SENDAI Lab

[オンライン開催]
機械学習と異常検知第1回

データサイエンス仙台ラボ

© 2023 Data Science Sendai Laboratory.

SENDAI Lab

データサイエンス仙台ラボ　存在意義とミッション

データサイエンス仙台ラボ

経営

支援

展開

[東京] コグラフ株式会社

データ分析

[仙台] 株式会社
イメージア・ソリューション

システム開発

データサイエンス

データ処理、データ分析、価値創造

データサイエンスを支える技術基盤

デジタルトランスフォーメーション(DX)
人工知能(AI)　　ビッグデータ

普及

連携

展望

仙台圏の企業

仙台圏の大学等

**仙台圏を
データサイエンスの先進地域に**

© 2023 Data Science Sendai Laboratory.

データサイエンス仙台ラボ　勉強会について

SENDAI Lab

データサイエンス仙台ラボ

展開

業界や業種を問わない
あらゆる分野への浸透

無償

興味やスキルの拡がり

データサイエンス

データ処理、データ分析、価値創造

データサイエンスを支える技術基盤

デジタルトランスフォーメーション(DX)
人工知能(AI)　　ビッグデータ

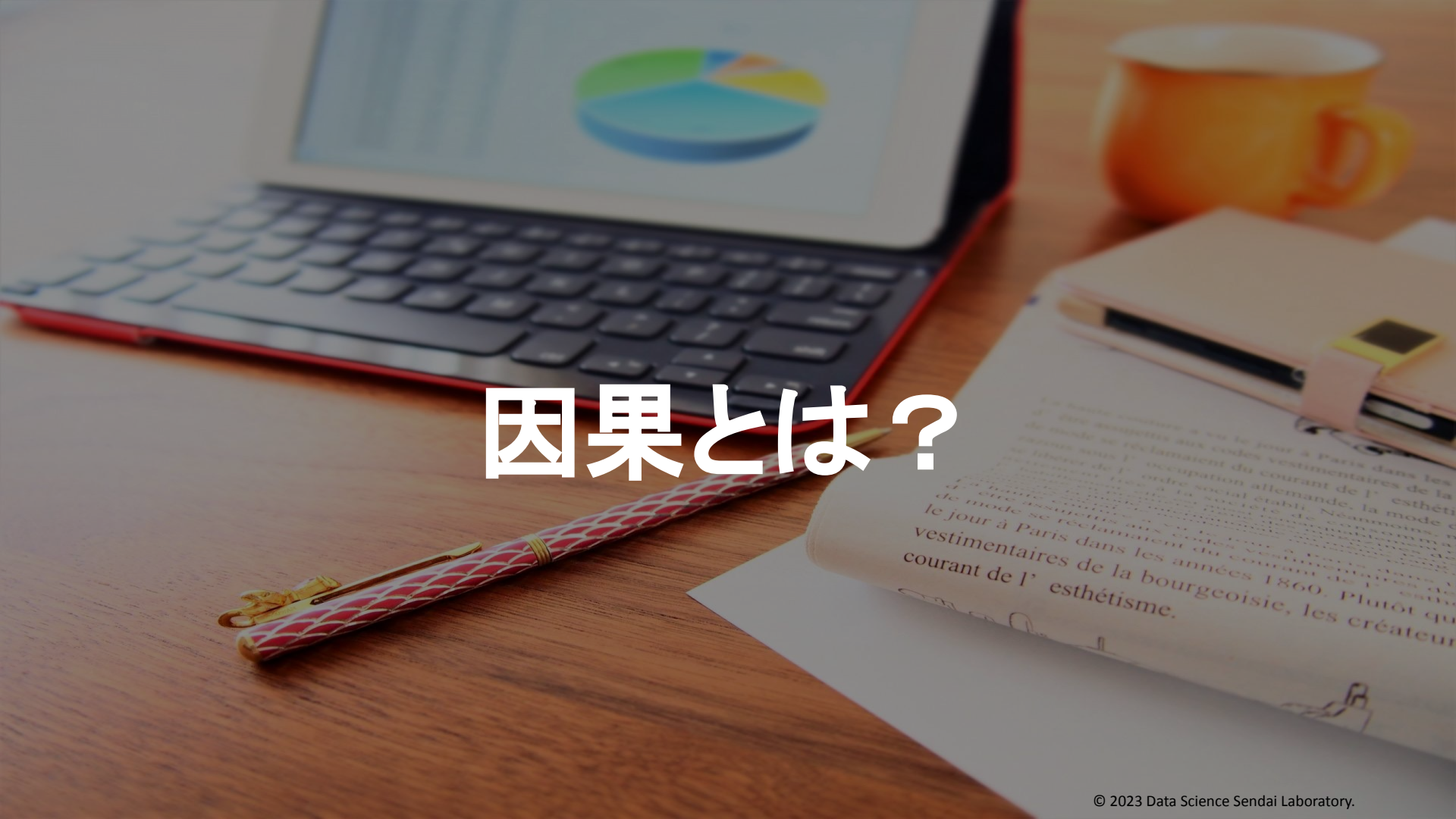
勉強会を通じて、
データサイエンスに触れる機会を創出

© 2023 Data Science Sendai Laboratory.

本勉強会について

- 勉強会開催時は、参加者のプライバシーを保護する観点より、原則ボイスとカメラはOFFとなります
- すべてのエッセンスを盛り込むことは難しく、“気づき”や“キーワード”など持って帰っていただくことを目的としています
- 質問や相談等、気軽にできる場を別途、オープンコミュニケーションツールであるSlackにて提供しています（こちらは基本、勉強会に参加された方のみに招待リンクを送付しています）
- 技術やスキルの習得を保証するものではなく、あくまでもご自身の興味関心や知識習得の入り口としてご活用ください

因果とデータサイエンス



因果とは？

© 2023 Data Science Sendai Laboratory.

© 2023 Data Science Sendai Laboratory.

因子・要因

因子・要因

因子・要因

因子・要因

因子・要因

原因

© 2023 Data Science Sendai Laboratory.

因子・要因

因子・要因

因子・要因

因子・要因

因子・要因

原因

結果

© 2023 Data Science Sendai Laboratory.

因子・要因

因子・要因

因子・要因

原因(因子・要因のかたまり)

(引き起こされた事象) 結果

結果

© 2023 Data Science Sendai Laboratory.

因子・要因

因子・要因

因子・要因

一般的には…

原因 → 結果

結果

© 2023 Data Science Sendai Laboratory.

因子・要因

因子・要因

因子・要因

データサイエンス分野では…

結果 → 原因 (因子・要因)

結果

© 2023 Data Science Sendai Laboratory.

データ
(結果)

© 2023 Data Science Sendai Laboratory.

データ
(結果)

統計解析

© 2023 Data Science Sendai Laboratory.

データ
(結果)

統計解析

因子・要因

因子・要因

原因

因子・要因

因子・要因

© 2023 Data Science Sendai Laboratory.

