機械・深層学習を用いたデータ分析・活用事例

AI(人工知能)・機械学習・深層学習の関係

Source：https://ainow.ai/2019/11/26/180809/

AI(人工知能)

機械学習

教師あり学習

「教師あり学習」は、あらかじめ問題（データ）とその答えを与え学習させる方法。教師あり学習は、新しい問題（データ）に対して、あらかじめ与えた分類方法で分類するため、一般的に「過去のデータから将来起こりそうな事象を予測すること」に使われる。例：天気予報、株価

ニューラルネットワーク

脳機能に見られるいくつかの特性に類似した数理的モデル

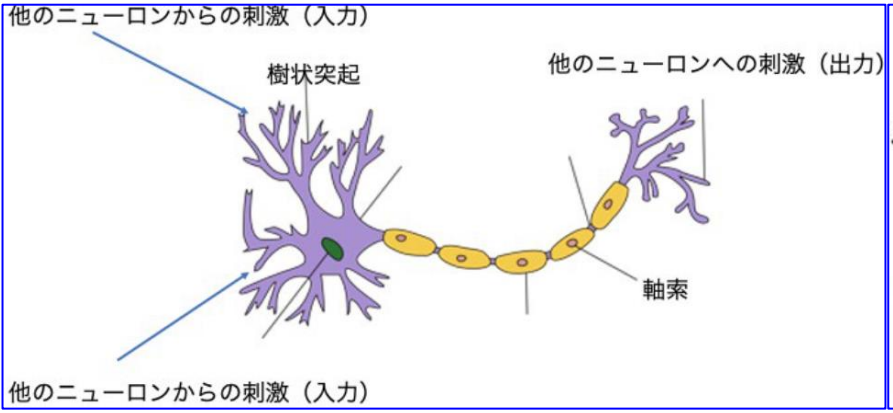
深層学習

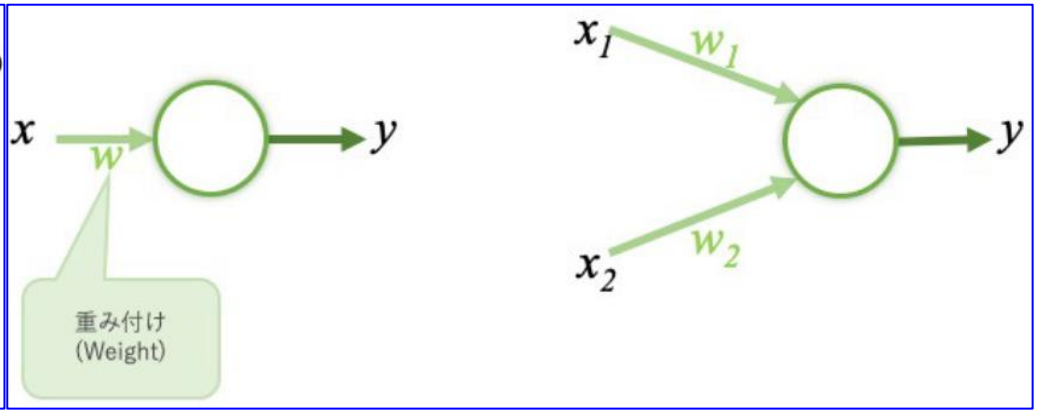
強化学習

「強化学習」は明確な正解がないときに、どの行動が「最適」か選択するように学習させる手法です。
事例：自動運転

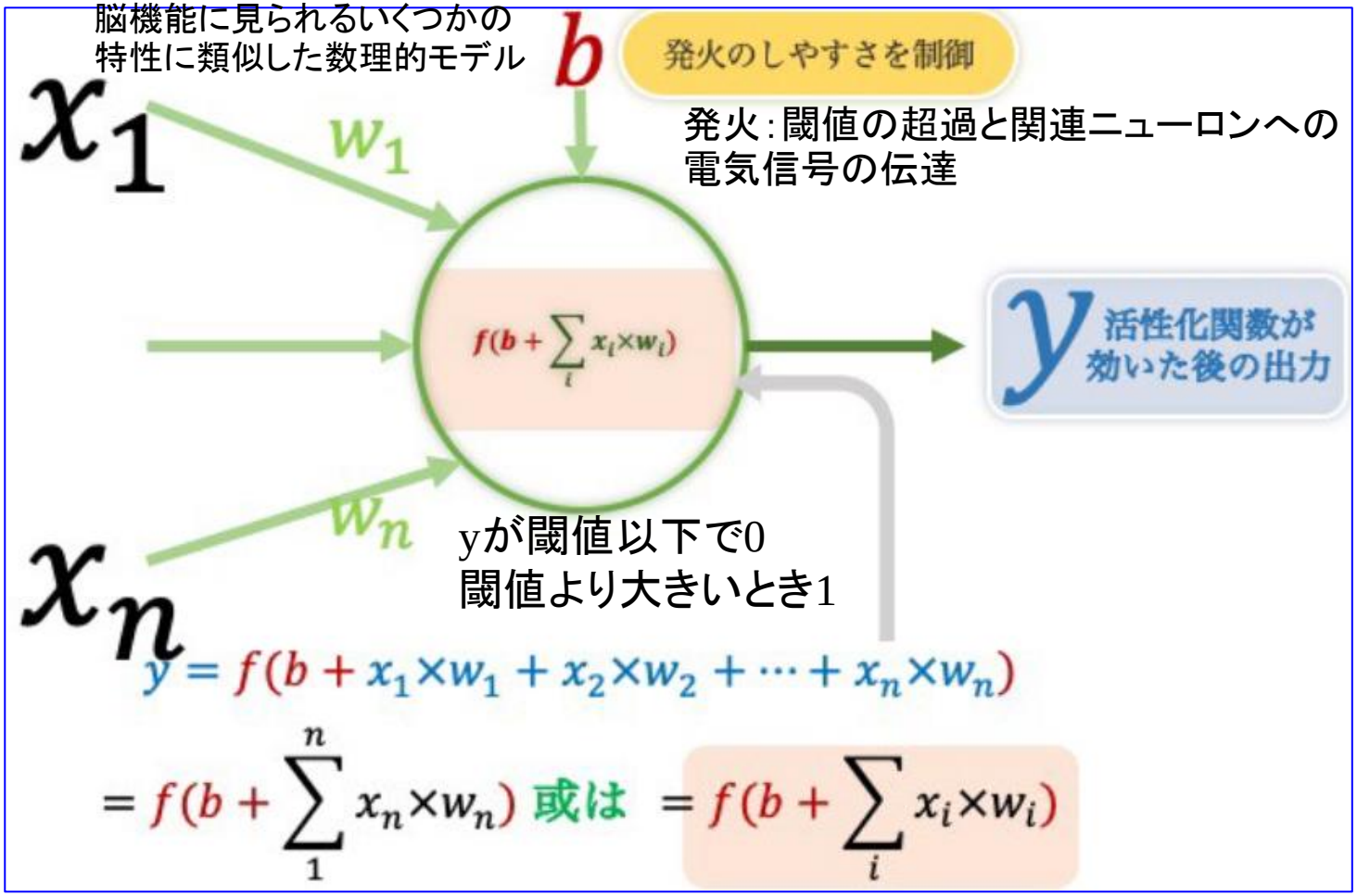
教師なし学習

「教師なし学習」では、問題（データ）のみを与え、その答えは与えずに学習させる方法です。教師なし学習は、データを与えられたときにそのデータに潜む傾向や構造を抽出するために用いられます。例として、アソシエーション分析というものは教師なし学習の一つです。そのアソシエーション分析で有名な話が、紙おむつとビールです。





