Empowered by Innovation NEC

日本リアルオプション学会
「価値創造のイノベーションと戦略」研究部会

ビッグデータ時代における ビジネス向け機械学習

2014年10月27日

日本電気株式会社
中台慎二、森永聡

自己紹介

森永　聡

1994 **東京大学大学院　修士課程修了**
　　NEC入社
1999 **工学博士(東京大学)**
2000～01 **金融庁出向** (05～08 **兼務**)
2014**現在**　**情報・ナレッジ研究所　部長**
　　(データ＆テキストマイニング)

中台　慎二

2003 **東京大学大学院　修士課程修了**
　　NEC入社
2006～08　**金融デリバティブのHPC適用** **(スパコン)**
2012～13　**UC Berkeley　客員研究員**
2014**現在**　**情報・ナレッジ研究所　主任**
　　(データ＆テキストマイニング)

はじめに

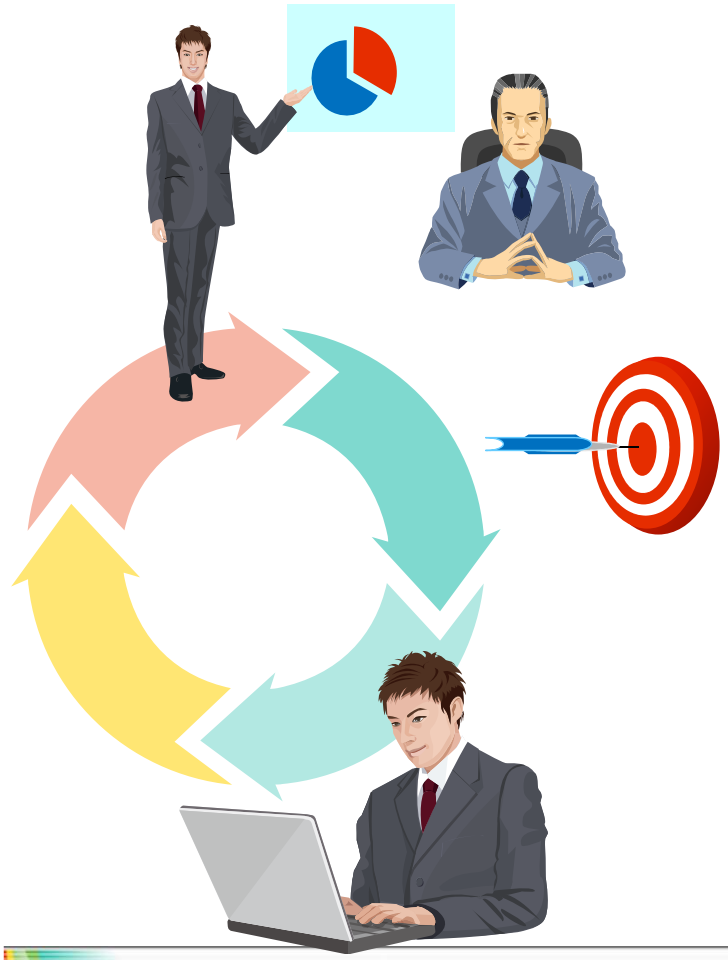
巷のキーワード： ビッグデータとデータサイエンティスト

$$\gamma_{ijk} \propto \exp(\mathbb{E}_{q(z^{-ij})}[p(x, z^{-ij}, z_{ij} = k)])$$

$$\gamma_{ijk} \propto \frac{n_{kw}^{-ij} + \beta}{n_{k}^{-ij} + V\beta}(n_{jk}^{-ij} + \alpha)$$

本日のテーマ1

データサイエンティストにとっての 強力な武器（学習エンジン）とは？

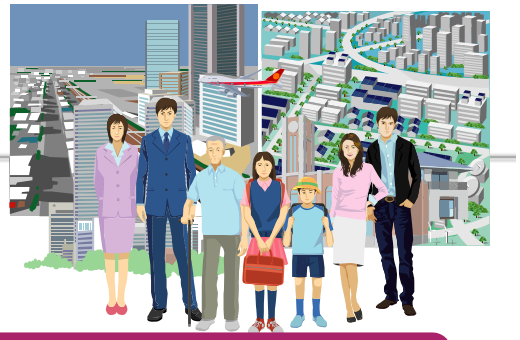


- 説明できる
- 精度が良い
- 速く回せる

異種混合

配布資料なし

本日のテーマ2



ビジネスへの適用事例のご紹介

電力需要予測

商品需要予測

適正価格予測

品質予測

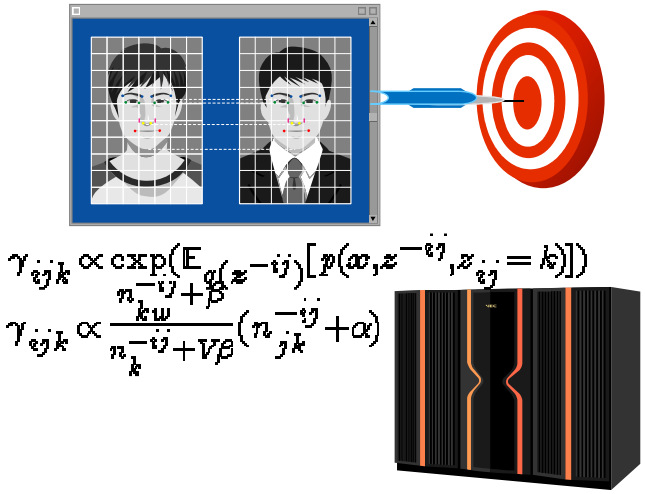
劣化予測

離反予測

配布資料なし

本日のテーマ3

機械学習の限界： 人の感覚は分からない



$$\gamma_{ijk} \propto \exp(\mathbb{E}_{q(z^{-ij})}[p(w, z^{-ij}, z_{ij} = k)])$$

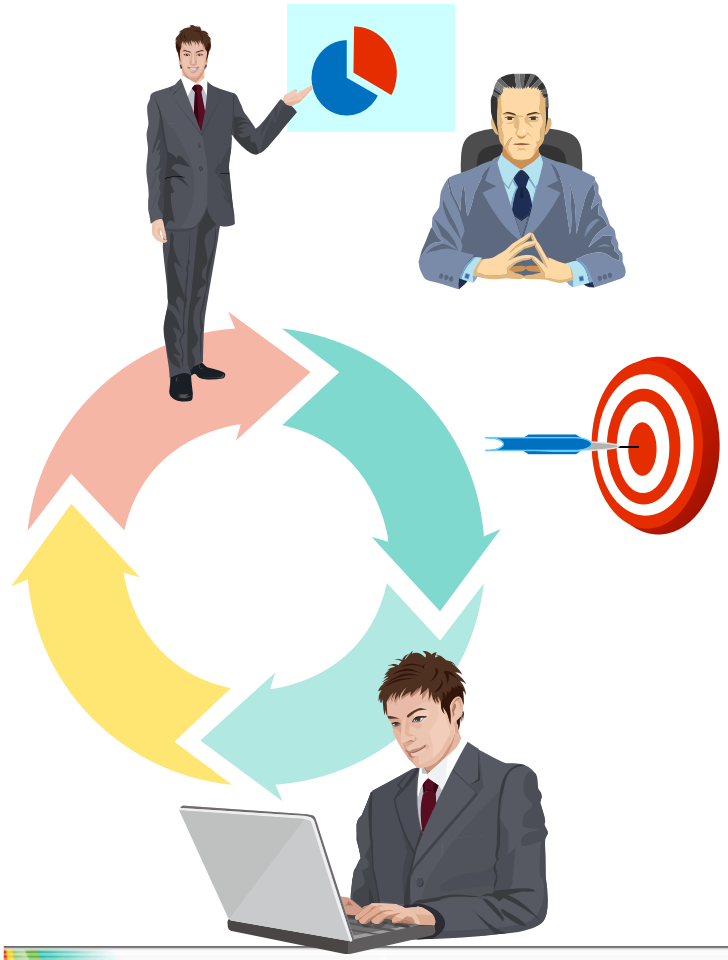
$$\gamma_{ijk} \propto \frac{n_{kw}^{-ij} + \beta}{n_k^{-ij} + V\beta}(n_{jk}^{-ij} + \alpha)$$