データ
(結果)

データサイエンスは
データを扱うことが目的ではない

因子・要因

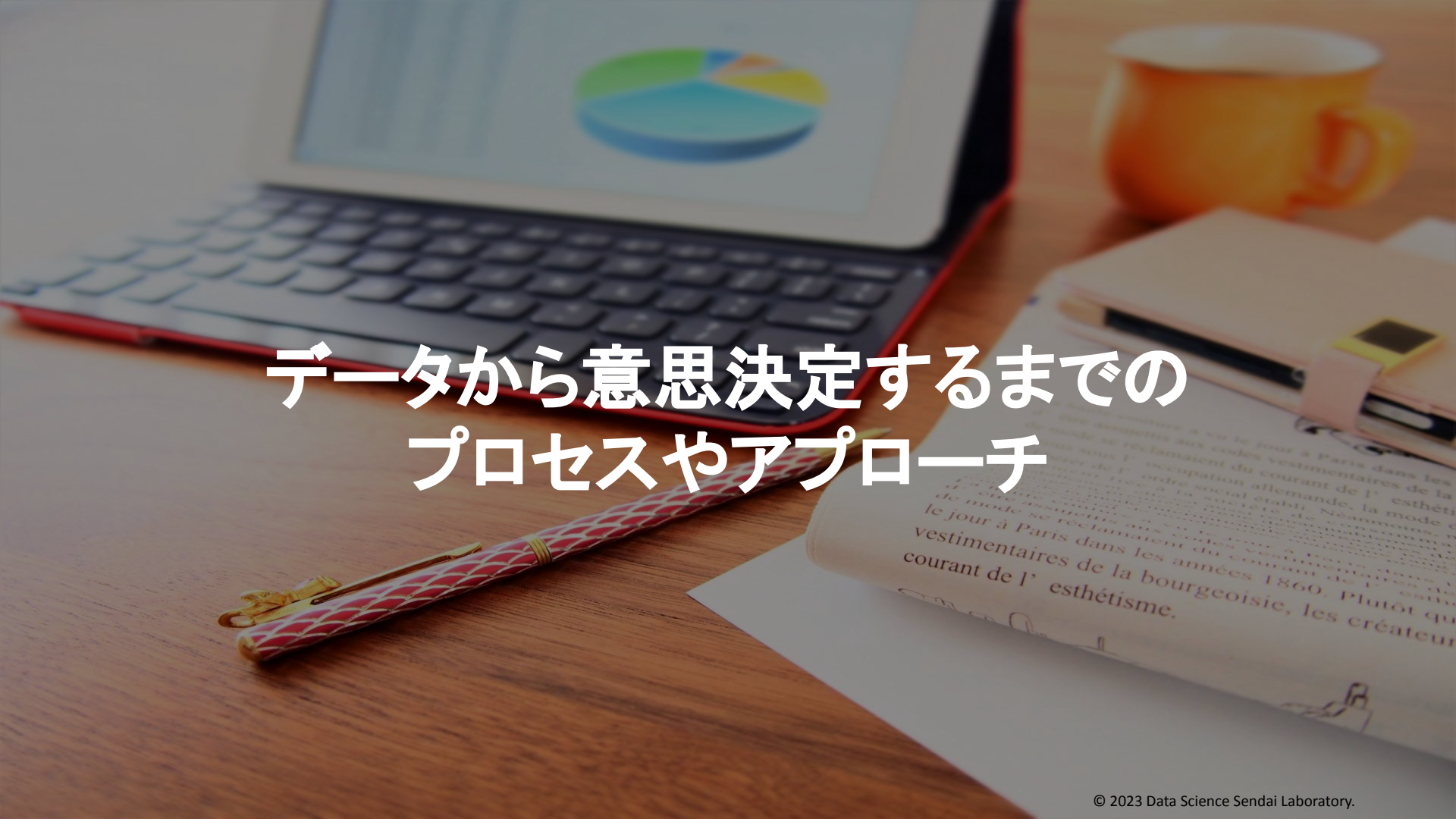
因子・要因

因子・要因

原因

因子・要因

© 2023 Data Science Sendai Laboratory.



データから意思決定するまでの
プロセスやアプローチ

© 2023 Data Science Sendai Laboratory.

1.
収集

© 2023 Data Science Sendai Laboratory.

© 2023 Data Science Sendai Laboratory.

© 2023 Data Science Sendai Laboratory.

© 2023 Data Science Sendai Laboratory.

© 2023 Data Science Sendai Laboratory.

© 2023 Data Science Sendai Laboratory.

1.
収集

2.
整形

3.
集計

5.
統計解析

4.
可視化

6.
機械学習

7.
考察

© 2023 Data Science Sendai Laboratory.

1.
収集

2.
整形

3.
集計

5.
統計解析

4.
可視化

6.
機械学習

7.
考察

8.
意思決定

© 2023 Data Science Sendai Laboratory.

プロセス

1. 収集
2. 整形
3. 集計
4. 可視化
5. 統計解析
6. 機械学習
7. 考察
8. 意思決定

© 2023 Data Science Sendai Laboratory.

データエンジニア

1. 収集
2. 整形
3. 集計
4. 可視化
5. 統計解析
6. 機械学習
7. 考察
8. 意思決定

MySQL SQLite hadoop NVIDIA DEEP LEARNING INSTITUTE

© 2023 Data Science Sendai Laboratory.

1.
収集

2.
整形

3.
集計

データアナリスト

pandas

5.
統計解析

4.
可視化

6.
機械学習

Keras

scikit learn

7.
考察

8.
意思決定

機械学習エンジニア
AIエンジニア

PyTorch

© 2023 Data Science Sendai Laboratory.