脳機能に見られるいくつかの
特性に類似した数理的モデル



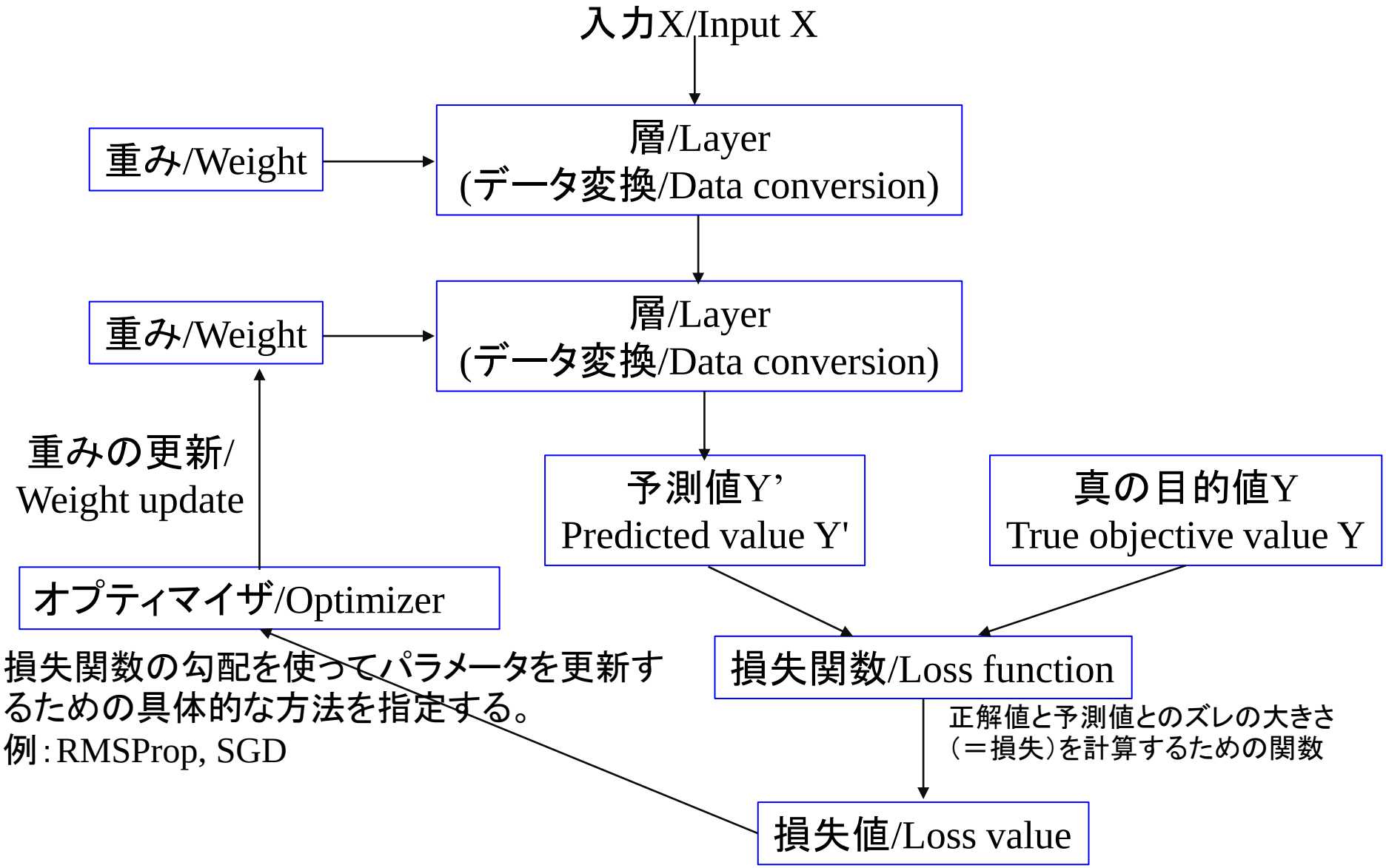
発火：閾値の超過と関連ニューロンへの
電気信号の伝達

yが閾値以下で0
閾値より大きいとき1

$$y = f(b + x_1 \times w_1 + x_2 \times w_2 + \cdots + x_n \times w_n)$$

$$= f\left(b + \sum_1^n x_n \times w_n\right) \text{ 或は } = f\left(b + \sum_i x_i \times w_i\right)$$