機械学習

Crowdsourcing

本日のテーマ1

データサイエンティストにとっての強力な武器（学習エンジン）とは？



説明できる

精度が良い

速く回せる

異種混合

データサイエンティストの分析プロセス

継続的なPDCA

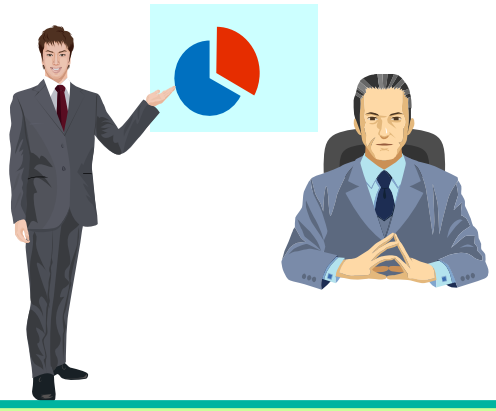
ビジネス課題の認識

課題のモデル化

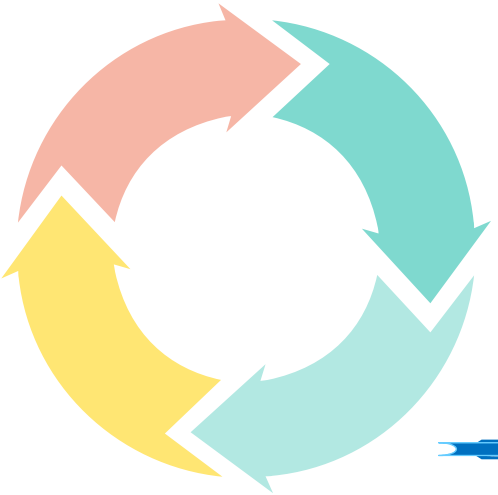
分析エンジンの実行

評価検証

レポーティング



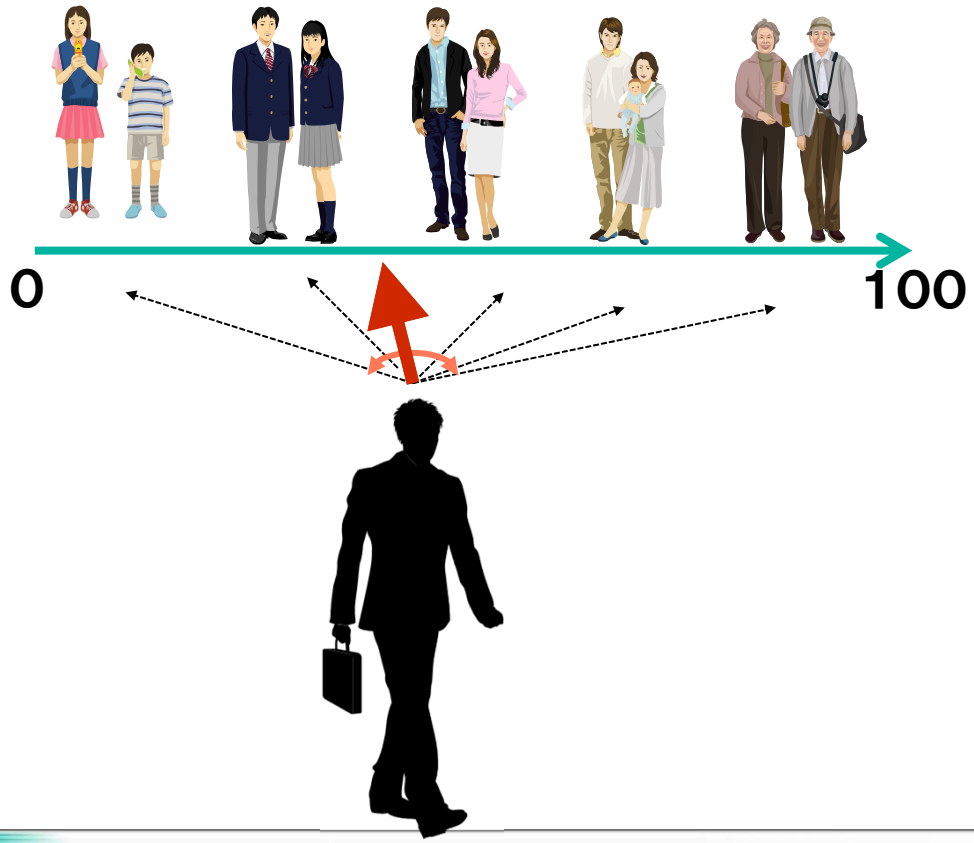
修正



【用語】 課題のモデル化： 回帰、判別

回帰 ： 連続値をあてる

- 例）0～100
- 年齢



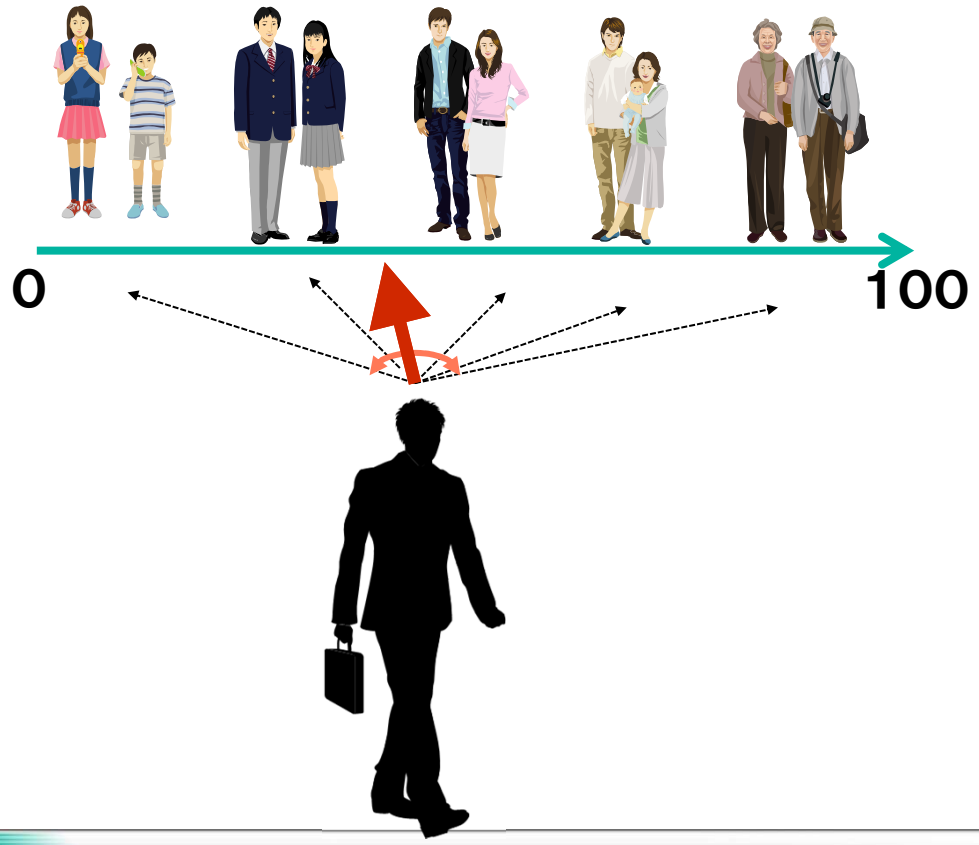
判別 ： 離散値をあてる

- 例）0 or 1
- 性別

【用語】 課題のモデル化：目的／説明変数

目的変数 ： 当てる対象

- 例）0～100
- 年齢



説明変数 ： 当てる為に使う変数

- 例）

- 画像データ
  - 図のシルエット
- 食事
  - お菓子、肉、洋食、和食

- 話し言葉
  - 僕、オレ、私、わし、