1.
収集

2.
整形

3.
集計

5.
統計解析

4.
可視化

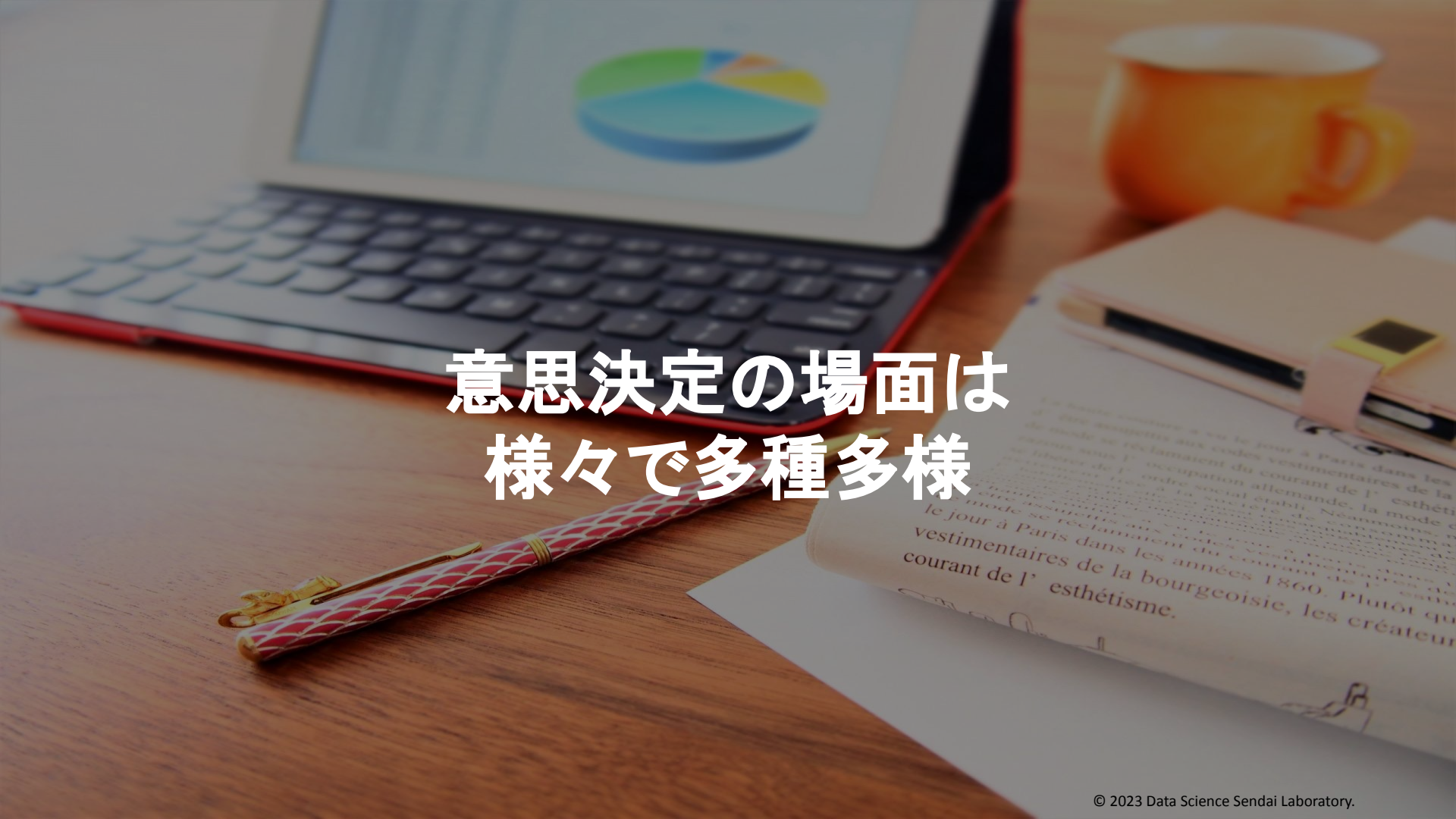
6.
機械学習

意思決定者
(経営者、現場のヒト)

7.
考察

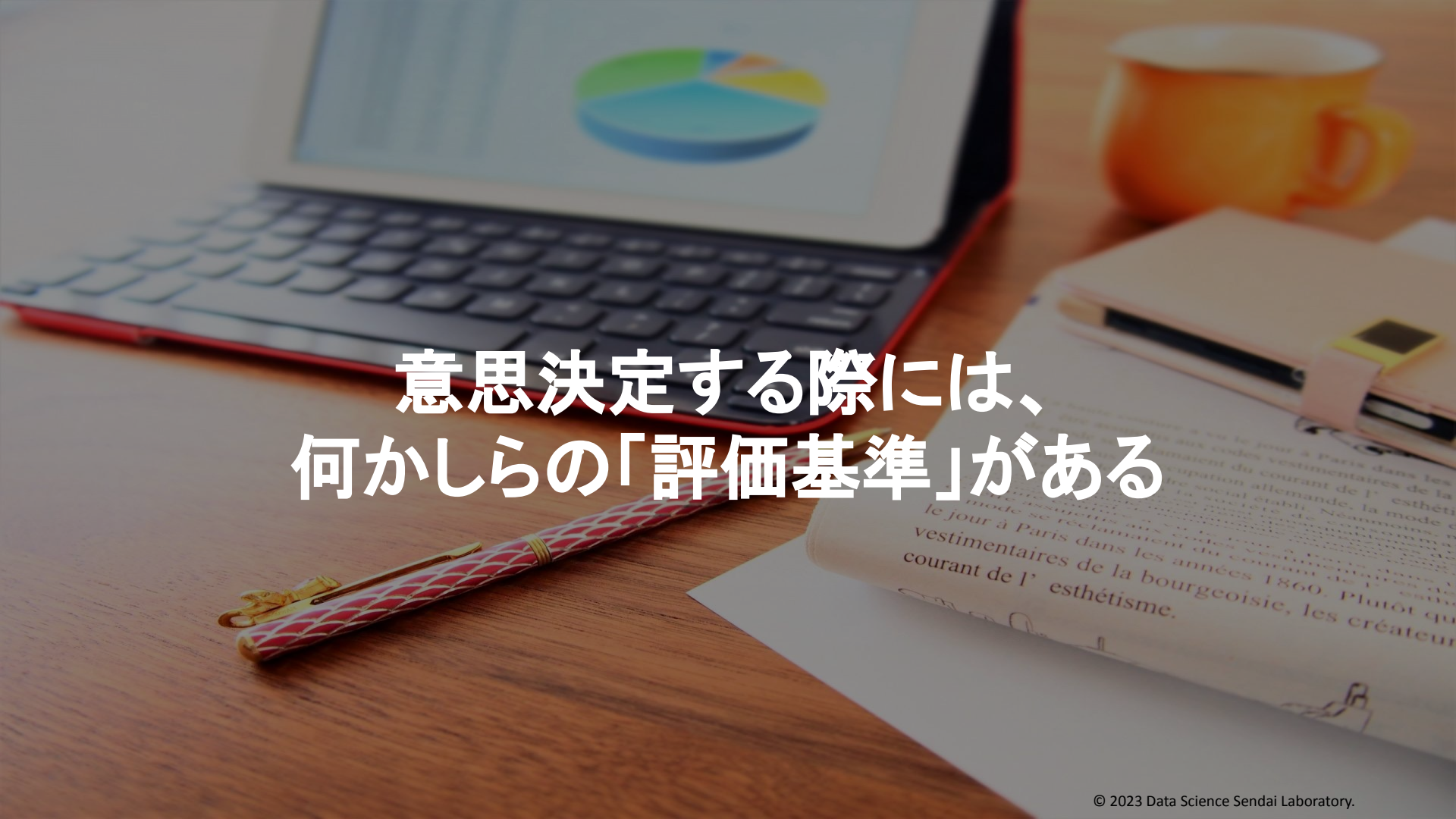
8.
意思決定

© 2023 Data Science Sendai Laboratory.