ニューラルネットワークの構造/Neural network structure



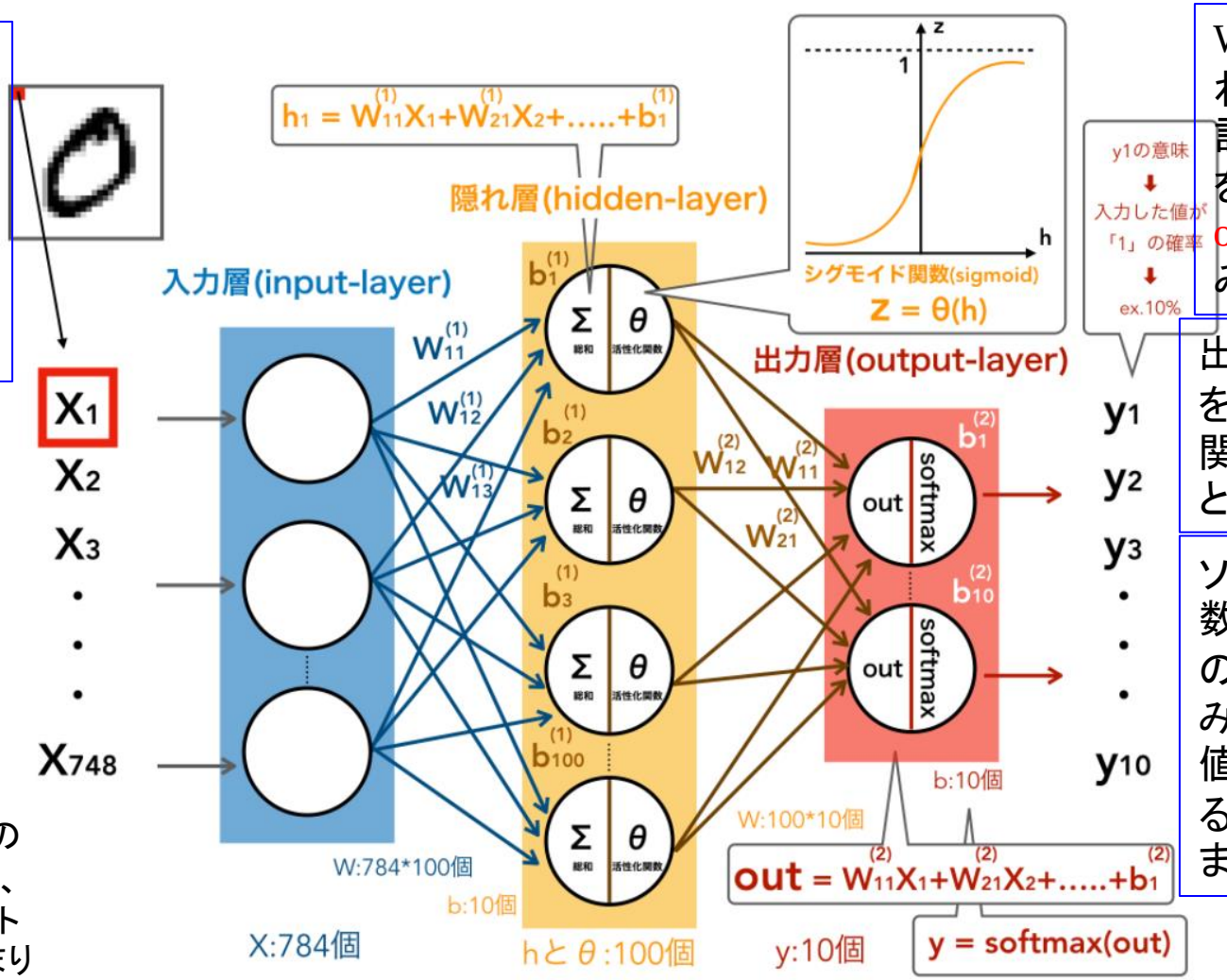
ニューラルネットワークの仕組み

入力値xは、このニューロン間の繋がりの強さを示す重みWと掛け合わせれ、伝播します。

Wとxが掛け合わせた784つの値の合計とバイアスbを加えて、hという隠れ層での値を生み出します。

このモデルでは、表現できるモデルが少ないため、活性化関数と呼ばれる関数にこのhを代入します。

28*28pixelのグレースケールの画像(pixel値は0-255)を28*28=748個の縦長のpixelの列に変換



Wとxが掛け合わせた値の合計にバイアスbを加えることでoutという値を生み出します。

出力値yは、outをソフトマックス関数に入れることで求めます。

ソフトマックス関数は、出力を0-1の値に落とし込み、出力された値の合計が1となるように値を返します。

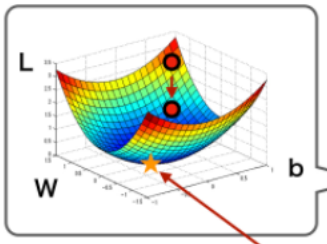
隠れ層は前の層のアウトプットであり、次の層のインプットになる変数の集まり

Source: https://miyabi-lab.space/blog/9

ニューラルネットワークの学習

予測値と正解ラベルとの誤差を評価するために損失関数を用います。

y^(予測値) ⇔ y(正解ラベル)



WとLのみで見ると..

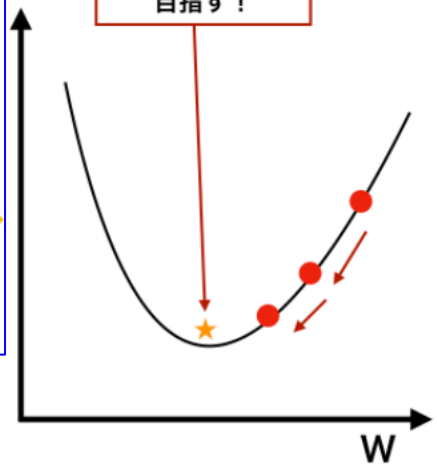
この2つの値の誤差を最小にすればOK！

損失関数(Loss-Function)

$L(W, b) = y^*\log\hat{y} + (1-y)\log(1-\hat{y})$

誤差を最小にするように重みWとバイアスbを更新していきます。

赤い点をWとbを更新することで動かしていき、最終的に星の描かれている最も低い位置、つまり誤差の最小となる地点を目指していきます。



$\frac{dL}{dW}$ = W方向の傾き

$\frac{dL}{dW} > 0$ (右上を向いているので)

※αは学習率(どのくらい割合で更新するか)

正の値(+)

$W(新しい) = W(古い) - \alpha\frac{dL}{dW}$

Wを新しい値に更新するためにW(新しい)=W(古い)-α*dL/dWという計算を行います。

W(新しい) < W(古い)

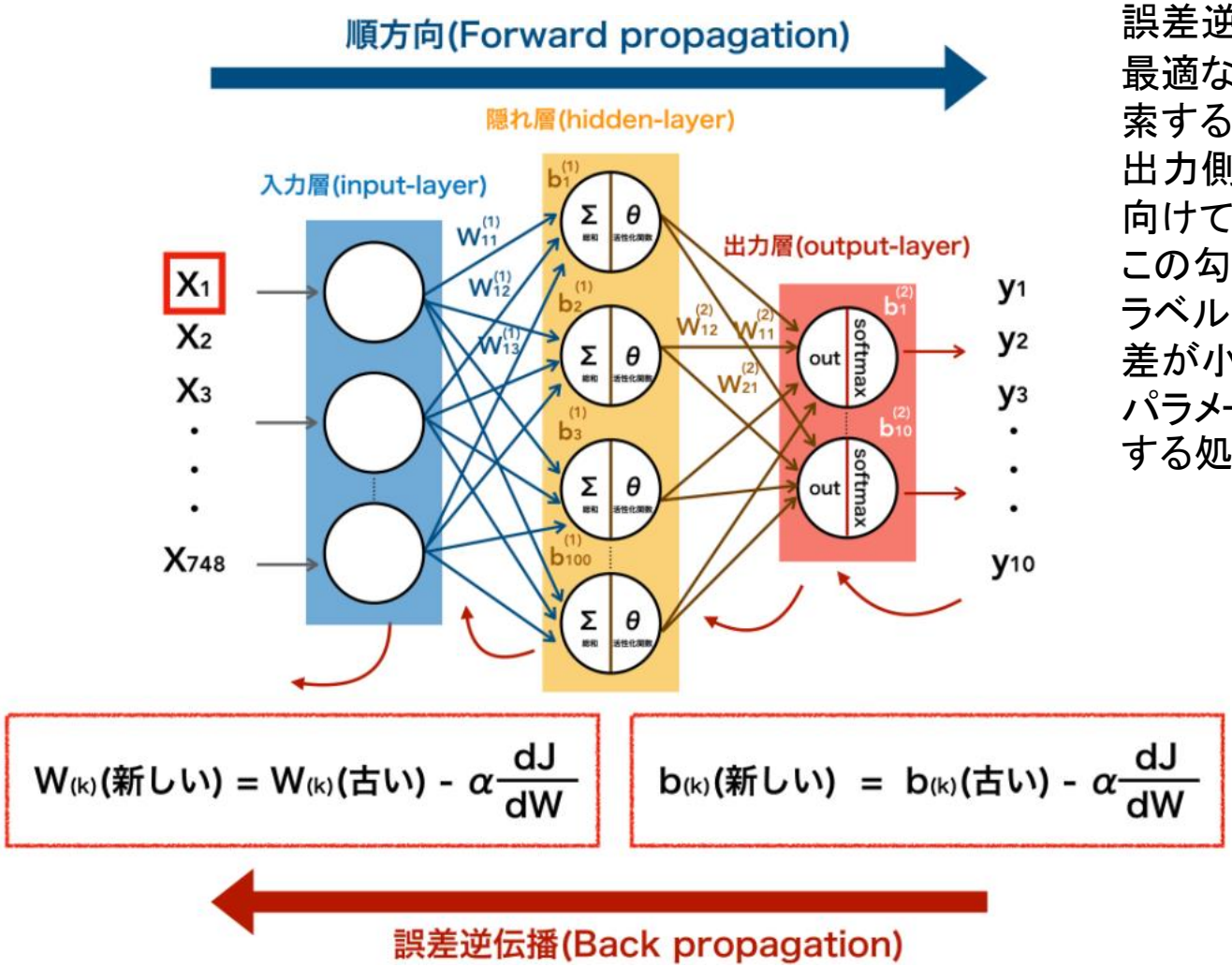
Wは小さくなるように更新させた!

bも同様に考えると..