- etc

本日のテーマ1

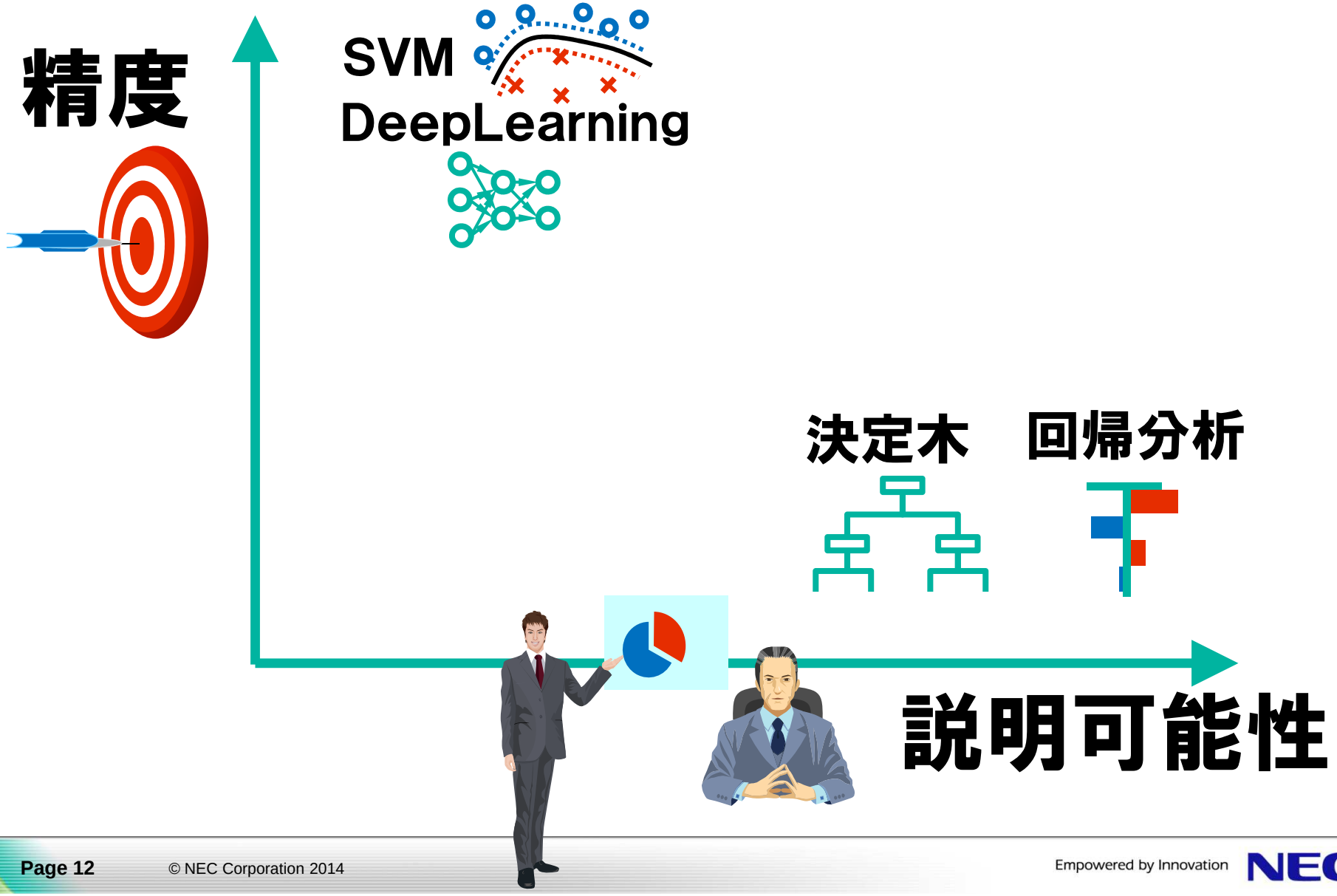
データサイエンティストにとっての強力な武器（学習エンジン）とは？

両立

- 説明できる
- 精度が良い
- 速く回せる

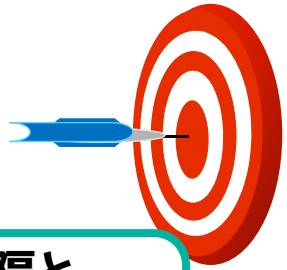
異種混合

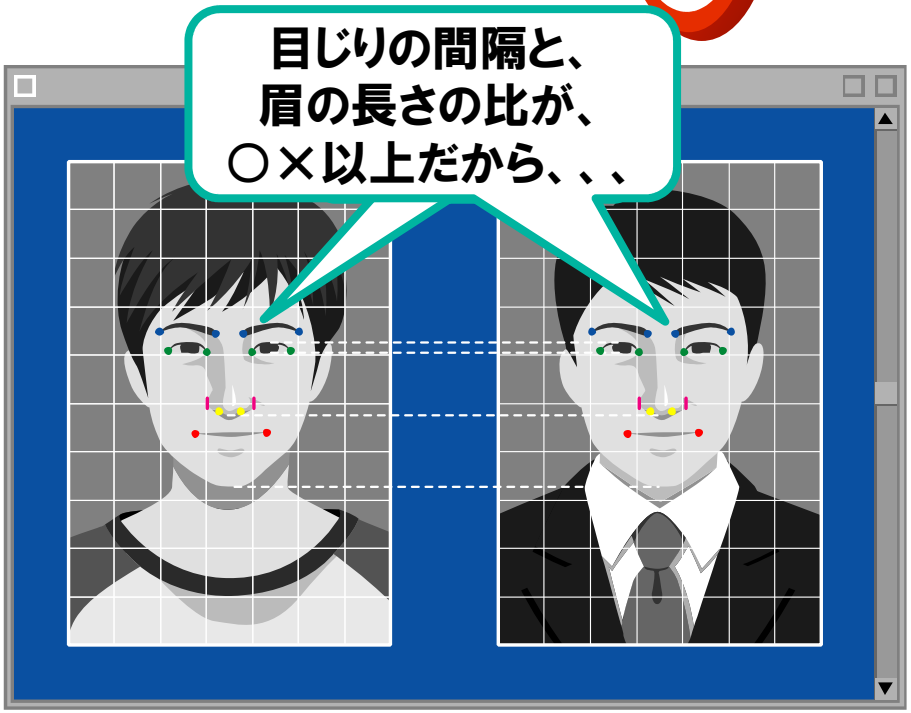
学習エンジンのマップ



学習エンジンの特徴に応じた使い分け

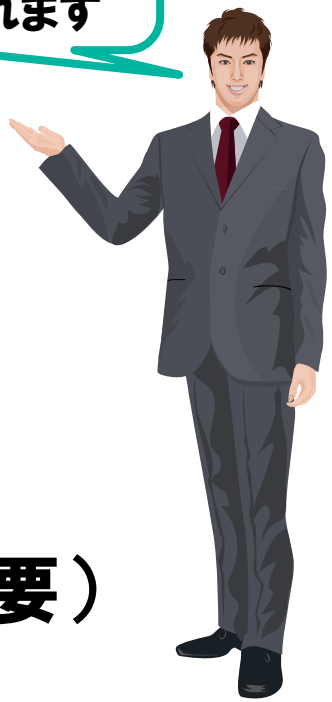
精度





画像認識など
（判定理由不要）

説明可能性（可読性）

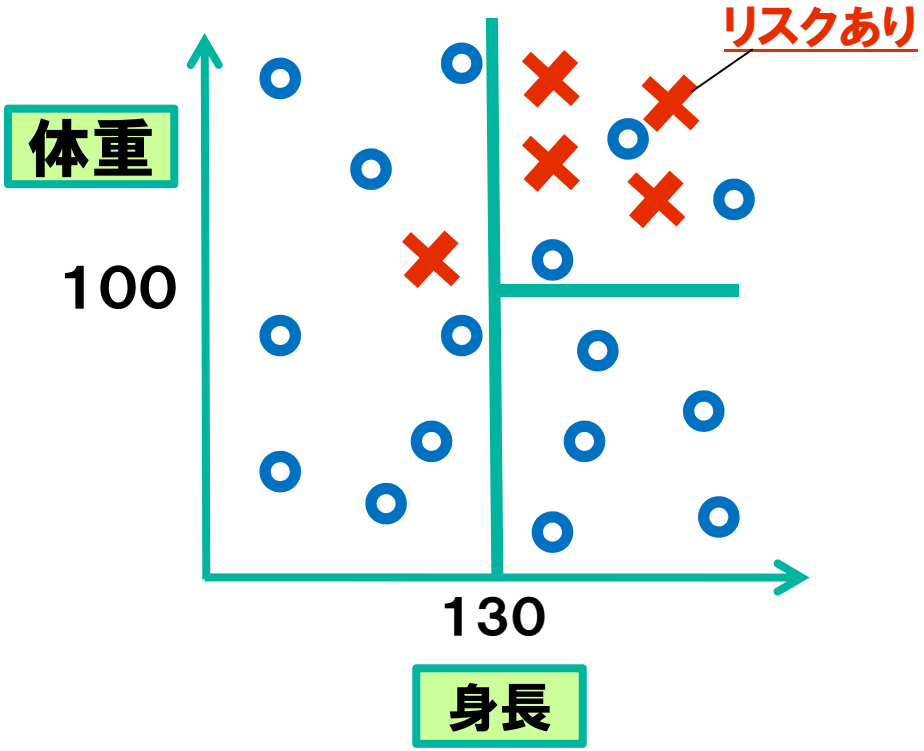


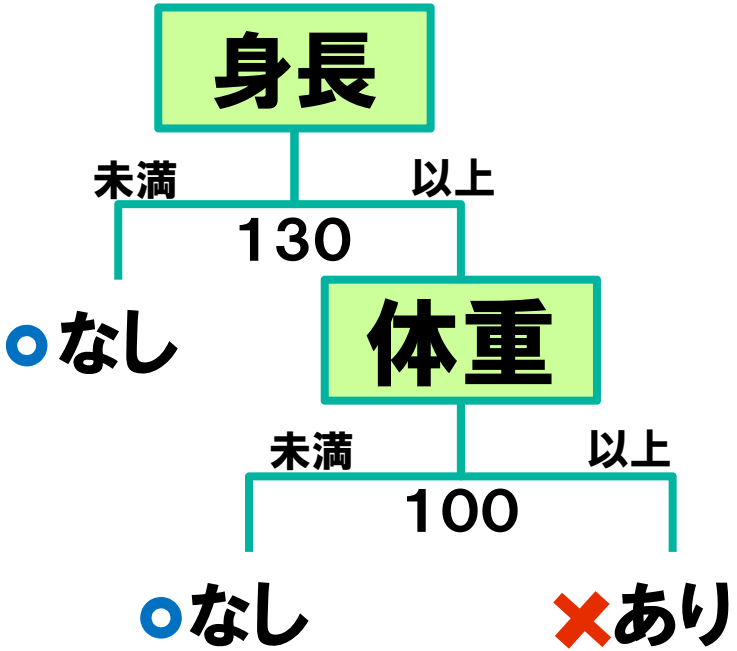
ビジネス分析
（レポーティングが重要）

可読性のある学習１： 決定木

例： 慎重と体重から、生活習慣病リスク

※説明のための疑似データ

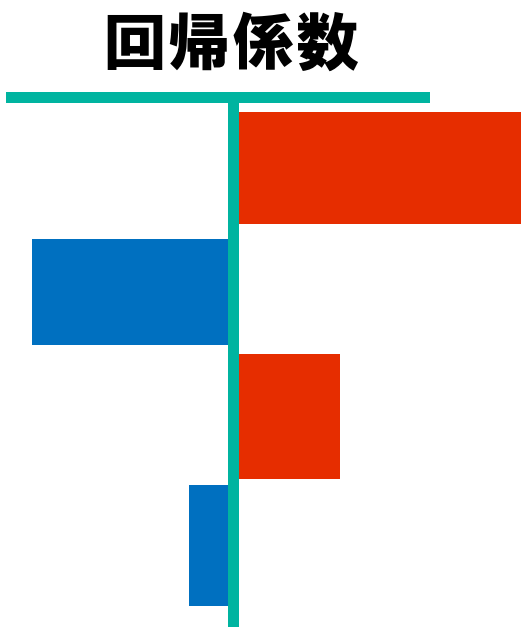




可読性のある学習２： 回帰分析