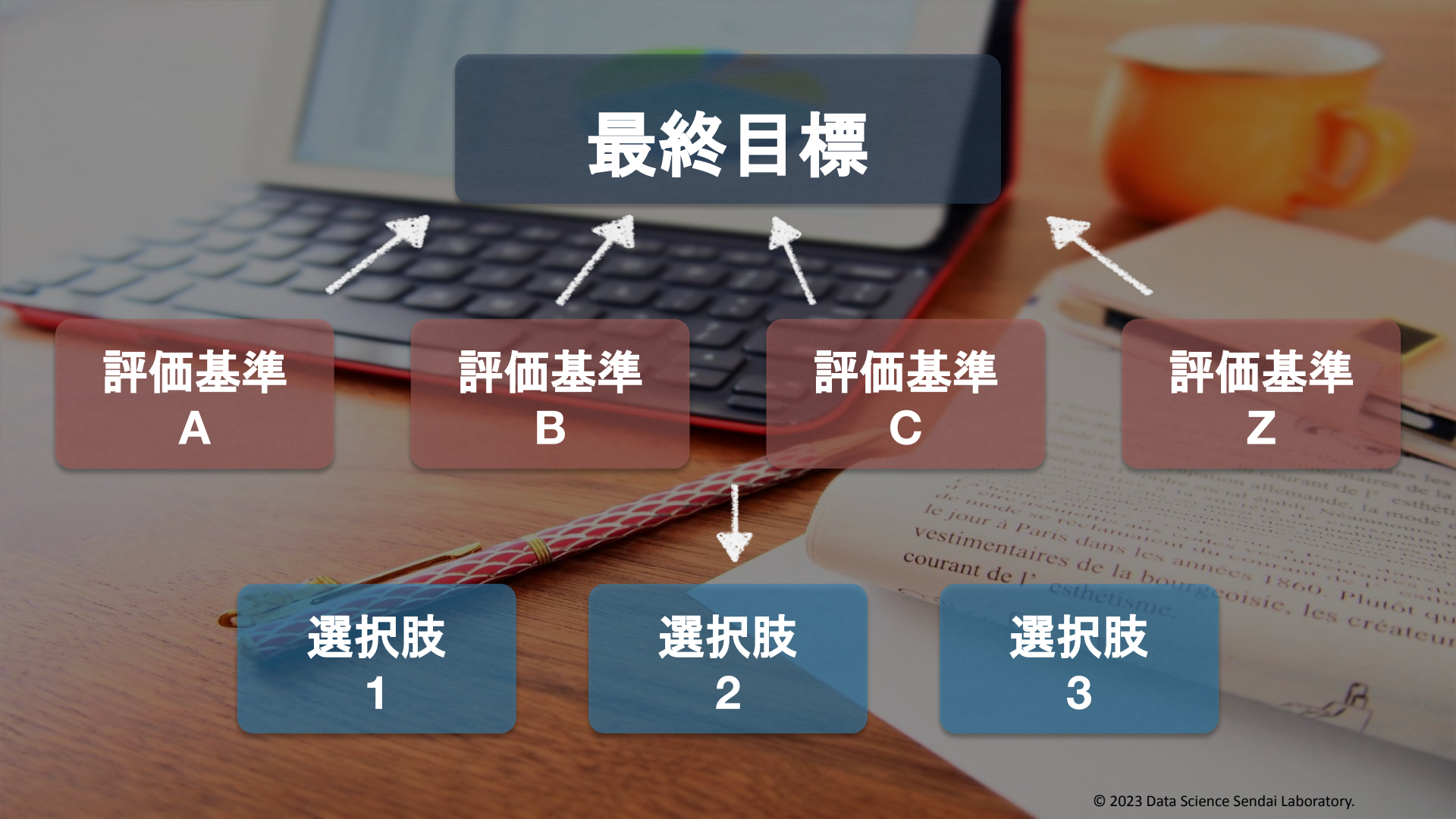
意思決定の場面は
様々で多種多様

© 2023 Data Science Sendai Laboratory.



意思決定する際には、
何かしらの「評価基準」がある

© 2023 Data Science Sendai Laboratory.



© 2023 Data Science Sendai Laboratory.

最終目標

データサイエンス
目的とアプローチが重要

選択肢
1

選択肢
2

選択肢
3

© 2023 Data Science Sendai Laboratory.

AI
(人工知能)

AI
(人工知能)

Machine Learning
（機械学習）

AI
(人工知能)

Machine Learning
（機械学習）

Supervised Learning
（教師あり学習）

AI
(人工知能)

Machine Learning
（機械学習）

Supervised Learning
（教師あり学習）

Non-Supervised Learning
（教師なし学習）

AI
(人工知能)

Machine Learning
（機械学習）

Supervised Learning
（教師あり学習）

Non-Supervised Learning
（教師なし学習）

Semi-Supervised Learning
（半教師あり学習）

AI
(人工知能)

Machine Learning
（機械学習）

Supervised Learning
（教師あり学習）

Non-Supervised Learning
（教師なし学習）

Semi-Supervised Learning
（半教師あり学習）

Reinforcement Learning
（強化学習）

AI
(人工知能)

Machine Learning
（機械学習）

Supervised Learning
（教師あり学習）

Non-Supervised Learning
（教師なし学習）

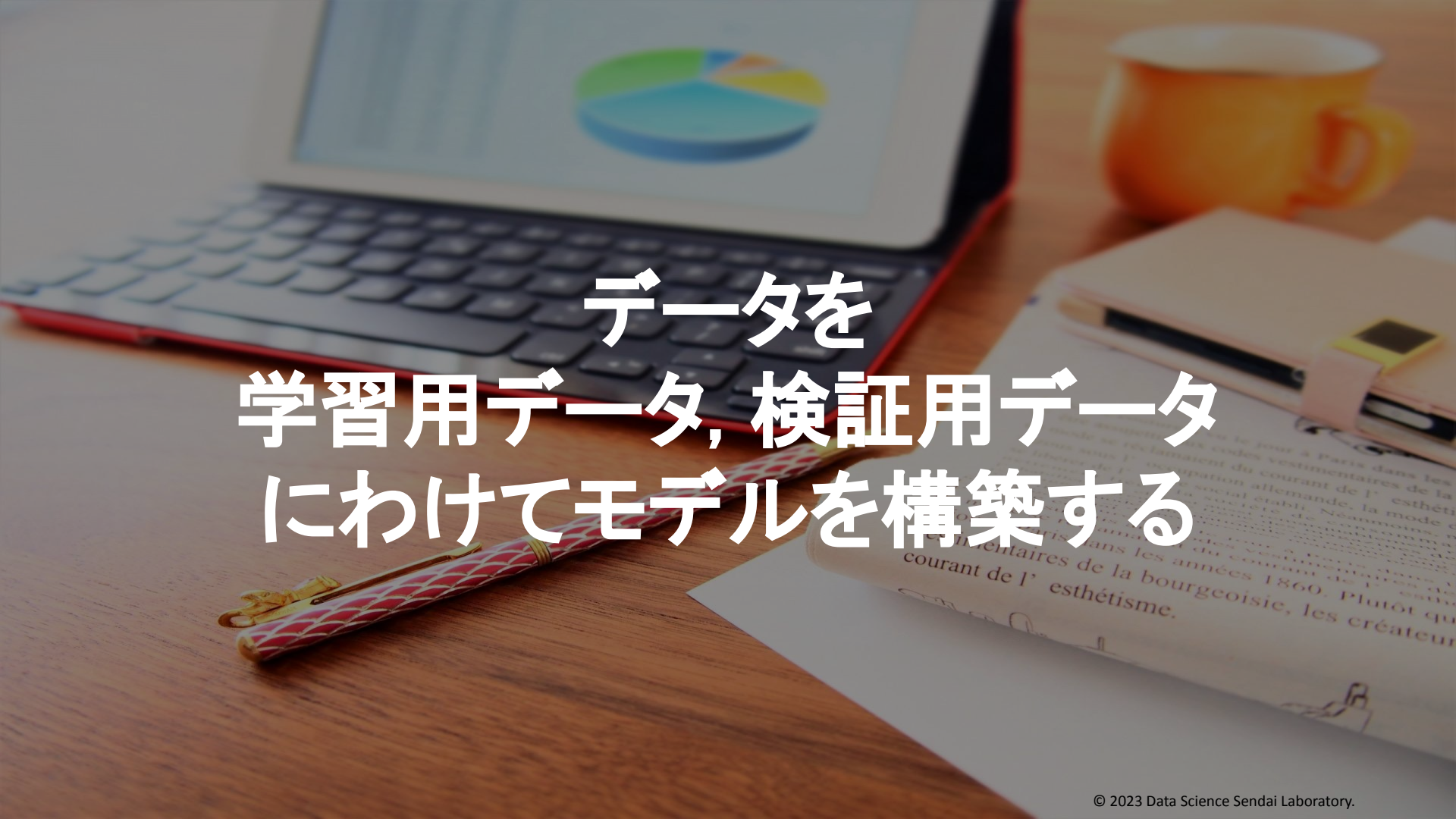
Semi-Supervised Learning
（半教師あり学習）

Reinforcement Learning
（強化学習）

Deep Learning
（深層学習）

教師あり学習

© 2023 Data Science Sendai Laboratory.