$b(新しい) = b(古い) - \alpha\frac{dL}{dW}$

バイアスbも同様にして更新していくことができます。

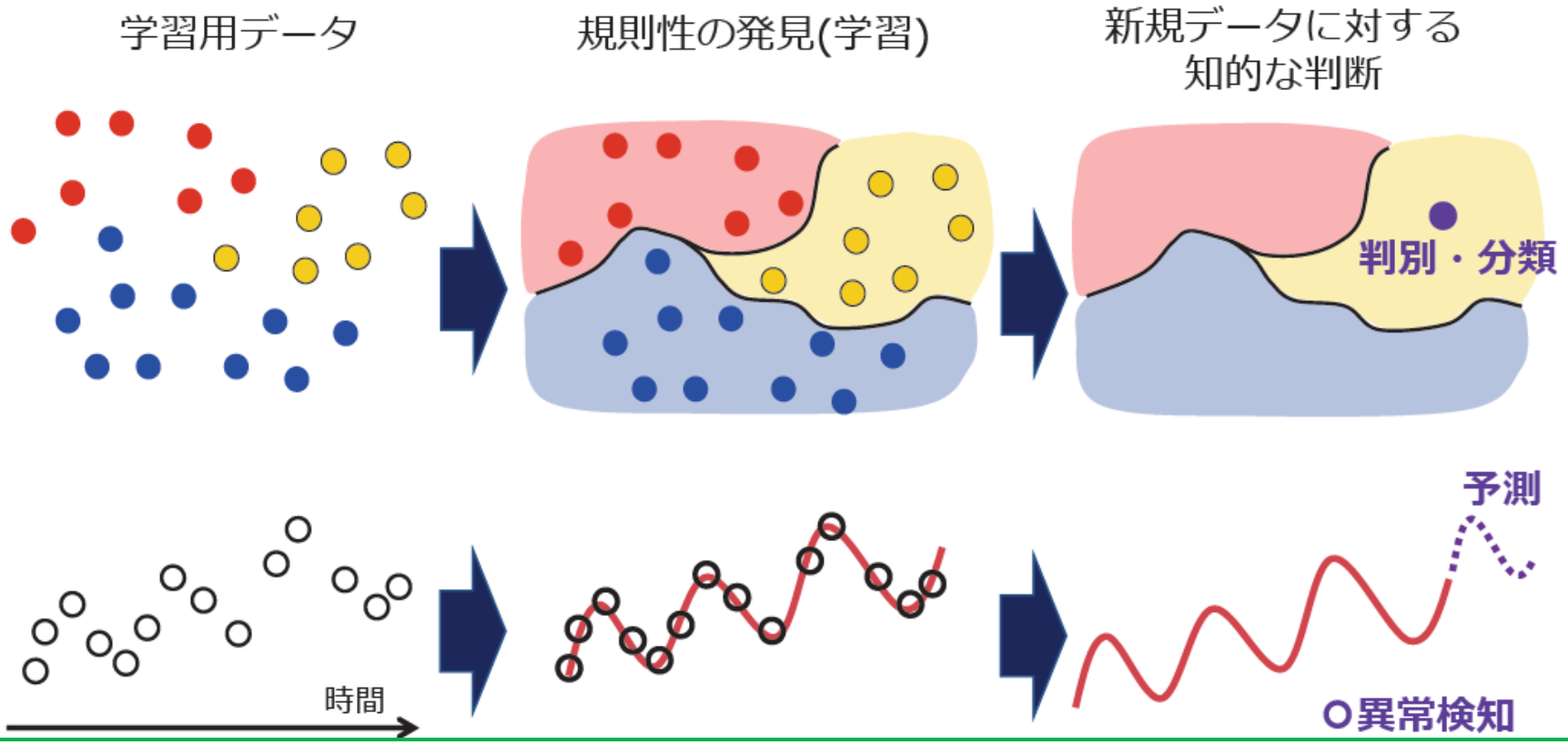
モデルを更新するための誤差逆伝播



誤差逆伝播法：
最適なパラメータを探索するための勾配を出力側から入力側に向けて逆順に計算し、この勾配をもとに教師ラベルと予測結果の差が小さくなるようにパラメータの値を更新する処理

予測をより精度を高めるために、後ろから順番にW,bを数式を使って更新します。

機械学習



機械学習とは「機械に大量のデータからパターンやルールを発見させ、それをさまざまな物事に利用することで判別や予測をする技術」のことである。

機械学習はデータの中のどの要素が結果に影響を及ぼしているのか（特徴量という）を人間が判断、調整することで予測や認識の精度をあげている。

深層学習はデータの中に存在しているパターンやルールの発見、特徴量の設定、学習なども機械が自動的に行うことが特徴である。人間が判断する必要がない。

7

機械学習で使われるアルゴリズム

▸分類（教師あり学習）▸回帰（教師あり学習）▸クラスタリング（教師なし学習）▸次元削減（教師なし学習）▸異常検知

問題を解決するための手順や計算方法

Source：https://ainow.ai/2019/11/26/180809/

教師あり学習

回帰

- 線形回帰
- サポートベクターマシン
- Elastic Net
- SVR

分類

- ランダムフォレスト
  - 決定木(閾値を設定して分類)：分類木、回帰木
- ロジスティック回帰
- サポートベクターマシン

教師なし学習

クラスタリング

- 群平均法
- K-means法
- Ward法

次元削減

- 主成分分析

異常検知

- K近傍法
- 主成分分析
- 1-class SVM

機械学習における分類

顧客の年齢、職業、婚姻状況、教育レベル、デフォルト(債務不履行)の有無、残高、住宅ローンの有無、個人ローンの有無、顧客へ最後に連絡した際の連絡手段、最後に連絡した日付、最後に連絡した月、顧客への最後の連絡で接触した時間、現在のマーケティングキャンペーンにおける顧客への連絡回数、以前のマーケティングキャンペーンにおける顧客への最終連絡日からの経過日数、以前のマーケティングキャンペーンにおける顧客への連絡回数、以前のマーケティングキャンペーンの結果などの個人情報から、定期預金の契約有無(yes or no)どちらに属するかを分類

age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
30	unemployed	married	primary	no	1787	no	no	cellular	19	oct	79	1	-1	0	unknown	no
33	services	married	secondary	no	4789	yes	yes	cellular	11	may	220	1	339	4	failure	no
35	management	single	tertiary	no	1350	yes	no	cellular	16	apr	185	1	330	1	failure	no
30	management	married	tertiary	no	1476	yes	yes	unknown	3	jun	199	4	-1	0	unknown	no
59	blue-collar	married	secondary	no	0	yes	no	unknown	5	may	226	1	-1	0	unknown	no
35	management	single	tertiary	no	747	no	no	cellular	23	feb	141	2	176	3	failure	no
36	self-employed	married	tertiary	no	307	yes	no	cellular	14	may	341	1	330	2	other	no