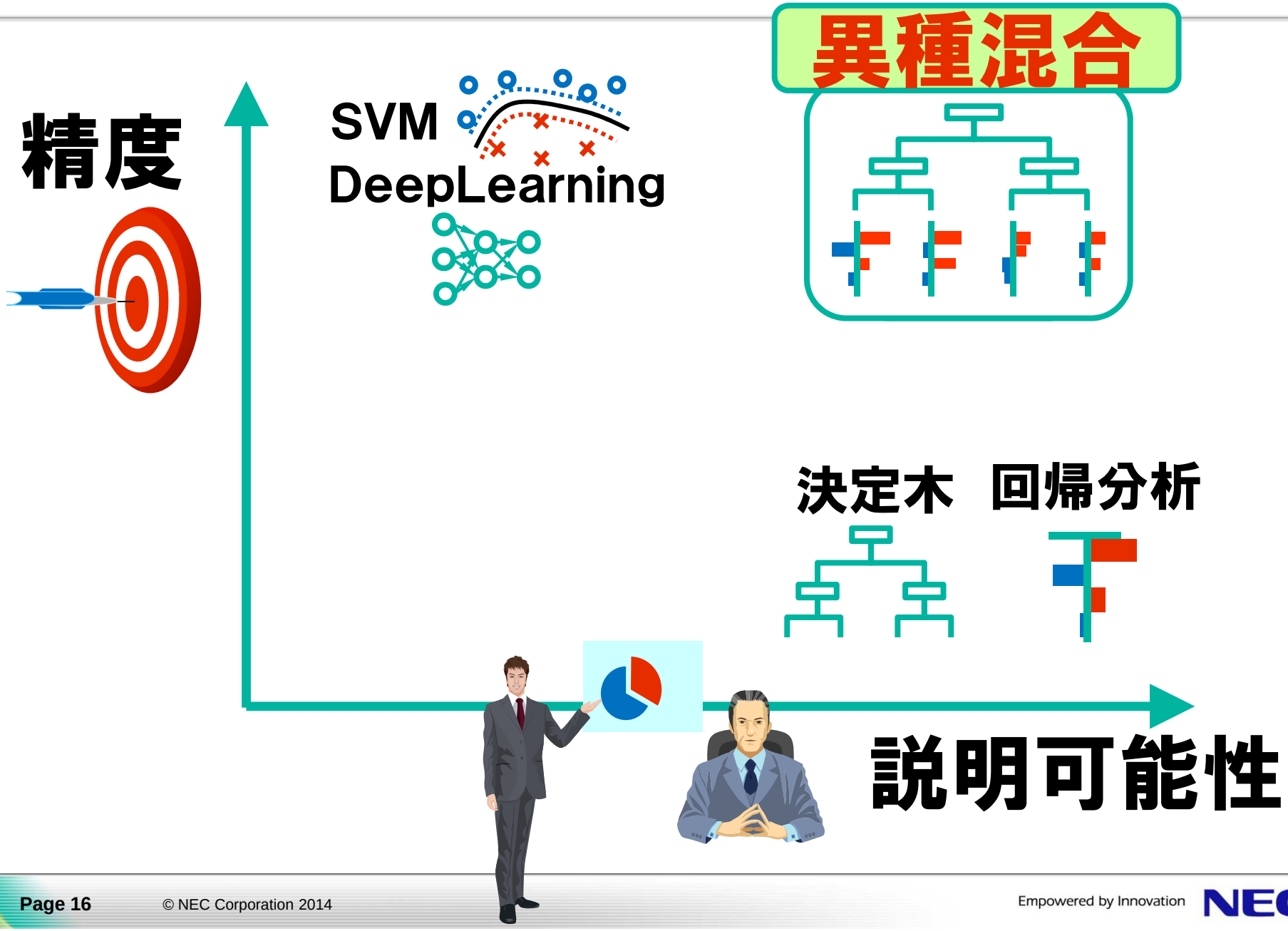
例

- 生活習慣病の発症率＝
  5×体重
  −3×運動日数
  +2×年齢
  −1×野菜を食べる



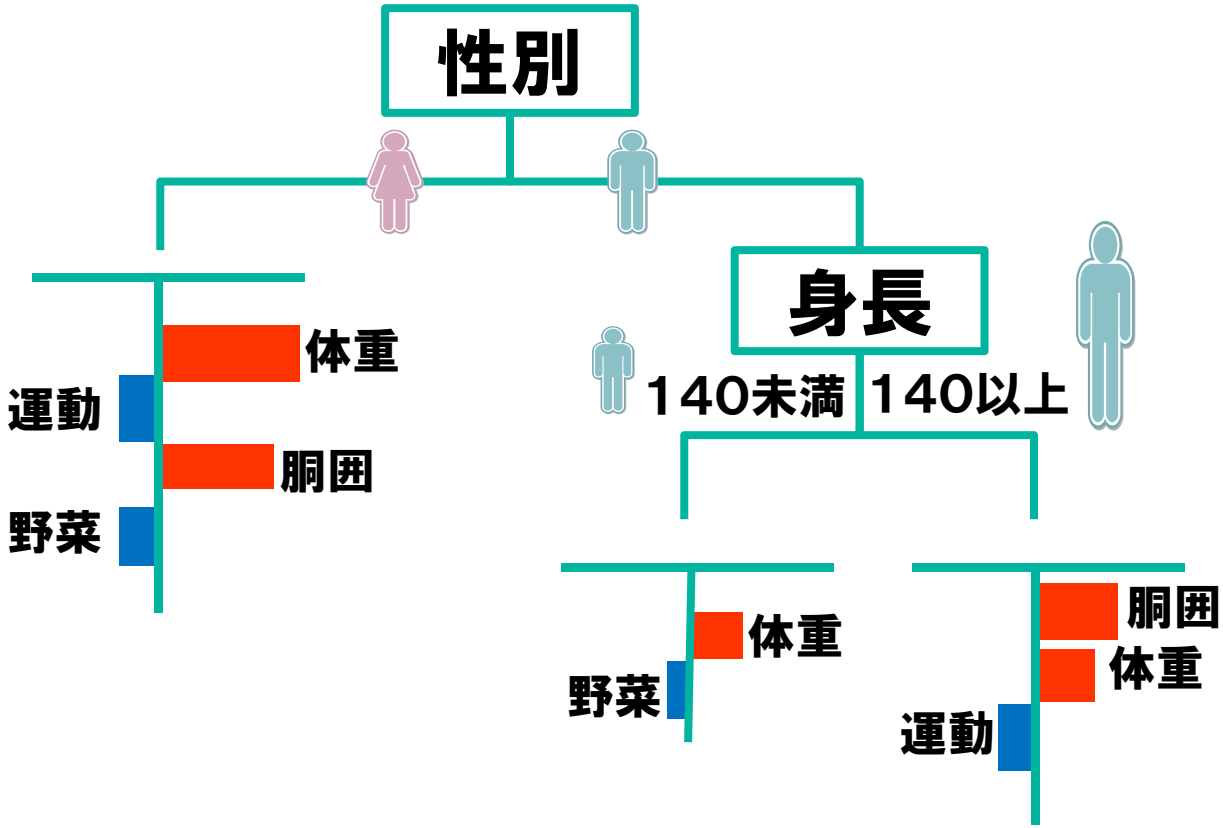
※説明のための疑似データ

学習エンジンのマップ



決定木の各葉に回帰式

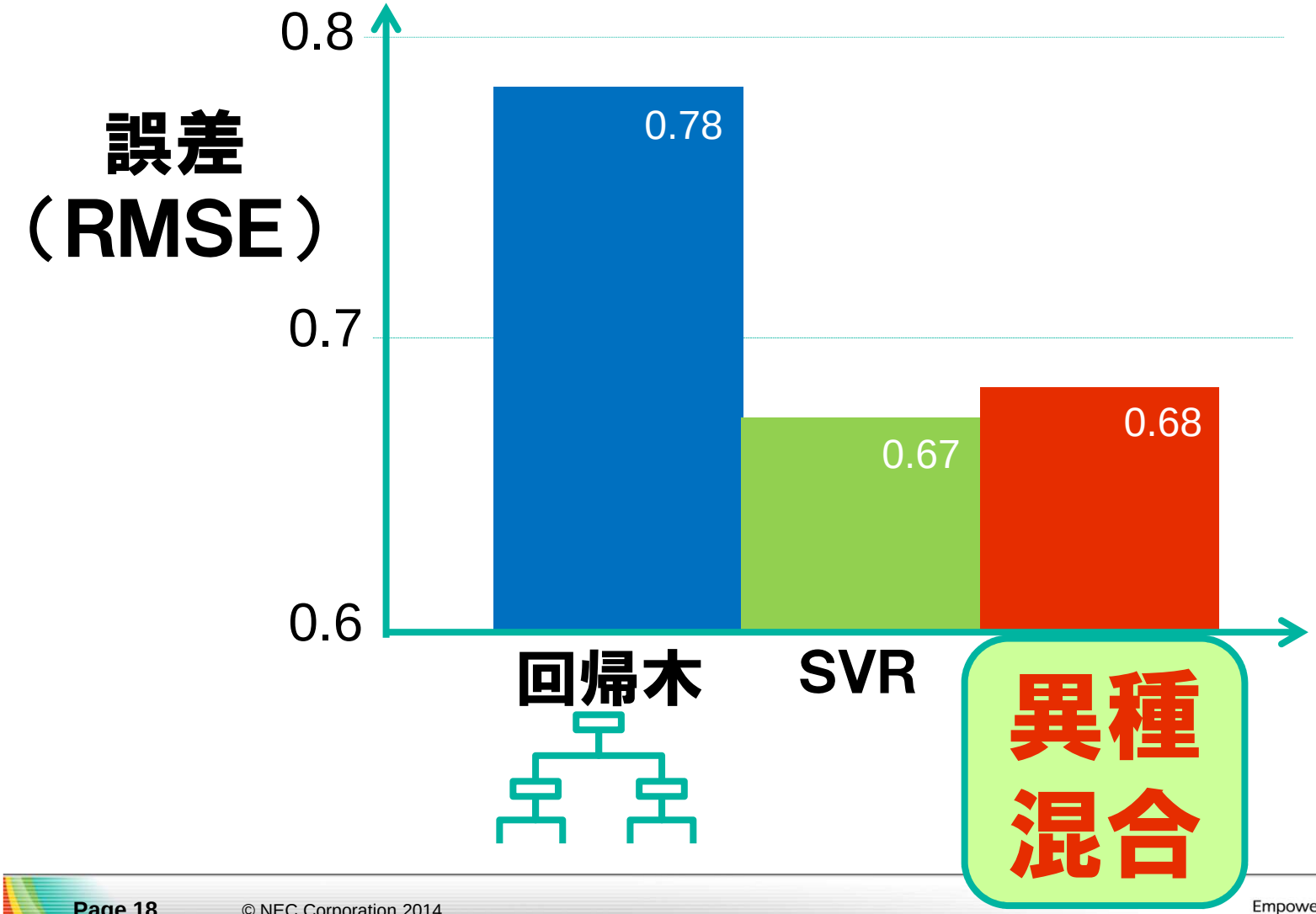
生活習慣病
～　成人病



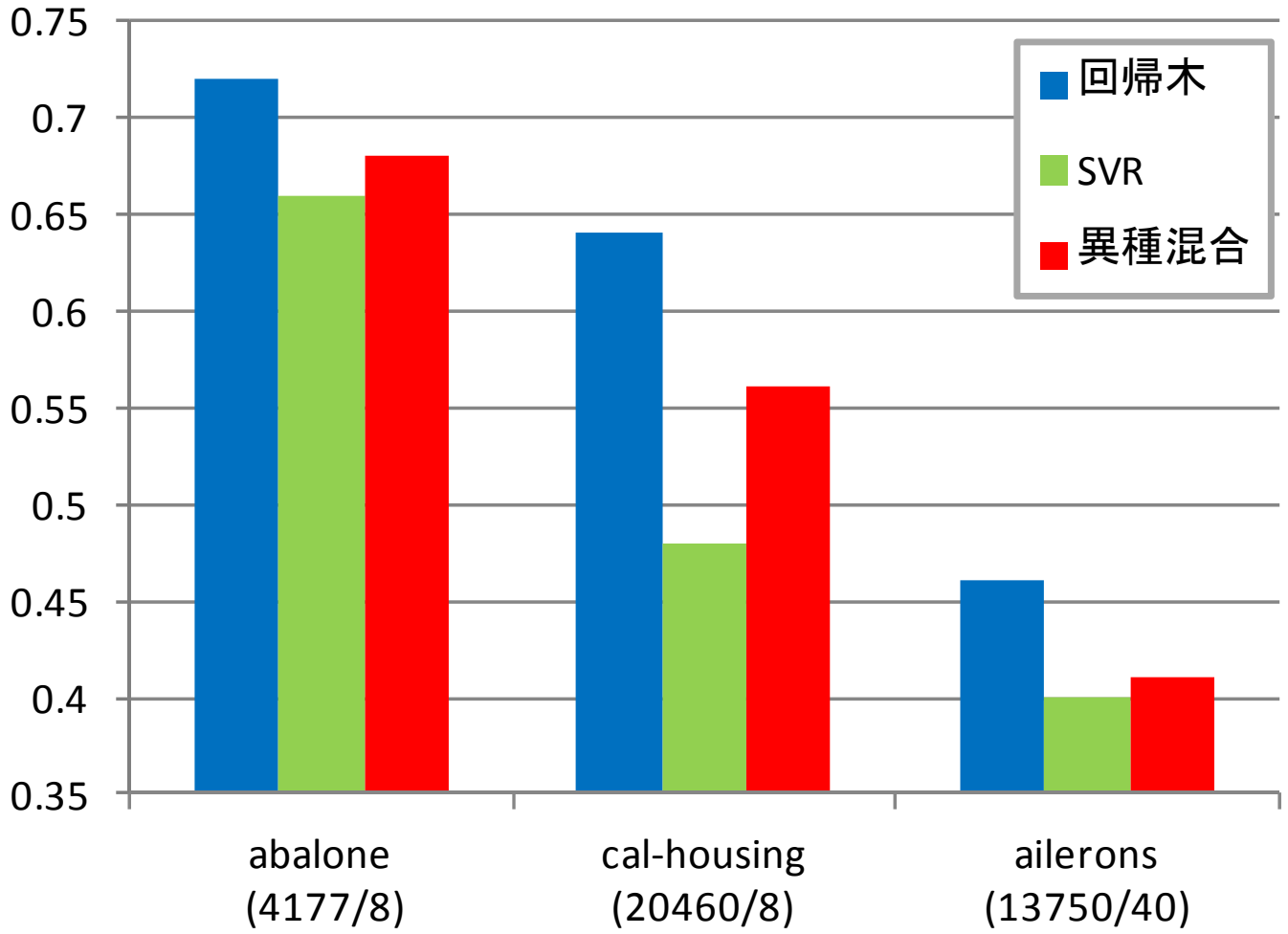
※説明のための疑似データ

精度比較（回帰）

データ： bank32nh（8192**サンプル**／32**次元**）

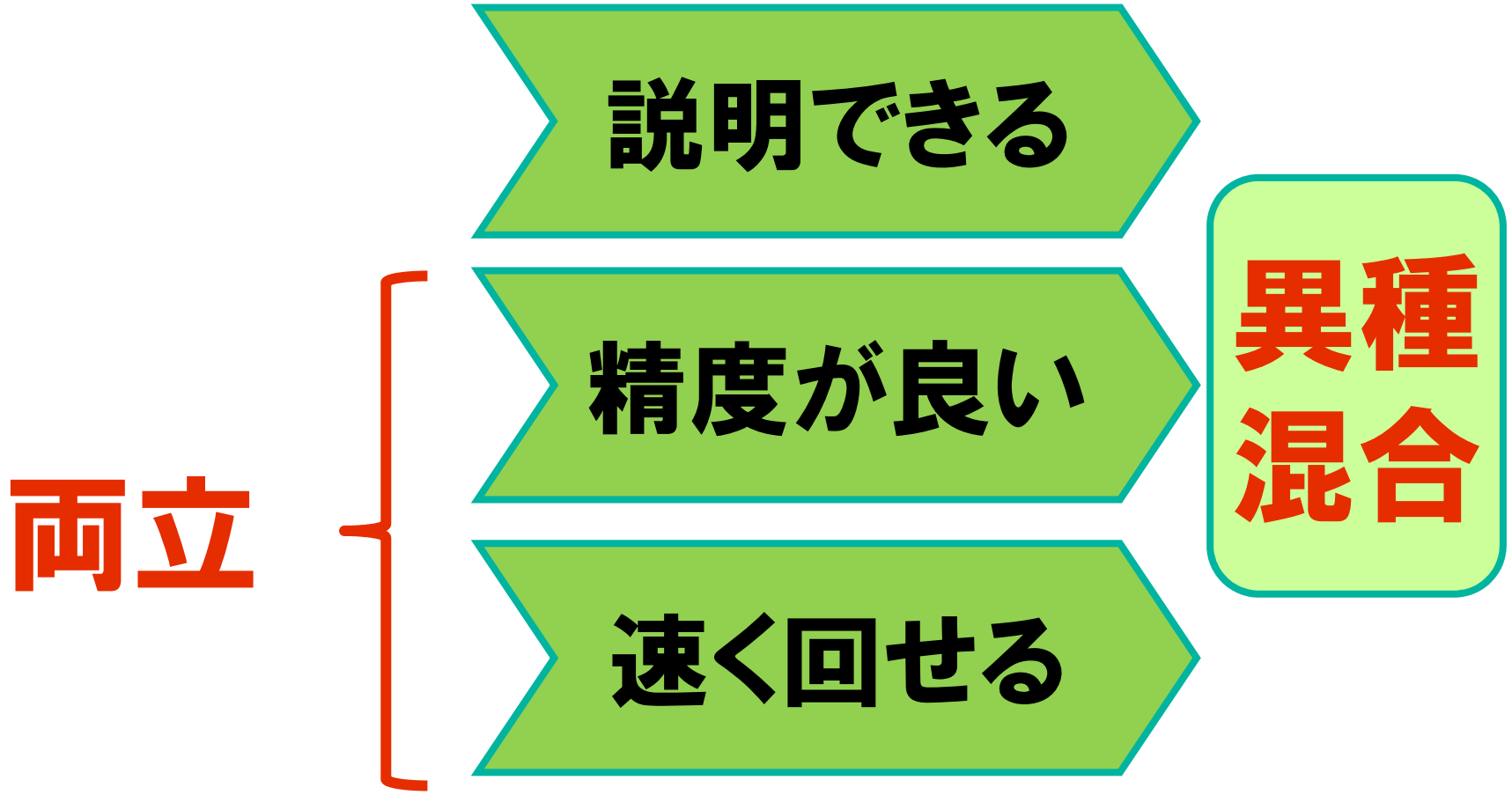


他の実験結果



本日のテーマ1

データサイエンティストにとっての強力な武器（学習エンジン）とは？



一般に、精度を上げるには、時間が必要

ビジネス分析でも、精度は、やはり重要

保険の勧誘

- 子供が生まれたタイミングで、保険に入る人が多い
- 「結婚した人」を、クレジットカードのデータから検知し、アプローチ

価値の試算例

1%↑ ＝ 1，000万円↑

精度を上げるには
