影響は強い

COLT / AISTATS

COLT (Conference on Learning Theory)

- ホームページ，論文集，DBLP，1988年〜，6〜8月開催
- 学習理論のトップ会議
- 限界や収束性とかを論じる非常に理論よりの機械学習の会議で，定理や証明のない論文は扱わない
- Association for Computational Learningが主催

AISTATS (International Conference on Artificial Intelligence and Statistics)

- ホームページ，論文集，DBLP，1995年〜，4〜5月開催
- 90年代までは人工知能の応用としての側面が強かったが，00年代にNeurIPSのコミュニティに近くなり理論系の論文が集まるようになった

UAI / WSDM

UAI (Uncertainty in Artificial Intelligence)

- ホームページ，論文集，DBLP，1985年〜，6〜7月開催
- 記号的機械学習が主流の時代に，ベイジアンネットなどの確率的な手法の研究者が始めた.
- ベイズ系の論文が集まる
- Association for Uncertainty in AI が主催

WSDM (International Conference on Web Search and Data Mining)

- ホームページ，論文集，DBLP，2008年〜，2月開催
- Web関係のデータマイニングの会議でKDDとコミュニティは重複
- ACM の SIGIR，SIGKDD，SIGMOD，SIGWEB などが合同で開催
- Web系インダストリ中心

PAKDD / ACML

PAKDD (Pacific-Asia Conference on Knowledge Discovery and Data Mining)

- ホームページ, DBLP, 1997年～, 4～5月開催
- アジア・オセアニア地域のデータマイニング系の会議

ACML (Asian Conference on Machine Learning)

- ホームページ, 論文集, DBLP, 2009年～, 11月開催
- アジア・オセアニアで開催される機械学習の会議

SDM / DS

SDM (SIAM Conference on Data Mining)

- ホームページ，論文集，DBLP，2001年～，4～5月開催
- Society for Industrial and Applied Mathematics が主催
- 応用数理系の学会が主催なので，他のデータマイニング系会議より数理モデルが明確な研究が好まれる

DS (Discovery Science)

- DBLP，1998年～，10～11月開催
- 日欧が組んで始めたが，今では運営の中心はヨーロッパに
- アルゴリズムより知識発見に重点をおくというコンセプトを掲げる
- ALTと常に共催

ALT / ILP

ALT (Algorithmic Learning Theory)

- DBLP, 1990年～, 10～11月開催
- 日欧が組んで始めたが, 今では運営の中心はヨーロッパに
- COLT と同様に理論系の機械学習の会議だが, 数理論理系の研究なども扱う
- DSと常に共催

ILP (International Conference on Inductive Logic Programming)

- DBLP, 1995年～, 8～9月開催
- 数理論理系の機械学習だが, 確率を取り込んだ確率論理にも拡張

ICMLA / BigData / DSAA

ICMLA (International Conference on Machine Learning Applications)

- ホームページ，DBLP，2002年〜，12月開催
- IEEE / Association for Machine Learning and Applications 主催
- 応用系を掲げる機械学習

BigData (International Conference on Big Data)

- 論文集，DBLP，2013年〜，10〜12月開催
- IEEE Computer Society 主催
- データマイニング系だがデータベース寄り

DSAA (Data Science and Advanced Analytics)

- ホームページ，DBLP，2014年〜，10〜11月開催
- IEEE Computational Intelligence Society 主催
- データマイニング系

RecSys / HCOMP

RecSys (ACM Conference on Recommender Systems)

- ホームページ，論文集，DBLP，2007年〜，9〜10月開催
- ACM が主催
- 機械学習，ヒューマン・コンピュータ・インターフェース，情報検索の分野が推薦システムを中心にまとまった会議
- 研究系の発表と，企業系の招待トラックとがあり，インダストリとアカデミアが半分ずつを占めるような形式

HCOMP (AAAI Conference on Human Computation and Crowdsourcing)

- ホームページ，論文集，DBLP，2013年〜，11月開催
- 人間計算 (human computation) を対象にした会議
- 機械学習とヒューマン・コンピュータ・インターフェース分野が関わっている

IJCAI / AAAI

IJCAI (International Joint Conference on Artificial Intelligence)

- ホームページ，論文集，DBLP，1969年〜，7〜8月開催
- 機械学習を含めた人工知能分野全体を扱う会議で AAAI と同等
- 奇数年に開催されてきたが，2015年以降は毎年開催に
- 各国の人工知能関連の学会が持ち回りで開催する

AAAI (AAAI Conference on Artificial Intelligence)

- ホームページ，論文集，DBLP，1980年〜，1〜2月開催
- 機械学習を含めた人工知能分野全体を扱う会議で IJCAI と同等
- アメリカ人工知能学会 National Conference on AI だったが，
  2007年に国際学会への変更に伴い AAAI Conference on AI に変更
- 夏開催だったが，2015年の IJCAI 毎年開催への変更に伴って冬開催

ICLR / IJCNN

ICLR (International Conference on Representation Learning)

- ホームページ，2013年〜，4月開催
- 深層学習専門の会議
- 深層学習が有望視され始めた2013年に，特徴を獲得するという意味で表現学習の会議として設立された

IJCNN (International Joint Conference on Neural Networks)