説明変数

目的変数

決定木を用いて分類可能

決定木は分類と回帰に使用されるノンパラメトリックな教師あり学習方法です。

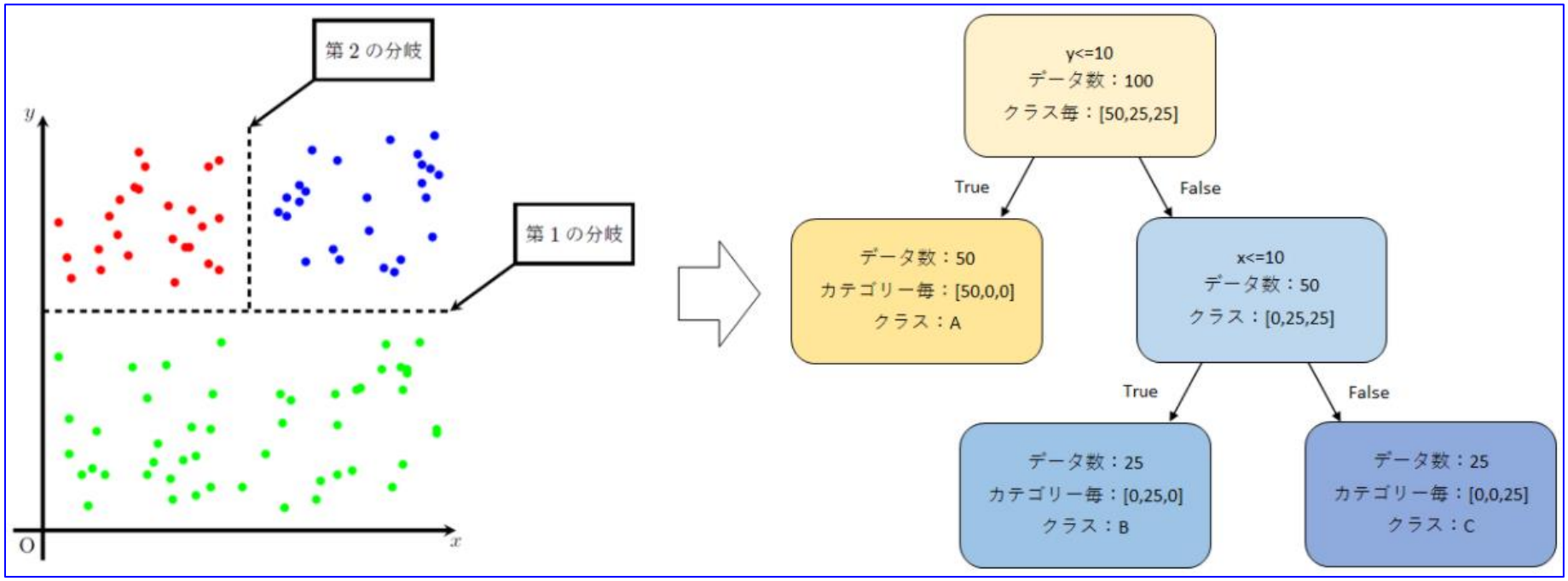
ノンパラメトリック検定は母集団の分布を仮定しない検定法

目標は、データの特徴量から推測される単純な決定ルールを学習することにより、ターゲット変数の値を予測するモデルを作成する。

「決定木分析」とは、ある目的に対して関連の強い項目から順に分岐させ、ツリー状に表す分析手法

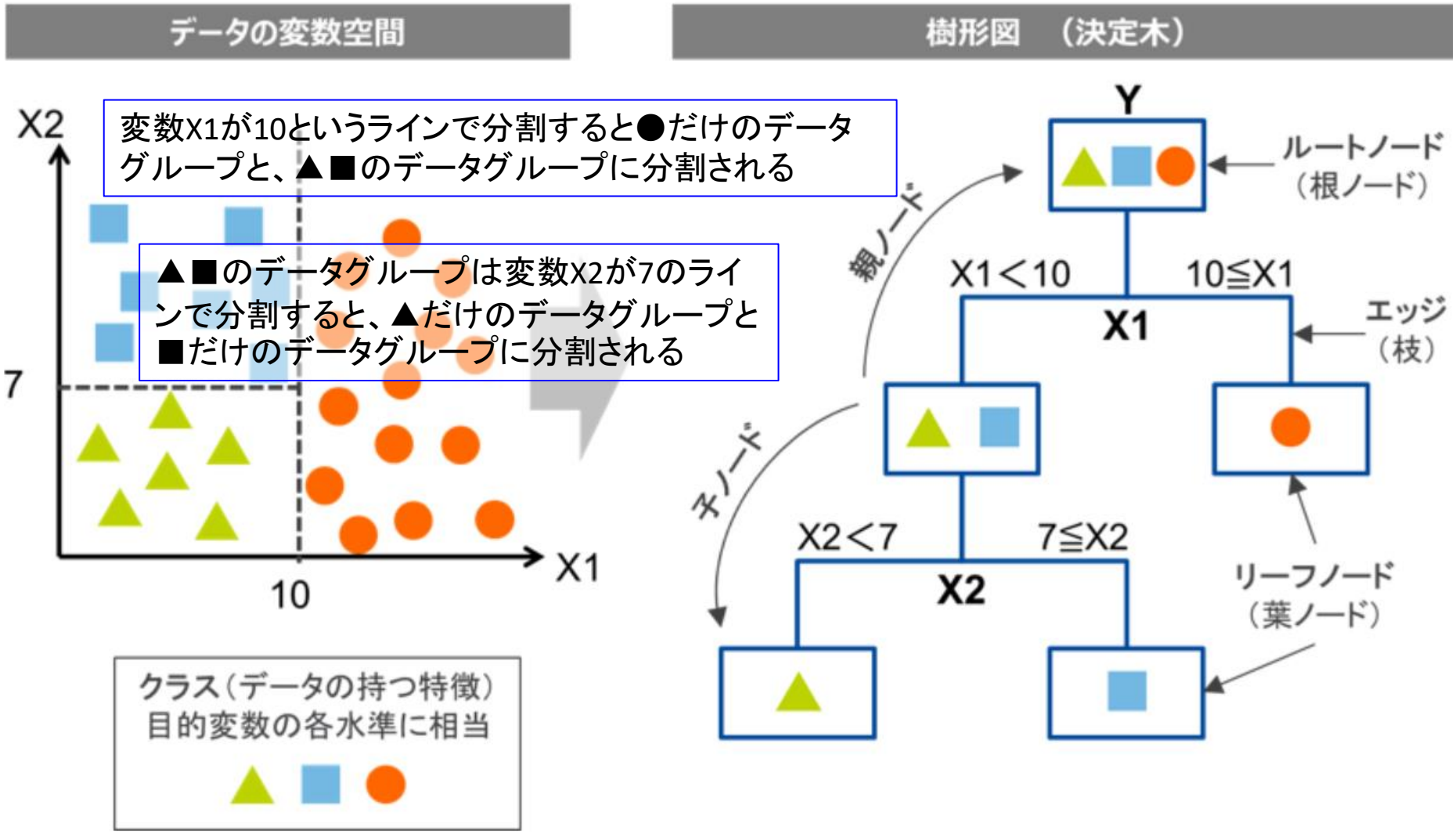
目的変数と説明変数の関係に着目し、目的変数が変化する説明変数の閾値を探し出して分岐を繰り返す。