どうしたら良いんだ？

他にも、分析タスクは、
山ほどあるのに

精度を上げるために悩むこと

説明変数の設計

説明変数の
組を変えてみる

発症率 ＝ BMI ＋ 年齢

or

発症率 ＝ 身長 ＋ 体重

パラメータのチューニング

学習エンジンには、
様々なつまみがあり、
調整してみる

中
弱 強

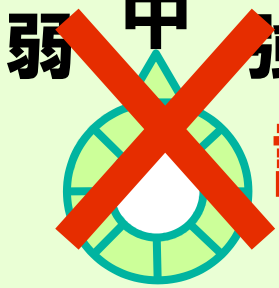
異種混合なら

説明変数の設計

説明変数の候補だけを入れればよい（組合せは不要）

パラメータのチューニング

自動で調整



調整不要

説明変数の設計における組合せ爆発

生活習慣病＝　体重　＋　身長

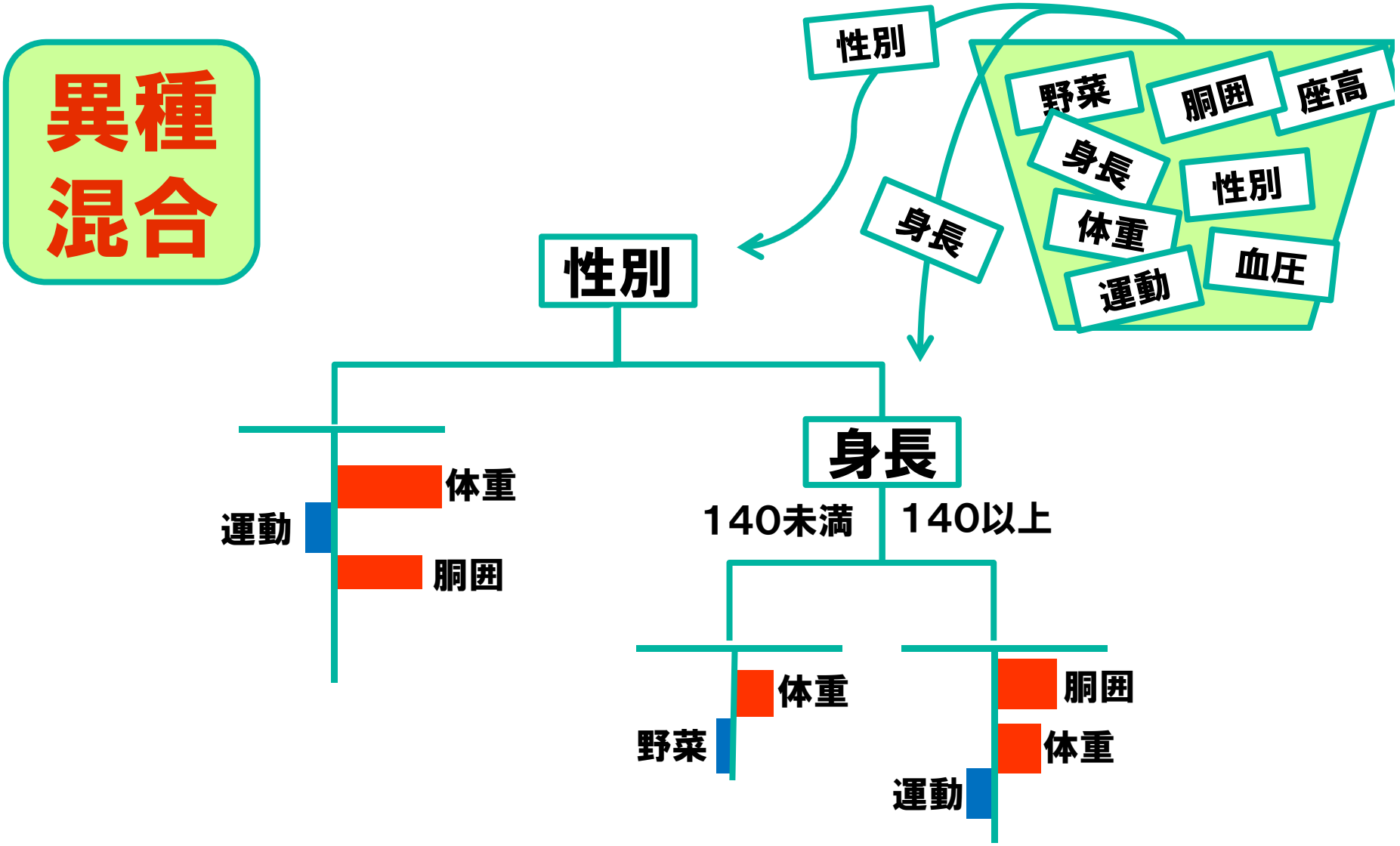
身長と体重じゃなくて、
BMI＝体重／身長^2
が効くんじゃない？

生活習慣病＝　BMI　＋　体重

生活習慣病＝　BMI　＋　体重　＋　身長

異種混合なら

自動で説明変数を組合せる



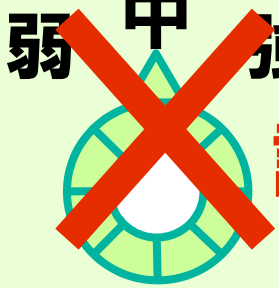
異種混合なら

説明変数の設計

説明変数の候補だけを入れればよい（組合せは不要）

パラメータのチューニング

自動で調整