データを
学習用データ, 検証用データ
にわけてモデルを構築する

© 2023 Data Science Sendai Laboratory.

学習用データ

検証用データ

© 2023 Data Science Sendai Laboratory.

学習用データ

説明変数
これら値が、尤もらしさ
を見出すもととなる

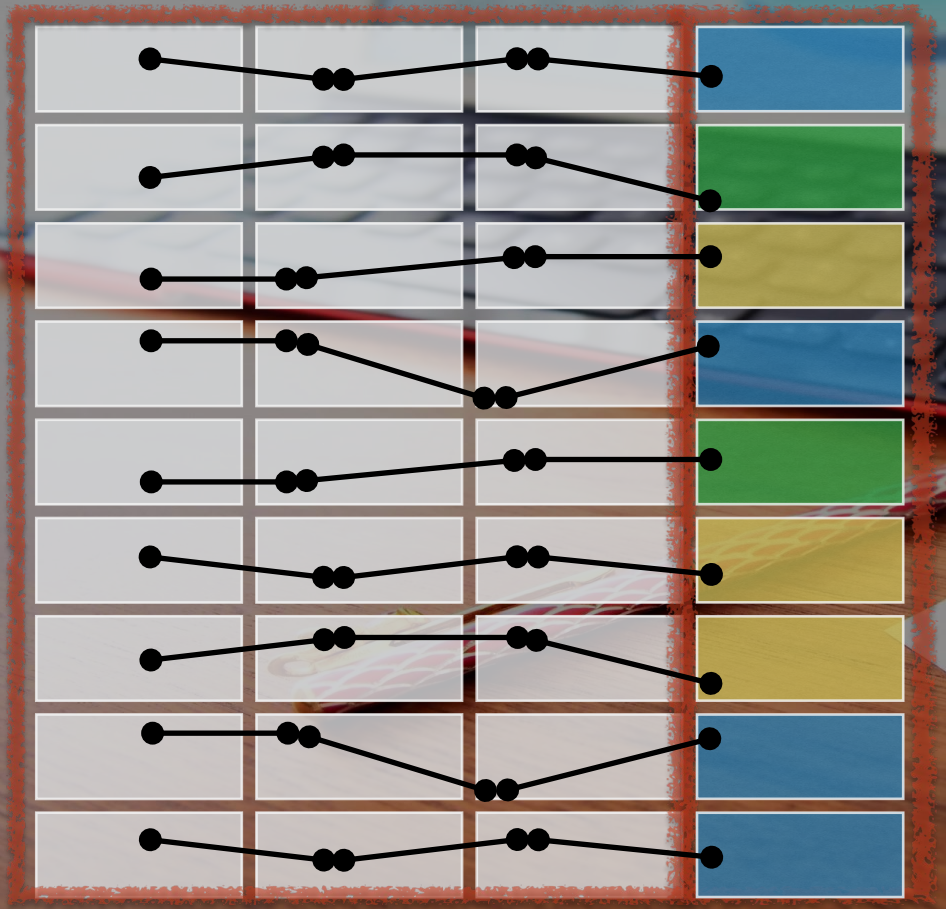
© 2023 Data Science Sendai Laboratory.

学習用データ

目的変数
いわゆる、正解ラベル
というもの。

© 2023 Data Science Sendai Laboratory.

学習用データ



パターン(説明変数)と
正確ラベル(目的変数)
をもとにモデル構築

© 2023 Data Science Sendai Laboratory.

学習用データ

学習した
モデル

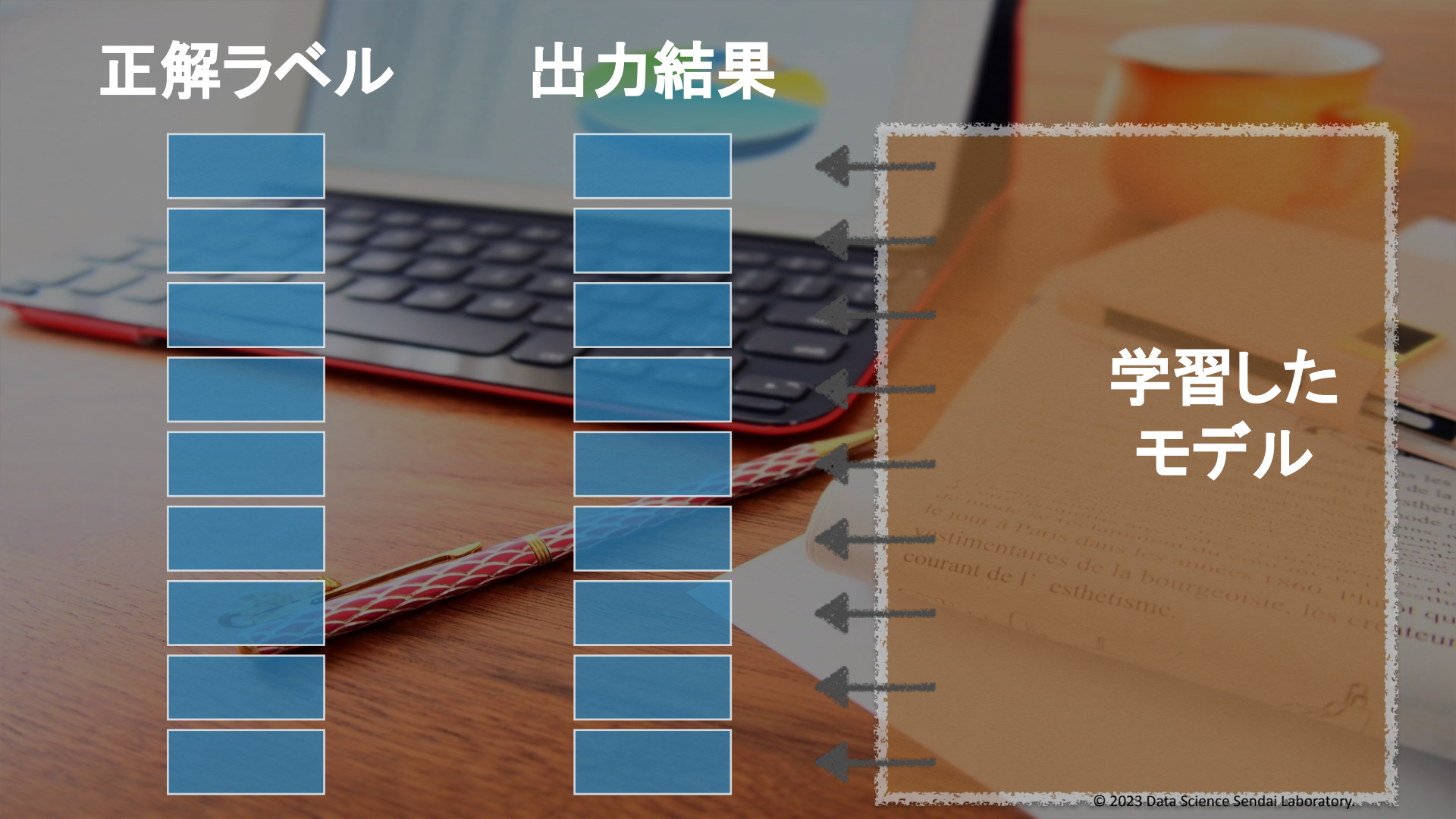
© 2023 Data Science Sendai Laboratory.