最初は大きな分岐からスタートし、徐々に小さい分岐に移っていく。



Source: https://pictblog.com/ai-bunrui-9

決定木のイメージ図



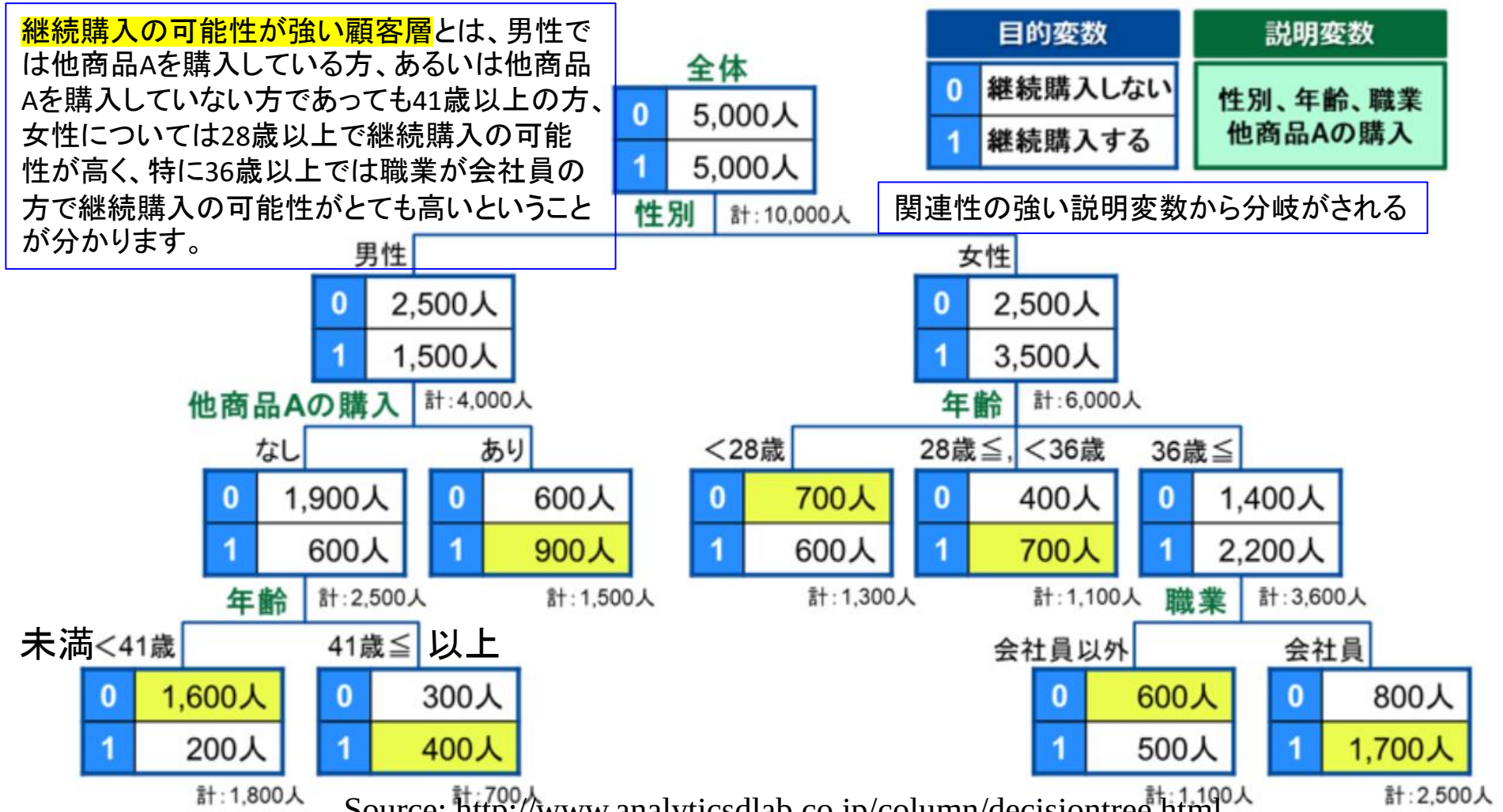
Source: http://www.analyticsdlab.co.jp/column/decisiontree.html

全10,000人の顧客の購買データ=商品を継続して購入しなかった人が5,000人＋継続して購入した人が5,000人。継続購入が目的変数：0：継続購入しない、1：継続購入するという2つのクラスを持つ質的変数。説明変数：性別、年齢、職業、また他商品Aを購入しているどうか、質的変数と量的変数の両方。

目的変数に対して効果的な切り方の閾値を計算

決定木の適用例（顧客購買データへの適用）

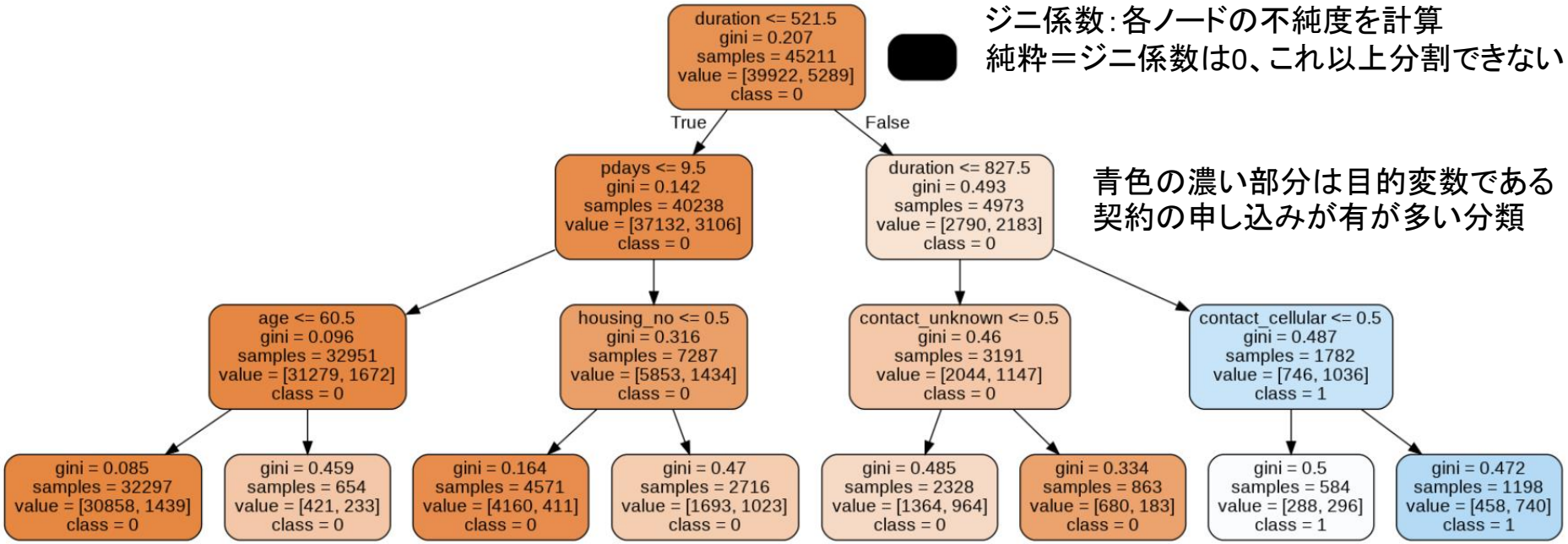
継続購入の可能性が強い顧客層とは、男性では他商品Aを購入している方、あるいは他商品Aを購入していない方であっても41歳以上の方、女性については28歳以上で継続購入の可能性が高く、特に36歳以上では職業が会社員の方で継続購入の可能性がとても高いということが分かります。



Source: http://www.analyticsdlab.co.jp/column/decisiontree.html

決定木の実装

目標：個人情報から、定期預金の契約有無(yes or no)どちらに属するかを分類



ジニ係数：各ノードの不純度を計算
純粋＝ジニ係数は0、これ以上分割できない

青色の濃い部分は目的変数である
契約の申し込みが有が多い分類

青色の濃い部分は、定額預金の契約が有（yes) が多い分類

読み方：duration <= 827.5　かつ　contact_cellular <= 0.5 である場合に、契約が有る可能性が高い
duration最終接触時間（秒）、durationが短い（例えば521秒以下）では契約が少ない傾向