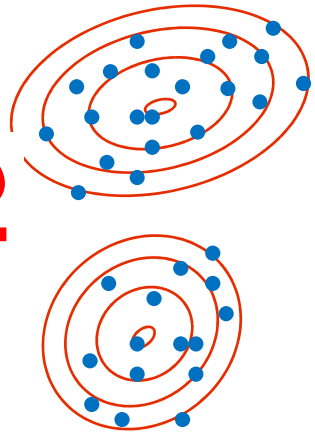
調整不要

様々なパラメータ

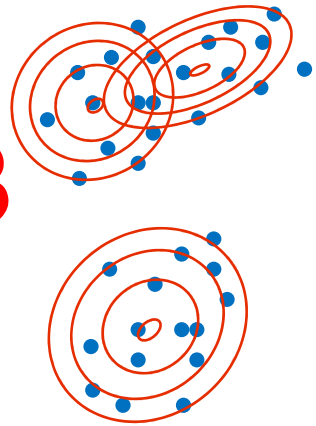
複雑性を決めるパラメータの
チューニングが**大変**なんだよね

例） 混合ガウス／K-meansの**クラスタ数K**

K＝2



K＝3

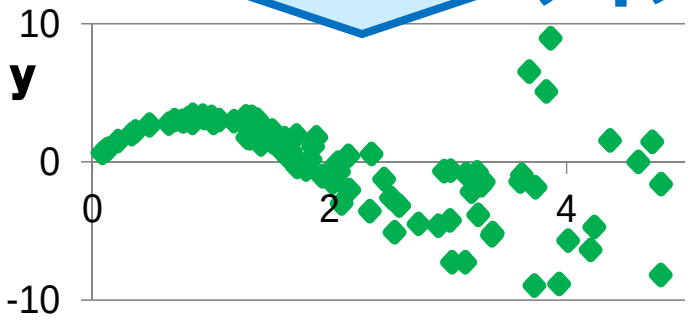


複雑性に関する回帰の実験

答え：$y = x^3 - 6x^2 + 8x$

真のモデル

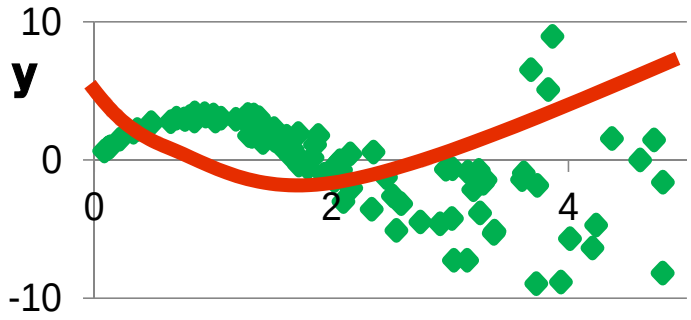
ノイズつきデータ生成

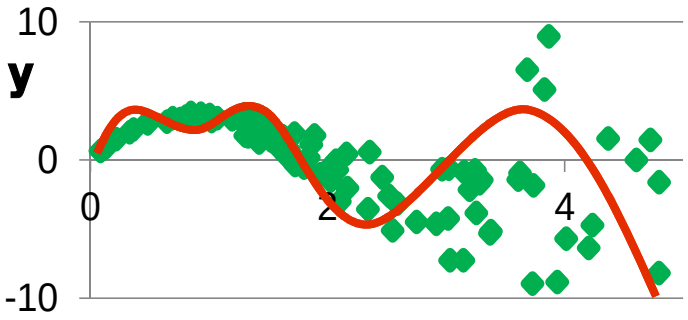


一致する？

シンプル　複雑性　複雑

回帰

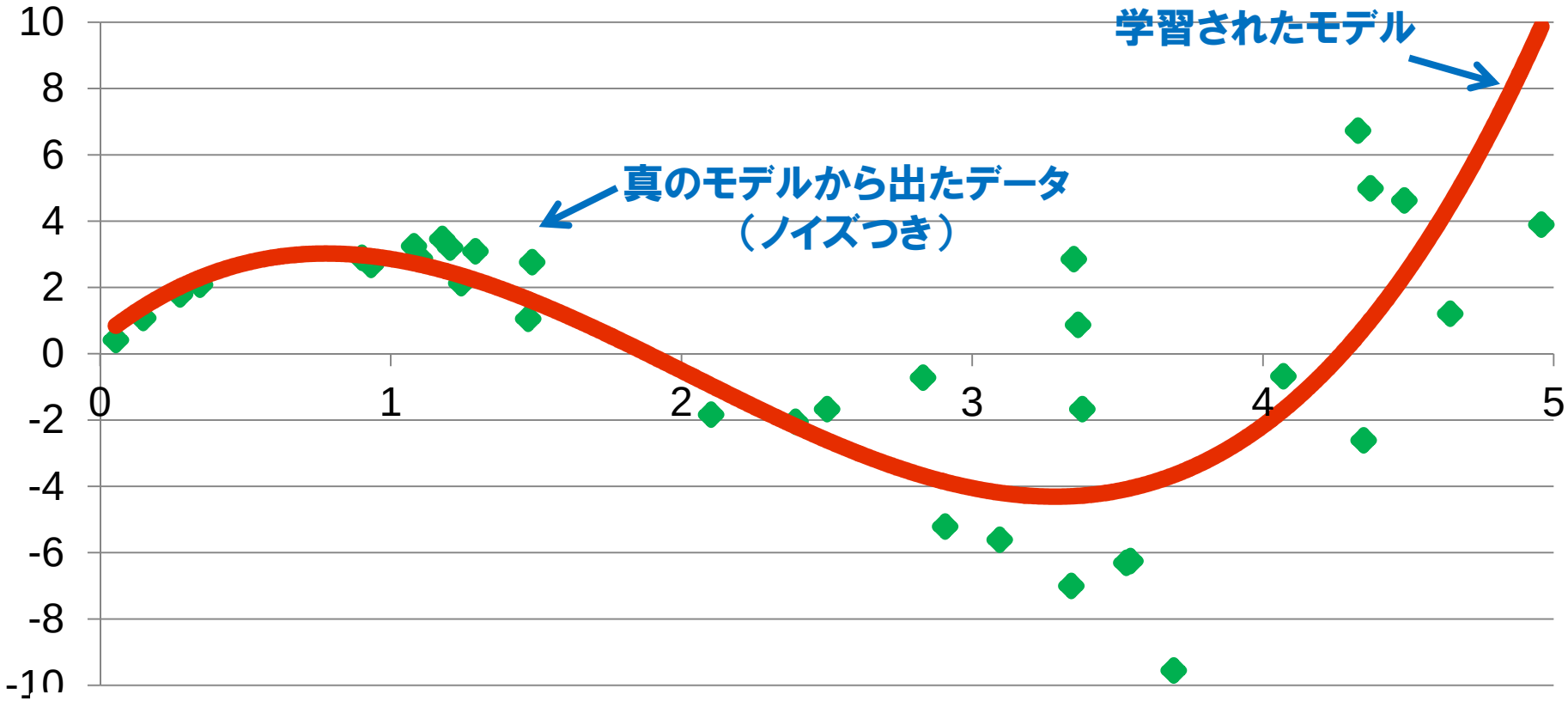




適切な複雑性で、学習できている場合

答え：$y = x^3 - 6x^2 + 8x$ ← 真のモデル

Y の値　　$y = 0.9\,x^3 - 5.7\,x^2 + 7.1\,x + 0.5$



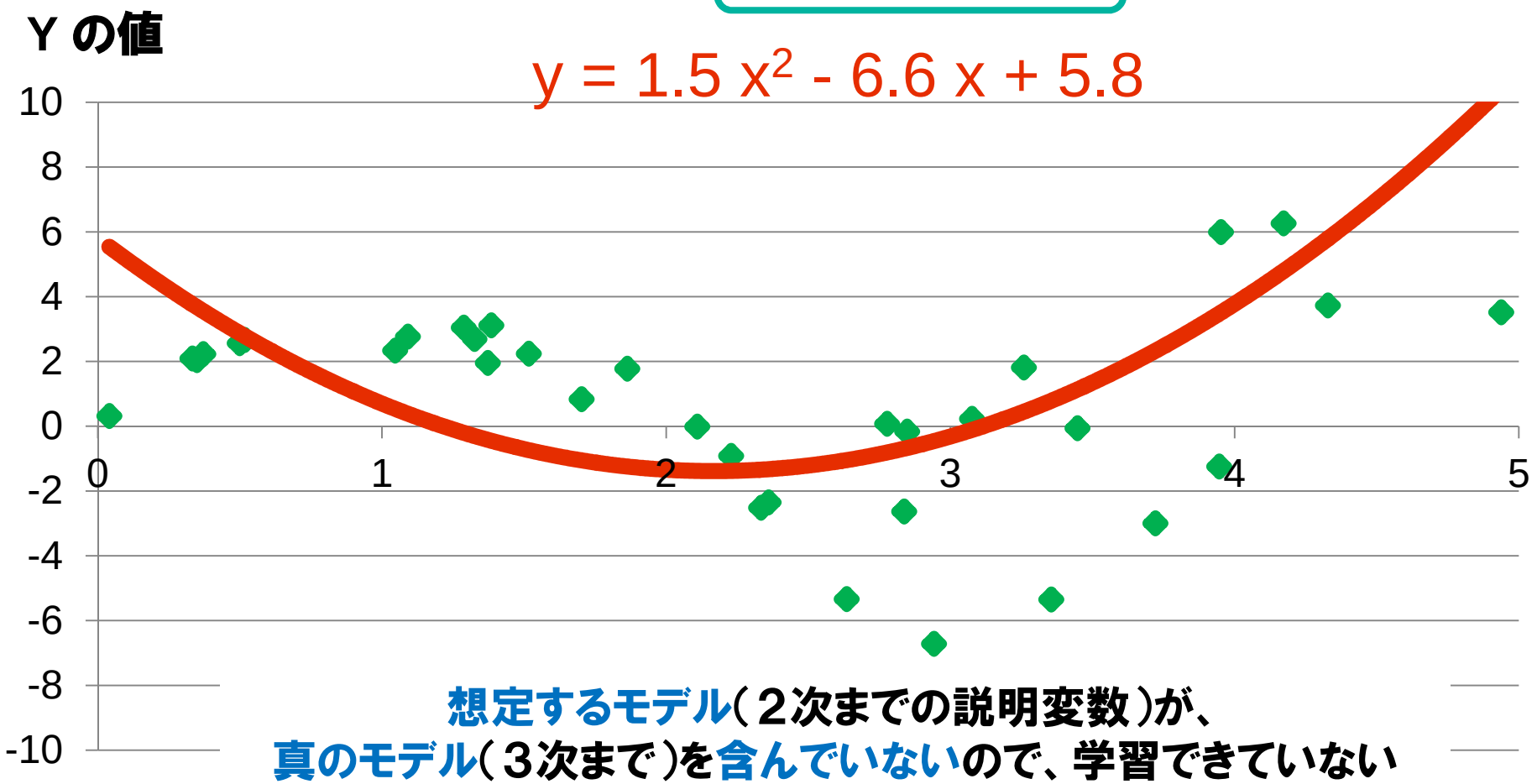