学習用データ

検証用データ

学習した
モデル

© 2023 Data Science Sendai Laboratory.



© 2023 Data Science Sendai Laboratory.

正解ラベル　　出力結果

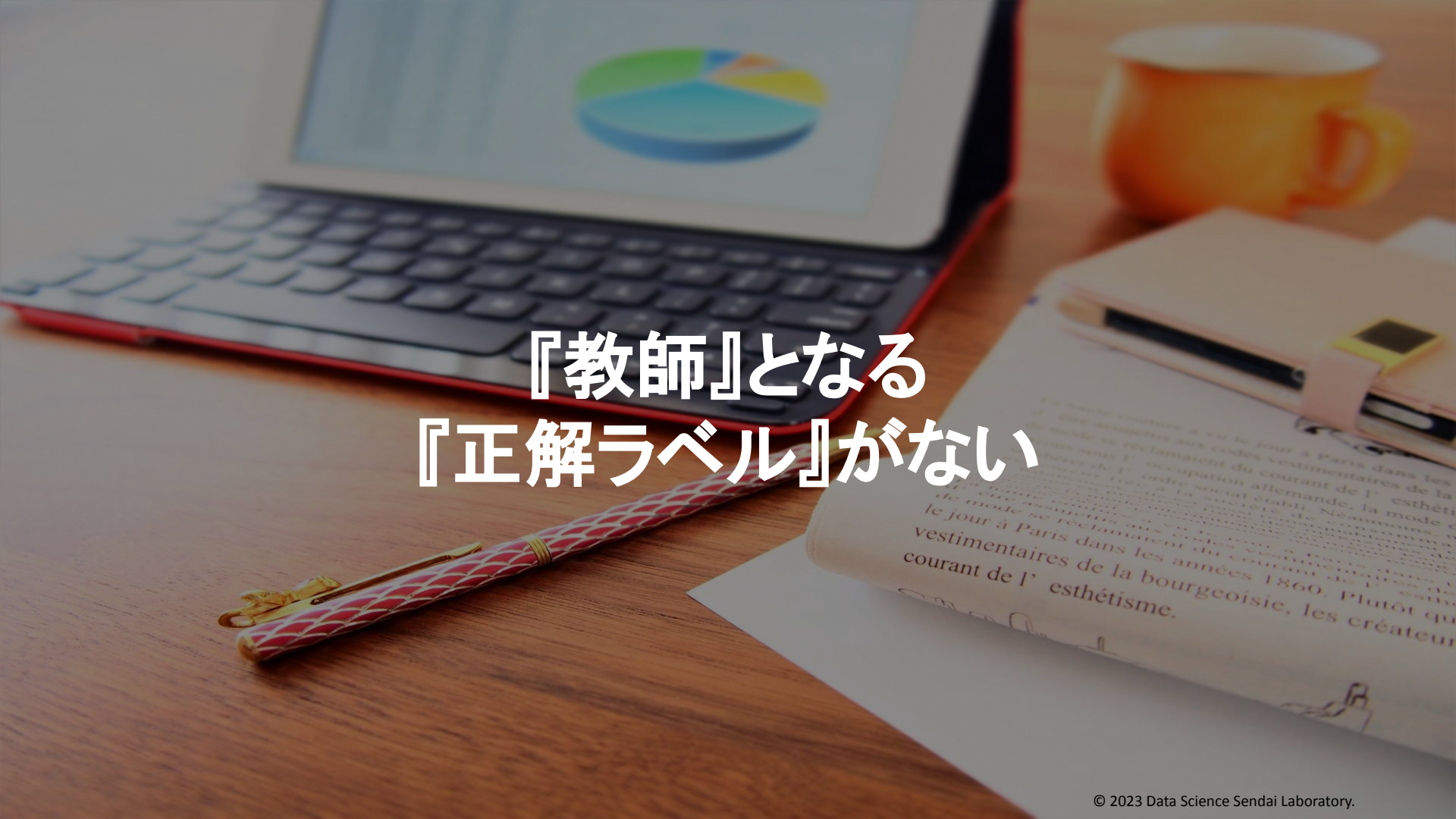
正解ラベルに対して
出力結果がどれくらい
正解しているか？
↓
正解率
↓
モデルの
予測精度

学習した
モデル

© 2023 Data Science Sendai Laboratory.