決定木の実装

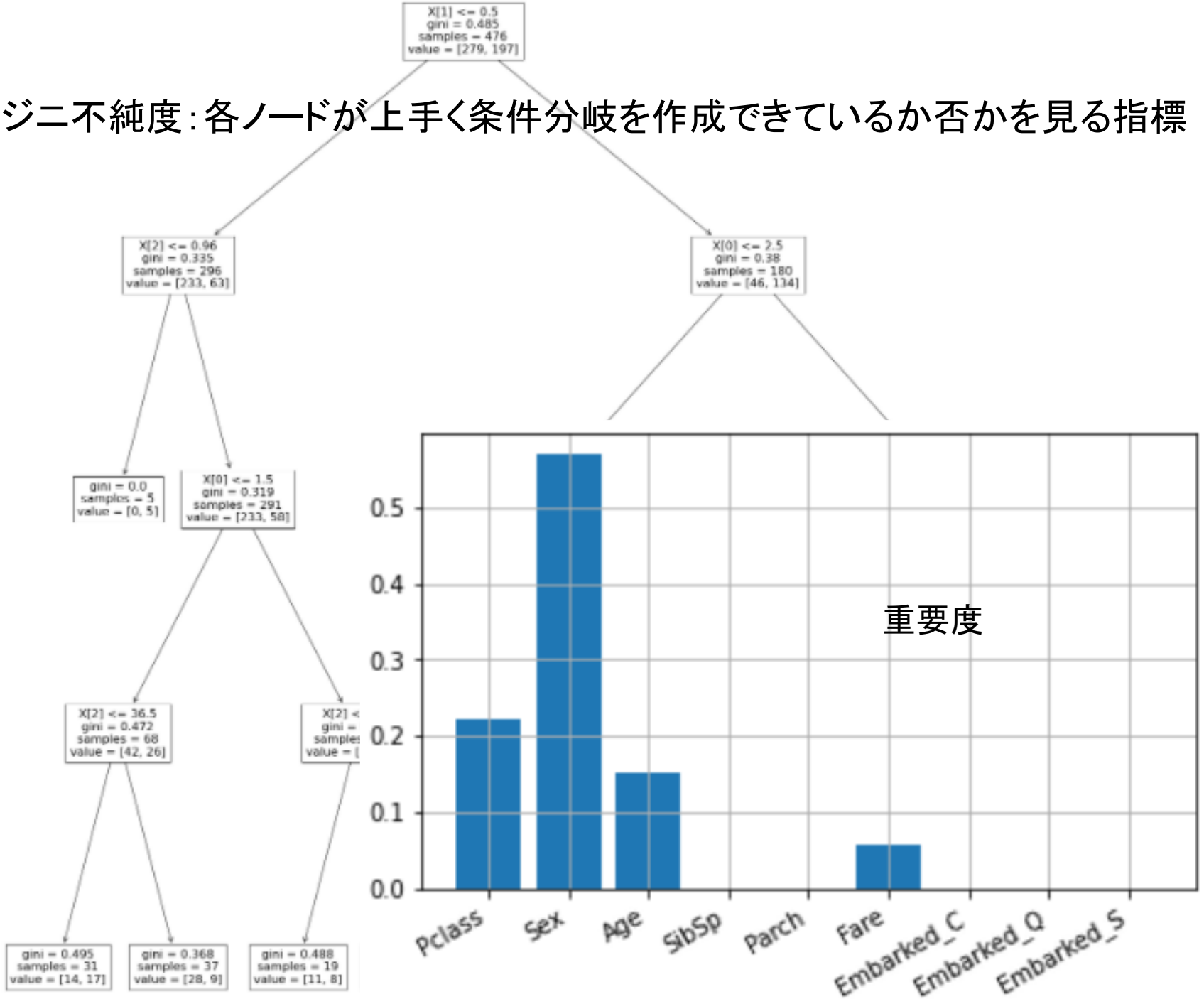
タイタニック号の乗客の情報から傾向を見つけて彼らの生存率を推定

	A	B	C	D	E	F	G	H	I	J	K	L
1	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
2	1	0	3	Braund, M	male	22	1	0	A/5 21171	7.25		S
3	2	1	1	Cumings,	female	38	1	0	PC 17599	71.2833	C85	C
4	3	1	3	Heikkinen	female	26	0	0	STON/O2.	7.925		S
5	4	1	1	Futrelle, M	female	35	1	0	113803	53.1	C123	S
6	5	0	3	Allen, Mr.	male	35	0	0	373450	8.05		S
7	6	0	3	Moran, Mr	male		0	0	330877	8.4583		Q
8							0	0	17463	51.8625	E46	S

カラム	説明
PassengerID	乗客番号
Survived	1: 生存, 0: 死亡
Pclass	乗客階級 （1,2,3の順で上級）
Name	名前
Sex	性別
Age	年齢
SibSp	兄弟/配偶者人数
Parch	両親/子供人数
Ticket	チケット番号
Fare	乗船料金
Cabin	部屋番号
Embarked	乗船した港　Cherbourg、Queenstown、Southampton

データセットは、train.csv（train用）/ test.csv（submit用）の2つ存在。
test.csvは、Survivedのみ情報が欠落しているため、これを推定

ジニ不純度：各ノードが上手く条件分岐を作成できているか否かを見る指標



各業種におけるビッグデータの活用事例

1. 電力需要予測
電力量＝ビルごとの従業員数×天候×カレンダー（電力の変動要因は季節・時間変動が支配的）

2. 商品需要予測
店舗ごとに売れ行きを予測し、自動発注するソリューションが有用

3. 適正価格予測
異種混合学習により、中古品価格＝中古品の型番×性能×特徴

多種多様なデータの中から精度の高い規則性を自動で発見し、その規則に基づいて、状況に応じた最適な予測を行う。

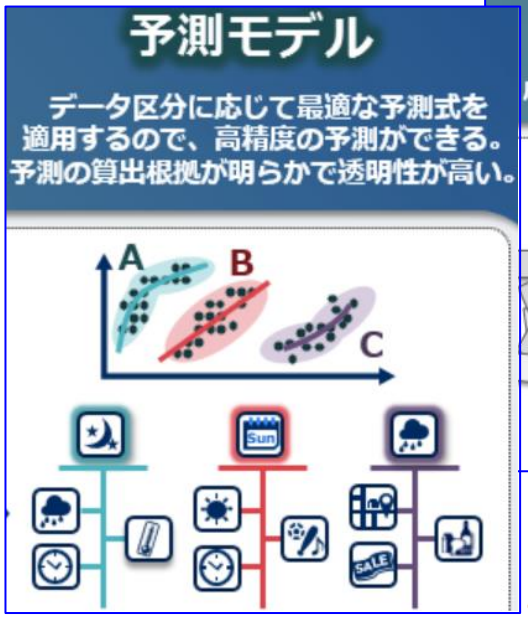
Source: 中台ら，ビッグデータ時代におけるビジネス向け機械学習，日本リアルオプション学会機関誌 リアルオプションと戦略 7(1)2015
Source: https://jpn.nec.com/ai/analyze/pattern.html

ハウス食品、NECとAIを利用した需給最適化に着手

11/4(木) 15:57 配信 0

ZDNet Japan

異種混合学習



ハウス食品およびグループのハウスウェルネスフーズとサンハウス食品は11月4日、NECと共同で需給・生産管理におけるサプライチェーンマネジメントシステムを4月に統合、共通基盤で全体最適の運用を開始したと発表した。これにより市場変動への対応高速化と食品ロスの削減を図る。