想定するモデル（３次までの説明変数）が、真のモデル（３次まで）を含んでいる

⇒　係数を、ほぼ当てることができる

シンプルすぎるモデルでは、当たらない

答え：$y = x^3 - 6x^2 + 8x$

必要な説明変数がない

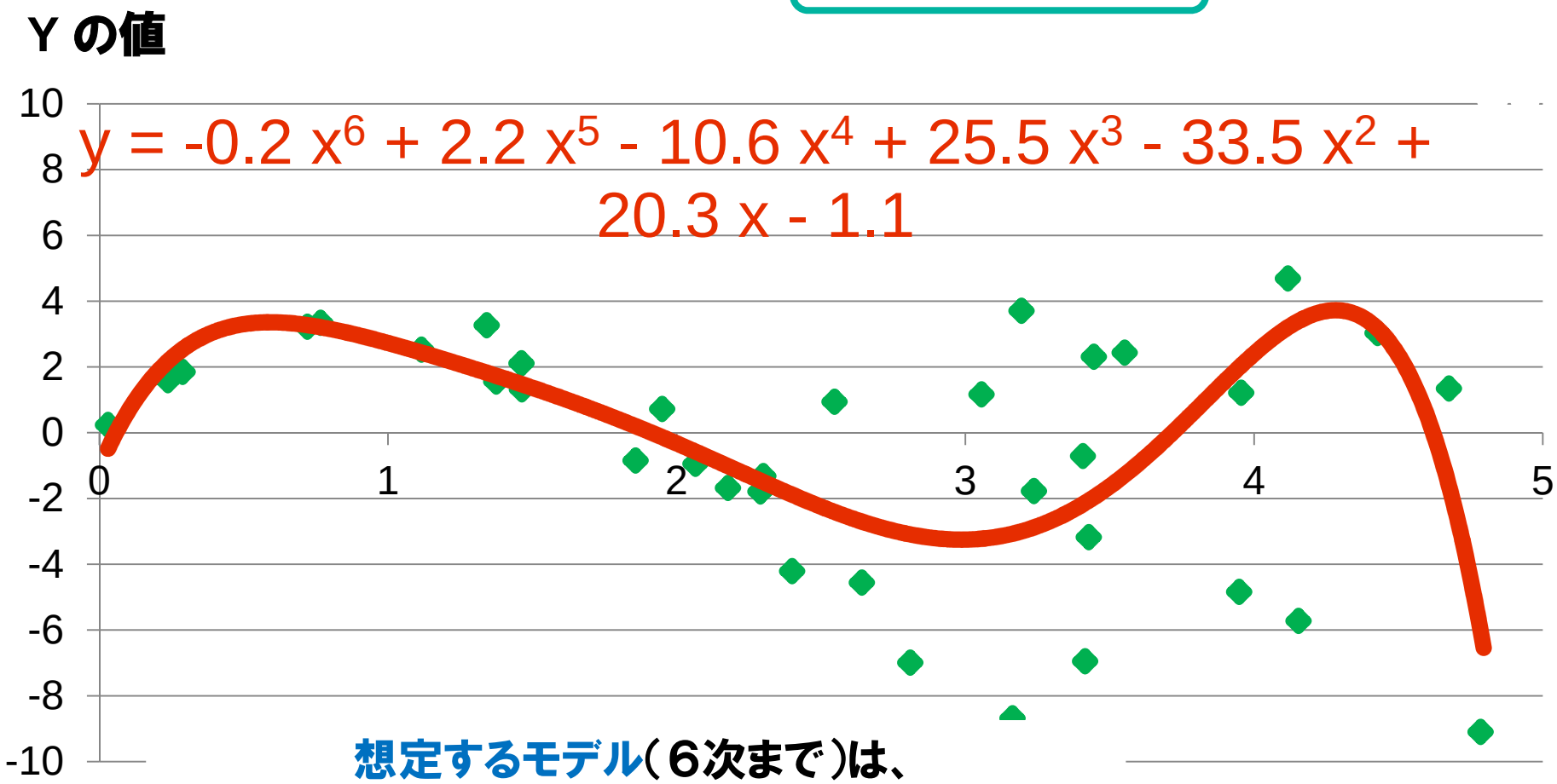


想定するモデル（２次までの説明変数）が、
真のモデル（３次まで）を含んでいないので、学習できていない

複雑な方が良い 訳ではない

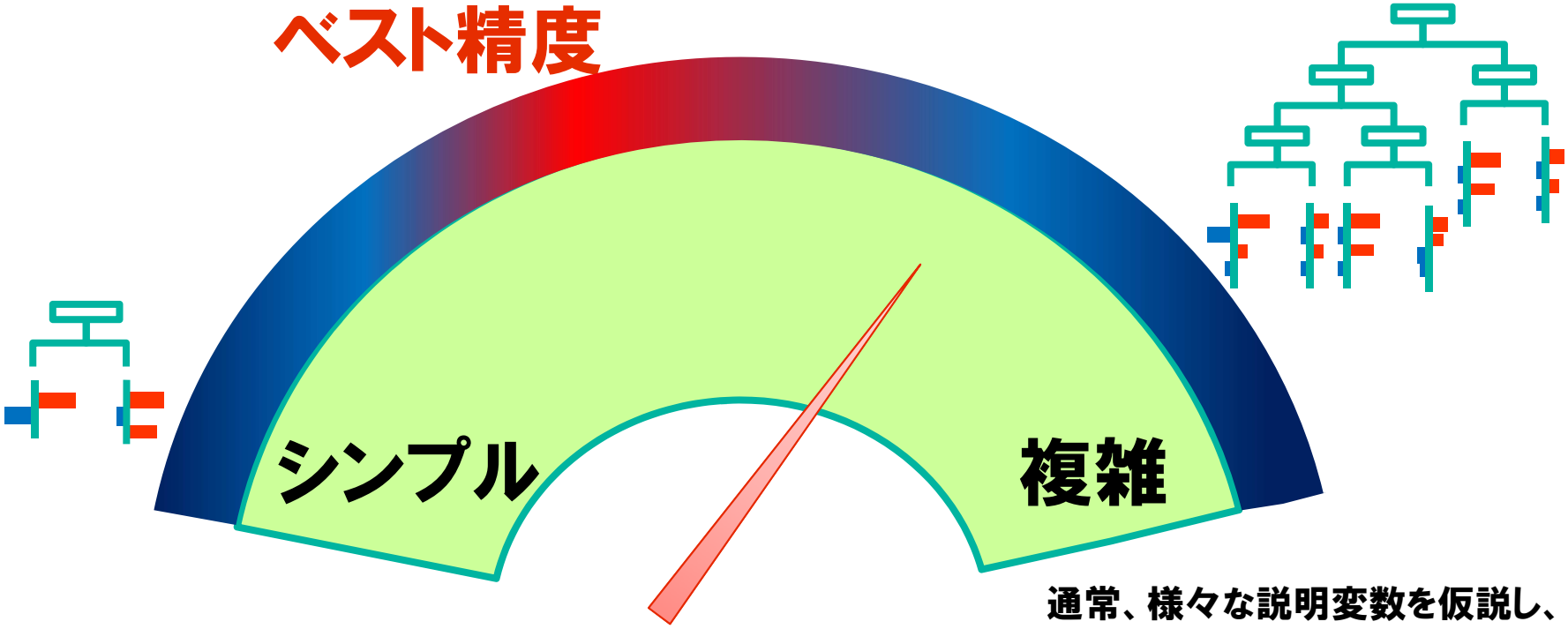
答え：$y = x^3 - 6x^2 + 8x$

無駄な説明変数がある



想定するモデル（６次まで）は、
真のモデル（１次まで）を含んでいるにも関わらず、
無駄な説明変数（２次～５次）があるため、
ノイズに引きずられ、過学習が起こる

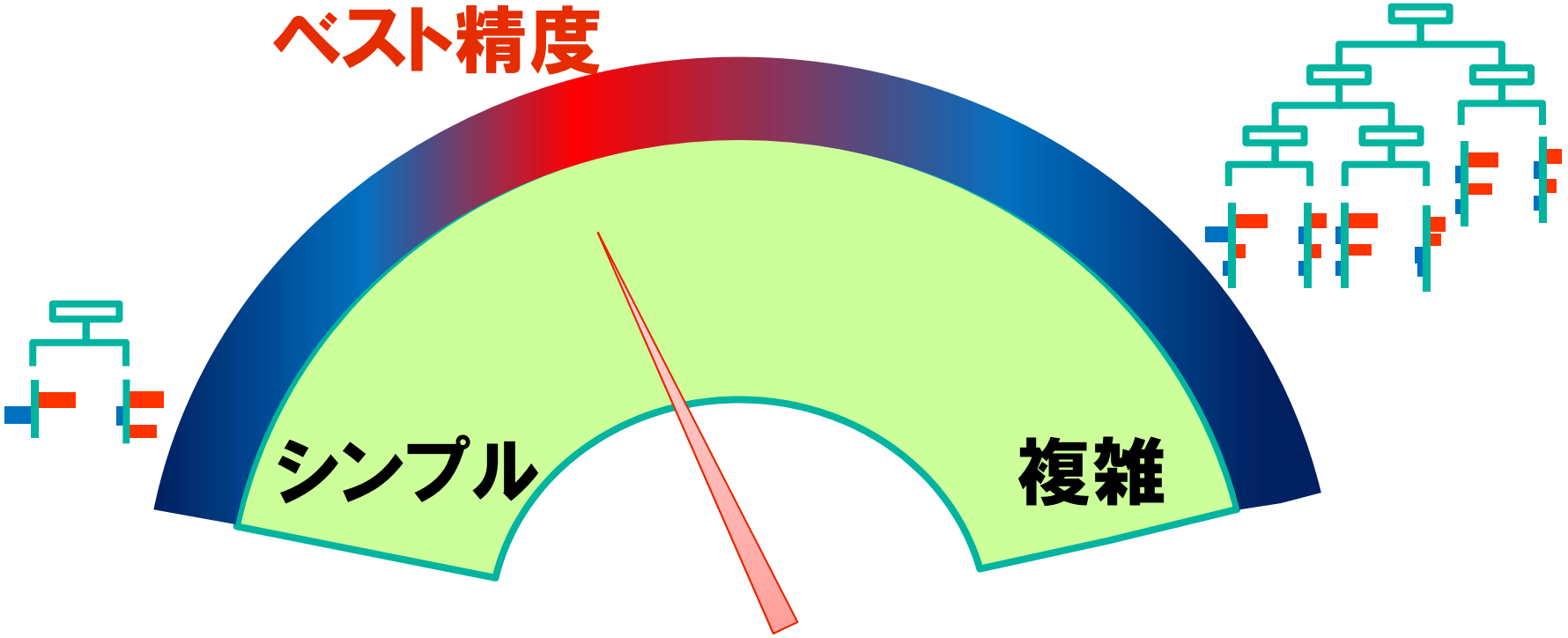
通常、「複雑性のつまみ」で調整する必要



通常、様々な説明変数を仮説し、
複雑性の1変数で制御する
（Lassoなど）

ベスト精度を
出すために、
調整しないと

適度な複雑さを、自動で決定

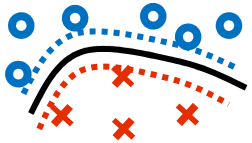