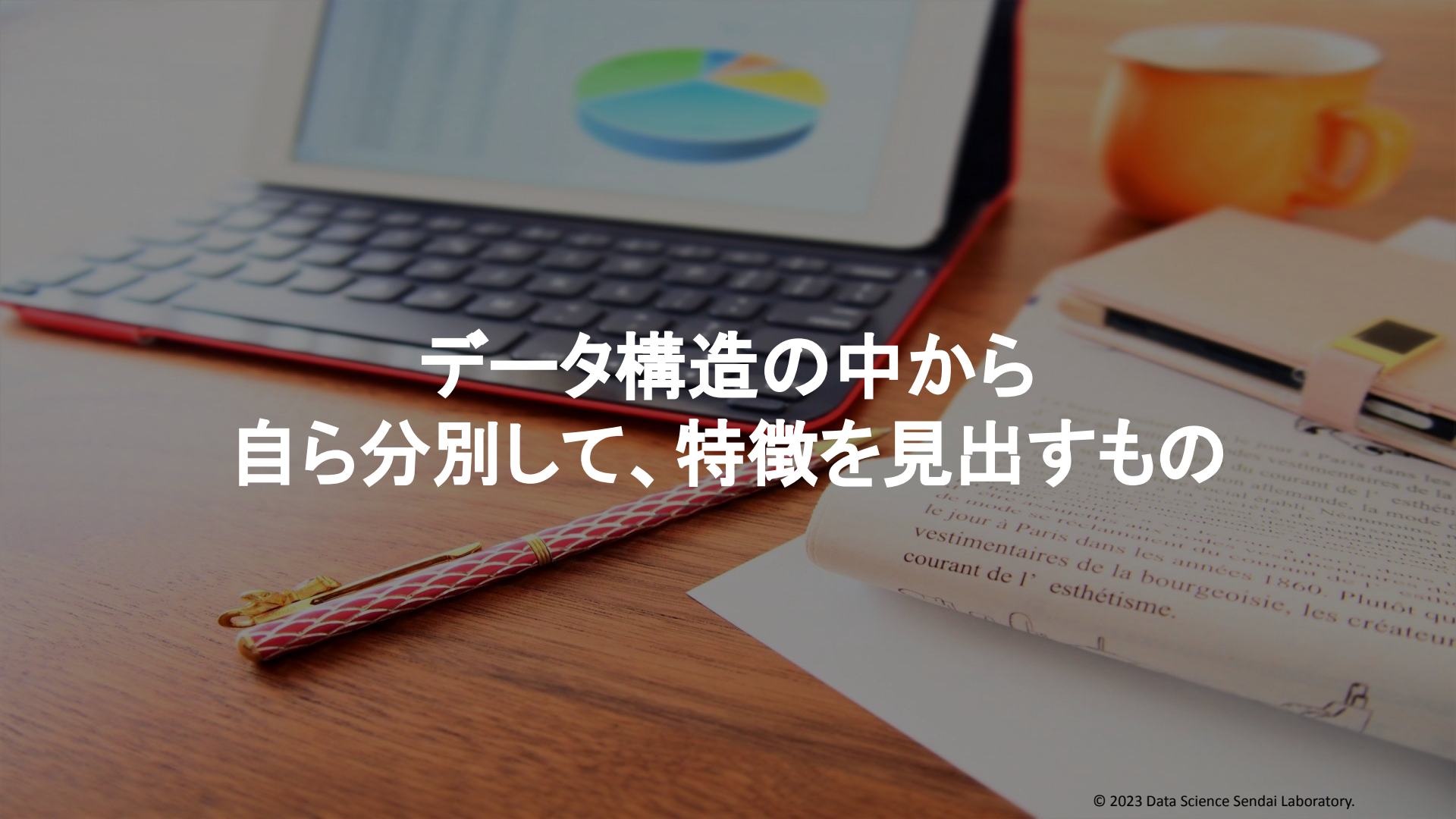
教師なし学習

© 2023 Data Science Sendai Laboratory.



『教師』となる
『正解ラベル』がない

© 2023 Data Science Sendai Laboratory.



データ構造の中から
自ら分別して、特徴を見出すもの

© 2023 Data Science Sendai Laboratory.

データ
かたまり

教師なし
学習モデル

主成分分析
オートエンコーダー
クラスタリング
自己組織化マップ
敵対的生成ネットワーク
DBSCAN

A

B

C

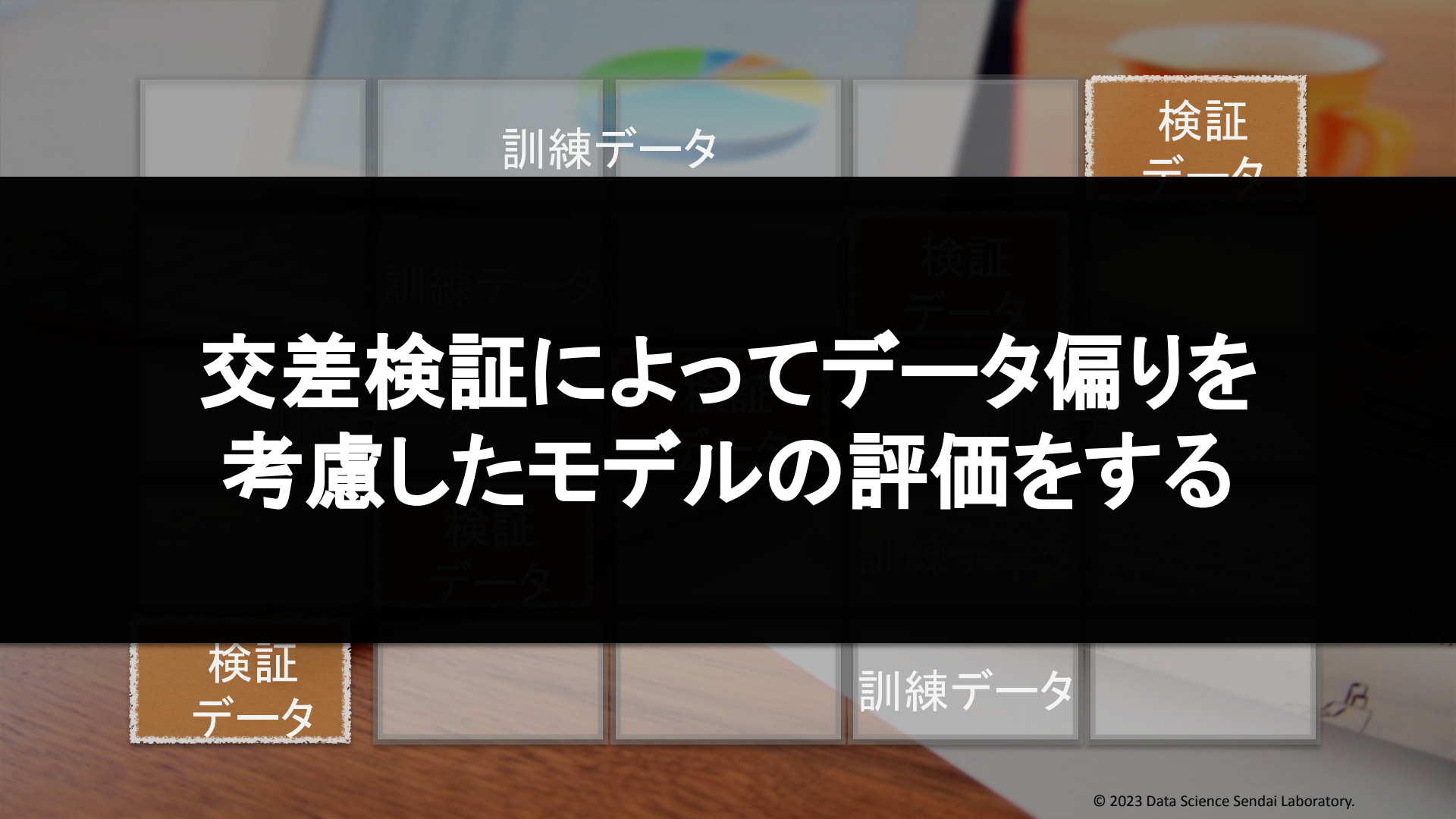
© 2023 Data Science Sendai Laboratory.

交差検証

© 2023 Data Science Sendai Laboratory.