ハウス食品グループの3社は、新たなサプライチェーンマネジメントシステムで、NECの生産管理システム「FlexProcess」と「需給最適化プラットフォーム」を採用した。FlexProcessによって需給予測から生産管理までの業務の統合し、倉庫や店舗など社内外の組織間の情報を緊密に結合、連動させることで市場ニーズの変化に迅速に対応する。需給最適化プラットフォームは、NECが手がける人工知能（AI）技術の1つ「異種混合学習技術」を利用。ここでは全国のエリアや倉庫ごとに傾向の違う商品の出荷数や販売数などを予測する数万もの予測モデルをAIで作成して、需給計画や発注業務の効率化を図る。

今回の施策により3社では、欠品件数で50%、製品・資材の廃棄ロスで10%、管理業務工数で60%の削減を目指すとしている。

データ区分に応じて
最適な予測
⇒高精度の予測

観測データ中に混在する複数の規則性を、自動的に分割・抽出、高い精度と解釈性を両立

Source: https://jpn.nec.com/ai/analyze/pattern.html
Source: https://news.yahoo.co.jp/articles/a72ed2860f7ff2d8fb752961479816f2537bb97f

各業種におけるビッグデータの活用事例

4. 品質予測

工場などでの生産管理においては、生産品の収率・歩留りが重要な経営指標になる。この指標を高くするには、合格ラインとなる品質以下となる生産物がどのような生産条件、環境、材料の状態であるかを見極め、生産工程の改善を行う必要がある。

品質＝生産条件×環境×材料の状態

5. 劣化予測

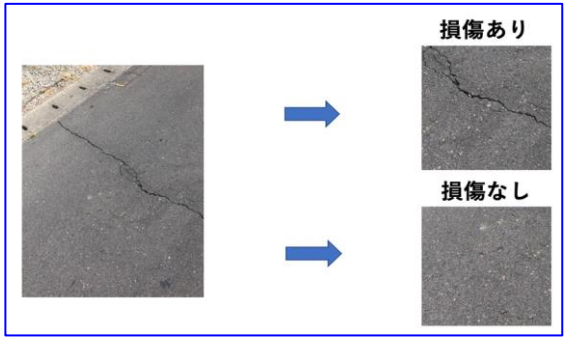
橋や道路、線路などの社会インフラの劣化によって発生する事故のニュースが後を絶たないが、安全のために劣化点検の頻度を増やし、保守要員を多くするのは、インフラ維持コストの増大となり、企業や個人、社会全体にとって望ましいものではない。

インフラの劣化程度を予測し、点検した方が良いと思われる時期でのみ、人が点検することでコストを抑える。

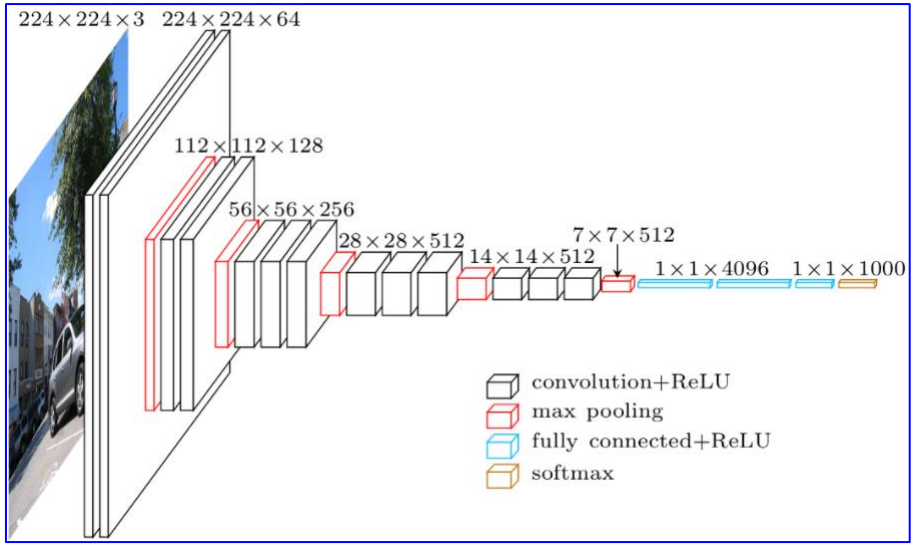
Source: 中台ら, ビッグデータ時代におけるビジネス向け機械学習, 日本リアルオプション学会機関誌 リアルオプションと戦略 7(1)2015

研究事例紹介：深層学習を用いた道路路面損傷有無の分類

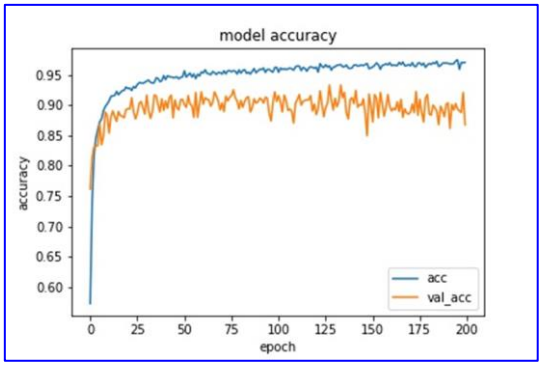
①データ収集・前処理
・道路路面の損傷画像をスマートフォンで撮影
・損傷有画像3,447枚,損傷無画像1,149枚
・Train, Validation, Test用に分割



②学習
・VGG16を用いて実験
・13層の畳み込み層と3層の全結合層の計16層からなる畳み込みニューラルネットワークであり1,000クラスを分類するモデル



③評価
・Testデータを用いて評価
・90%



各業種におけるビッグデータの活用事例

事例(ミツカン):気象情報とツイート情報の活用

日本気象協会（JWA）とTwitter Japanは、気象データとツイートデータを組み合わせて商品の需要を予測するサービスを提供し、ミツカンはそのデータを活用。これにより、余剰在庫を削減できた。
ツイートデータ活用で算出しているのは「体感指数」。気温だけで需要を分析すると、同じ気温なら分析結果も同じになるが、実際には、同じ30度でも5月の30度と8月の30度では湿度などが作用して感じ方が違うので、体感の数値化に取り組んでいる。

概要

- 主体：ミツカン
- 業種：食品
- 課題：
  ✓冷やし中華は基本的に夏にしか食べないので、夏の終わりに売れ残ったつゆはロスになってしまうことが課題。
- 活用しているデータ：気象データとツイートデータ
- 活用しているツール・技術：日本気象協会（JWA）とTwitter Japanの気象データとツイートデータを組み合わせて商品の需要を予測するサービス
- データ活用の体制：日本気象協会（JWA）とTwitter Japanは、気象データとツイートデータを組み合わせて商品の需要を予測するサービスを提供し、ミツカンはそのデータを活用している。

商品の需要＝気象×「暑い」ツイート数

効果、課題等

- 効果
  ✓商品(冷やし中華）の余剰在庫を35%削減できた。
- 課題：気温は地域によって異なるので、体感指数も当然地域ごとに変わる。ツイート数は人口に比例するので、現在の収集方法だと人口が少ない地方のツイートデータが足りない。そのため今後は、能動的にデータを収集する方法を考えなければならない。

「暑い」ツイートの時系列

Source: https://www.soumu.go.jp/johotsusintokei/linkdata/r02_05_houkoku.pdf