- ホームページ，論文集，DBLP，2000年〜，6〜8月開催
- ニューラルネットワークの国際会議
- 現在は IEEE Computational Intelligence Society と International Neural Network Society の共催

ICANN / ICONIP

ICANN (International Conference on Artificial Neural Networks)

- ホームページ, DBLP, 1991年〜, 4月開催
- ヨーロッパのニューラルネットワークの国際会議
- European Neural Network Society の主催

ICONIP (International Conference on Neural Information Processing)

- DBLP, 1994年〜, 9月開催
- アジア・太平洋のニューラルネットワークの国際会議

情報論的学習理論ワークショップ (IBIS)

情報論的学習理論ワークショップ (Information-Based Inductive Sciences; IBIS)

- ホームページ，1998年〜，11月開催
- 情報理論を核としたデータ科学のワークショップとして始まったが，機械学習や統計など国内のデータ科学関連の最大の会議になった
- 1998〜2000年は情報理論とその応用学会を中心に開催，2001年からは電子情報通信学会 情報論的学習理論 時限専門委員会が，2010年からは電子情報通信学会 情報論的学習理論と機械学習 研究専門委員会が主催
- 2010年以降は，発表原稿は電子情報通信学会の技術報告として発行

人工知能学会全国大会

人工知能学会全国大会 (Annual Conference of the Japanese Society for Artificial Intelligence; JSAI)

- ホームページ，1987年〜，5〜6月開催
- 日本の人工知能関連の学会として1986年7月24日に設立された人工知能学会は，翌年に第1回の全国大会を開催した
- 機械学習や自然言語処理などのコンピュータ科学分野の人工知能に加え，認知科学や人文科学系の話題まで広く扱う

参考文献

[Barr 15] Barr, A.: Google Mistakenly Tags Black People as 'Gorillas,' Showing Limits of Algorithms, The Wall Street Journal (2015), ⟨http://on.wsj.com/1CaCNlb⟩

[Bishop 06] Bishop, C. M.: *Pattern Recognition and Machine Learning*, Information Science and Statistics, Springer (2006)

[Bishop 08] Bishop, C. M.: パターン認識と機械学習 — ベイズ理論による統計的予測, 上下, 丸善出版 (2007–2008), [監訳：元田 浩 他；訳：神嶌 敏弘 他]

[Bottou 15] Bottou, L.: Two High Stakes Challenges in Machine Learning, The 32nd Int'l Conf. on Machine Learning, Invited Talk (2015)

[Domingos 15] Domingos, P.: *The Master Algorithm*, Basic Books (2015)

[Domingos 21] Domingos, P.: マスターアルゴリズム — 世界を再構築する「究極の機械学習」, 講談社 (2021), [訳] 神嶌 敏弘

[Idé 17] Idé, T., Katasuki, T., Morimura, T., and Morris, R.: City-Wide Traffic Flow Estimation From a Limited Number of Low-Quality Cameras, *IEEE Trans. on Intelligent Transportation Systems*, Vol. 18, No. 4, pp. 950–959 (2017)

[人工 15] 人工知能学会（編）：深層学習 — Deep Learning, 近代科学社 (2015)

[Kohavi 15] Kohavi, R.: Online Controlled Experiments: Lessons from Running A/B/n Tests for 12 years, The 21st ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining, Keynote (2015)

[Langford] Langford, J.: The NIPS Experiment: ⟨http://cacm.acm.org/blogs/blog-cacm/181996-the-nips-experiment/fulltext⟩

[McNee 06] McNee, S. M., Riedl, J., and Konstan, J. A.: Accurate Is Not Always Good: How Accuracy Metrics Have Hurt Recommender Systems, in *Proc. of the SIGCHI Conf. on Human Factors in Computing Systems*, pp. 1097–1101 (2006)

[Michalski 93] Michalski, R. S.: Inferential Theory of Learning as a Conceptual Basis for Multistrategy Learning, *Machine Learning*, Vol. 11, pp. 111–151 (1993)

[Mitchell 97] Mitchell, T. M.: *Machine Learning*, The McGraw-Hill (1997)

[Perlich 11] Perlich, C., Kaufman, S., and Rosset, S.: Leakage in Data Mining: Formulation, Detection, and Avoidance, in *Proc. of the 17th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, pp. 556–563 (2011)

[Perlich 16] Perlich, C.: Automated Machine Learning in the Wild, The 10th ACM Conf.

on Recommender Systems, Keynote (2016)

[Samuel 59] Samuel, A. L.: Some Studies in Machine Learning Using the Game of Checkers, *IBM Journal of Research and Development*, Vol. 3, pp. 211–229 (1959)

[Varian 13] Varian, H. R.: Predicting the Present with Search Engine Data, The 19th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining, Invited Talk (2013)

[Watanabe 69] Watanabe, S.: *Knowing and Guessing — A Formal and Quantitative Study*, John Wiley & Sons, Inc. (1969)

[Wolpert 96] Wolpert, D. H.: The Lack of A Priori Distinctions Between Learning Algorithms, *Neural Computation*, Vol. 8, pp. 1341–1390 (1996)

[Wolpert 97] Wolpert, D. H. and Macready, W. G.: No Free Lunch Theorems for Optimization, *IEEE Trans. on Evolutionary Computation*, Vol. 1, pp. 67–82 (1997)