必要十分な
複雑さを
自動で選択

NECの機械学習

精度

SVM



SVM発明者
（Prof. Vapnik）

at&t → 2002 → NEC 北米研

異種混合

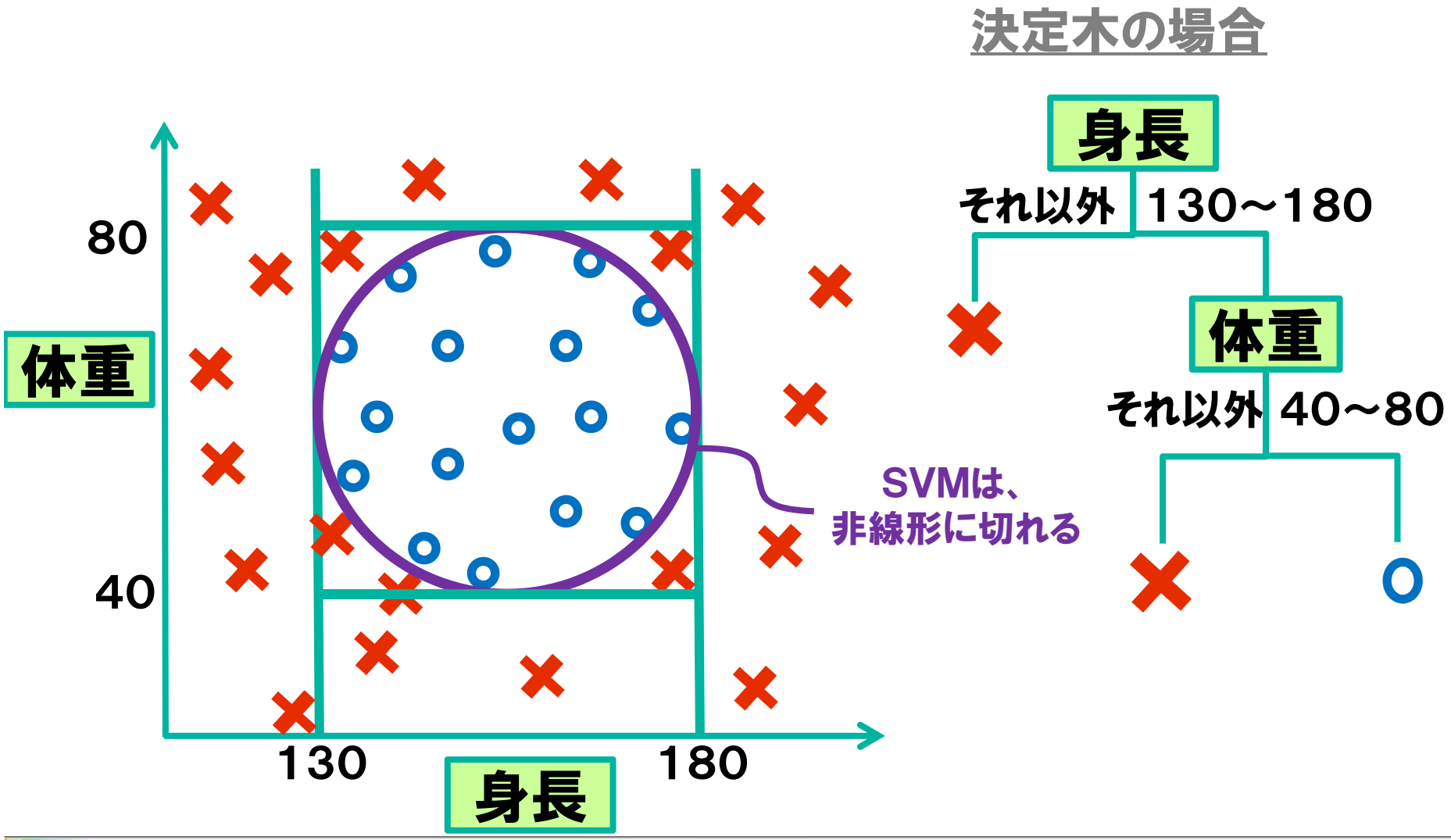
藤巻

NEC 北米研

機械学習のトップ学会
- AISTATS2012
- ICML2012
- NIPS2013
- ICML2014
- AISTATS2014
- NIPS2014

説明可能性

（ご参考） SVMの方が、精度が良い理由



Orchestrating a brighter world

世界の想いを、未来へつなげる。

未来に向かい、人が生きる、豊かに生きるために欠かせないもの。
それは「安全」「安心」「効率」「公平」という価値が実現された社会です。

NEC**は、ネットワーク技術とコンピューティング技術をあわせ持つ類のないインテグレーターとして**
リーダーシップを発揮し、卓越した技術とさまざまな知見やアイデアを融合することで、
世界の国々や地域の人々と協奏しながら、
明るく希望に満ちた暮らしと社会を実現し、未来につなげていきます。

Empowered by Innovation