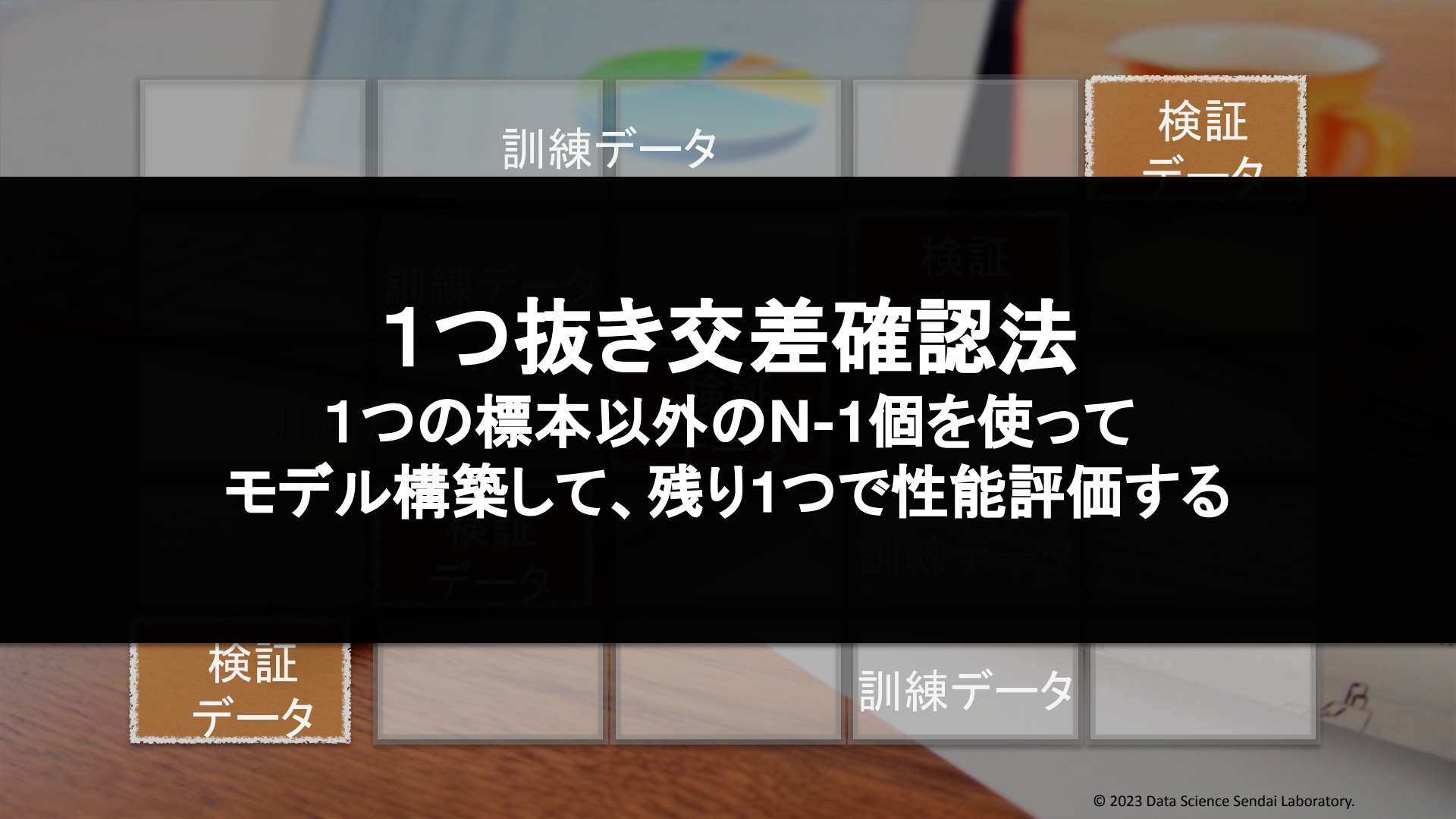
© 2023 Data Science Sendai Laboratory.

© 2023 Data Science Sendai Laboratory.



交差検証によってデータ偏りを考慮したモデルの評価をする

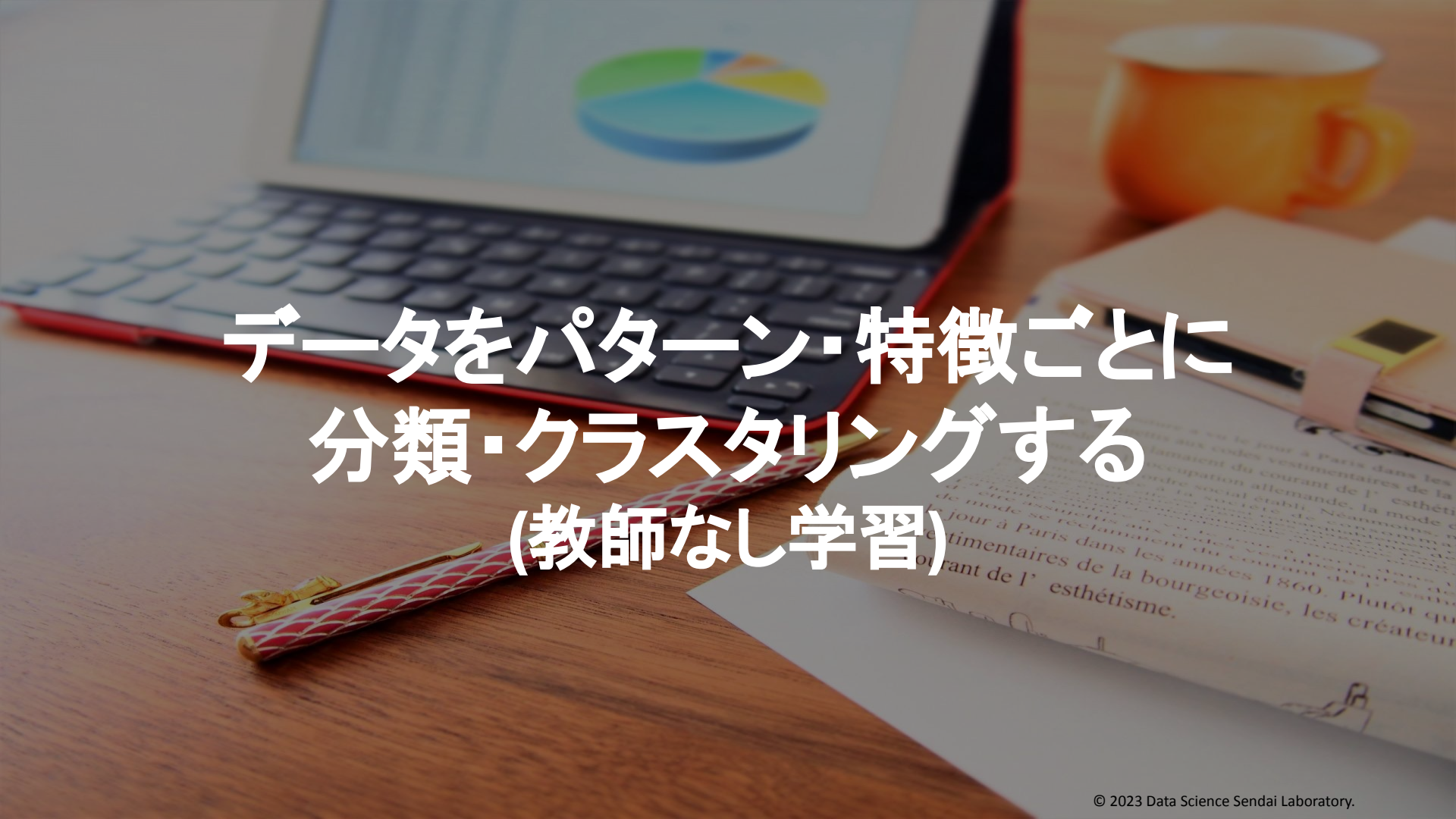
© 2023 Data Science Sendai Laboratory.



1つ抜き交差確認法

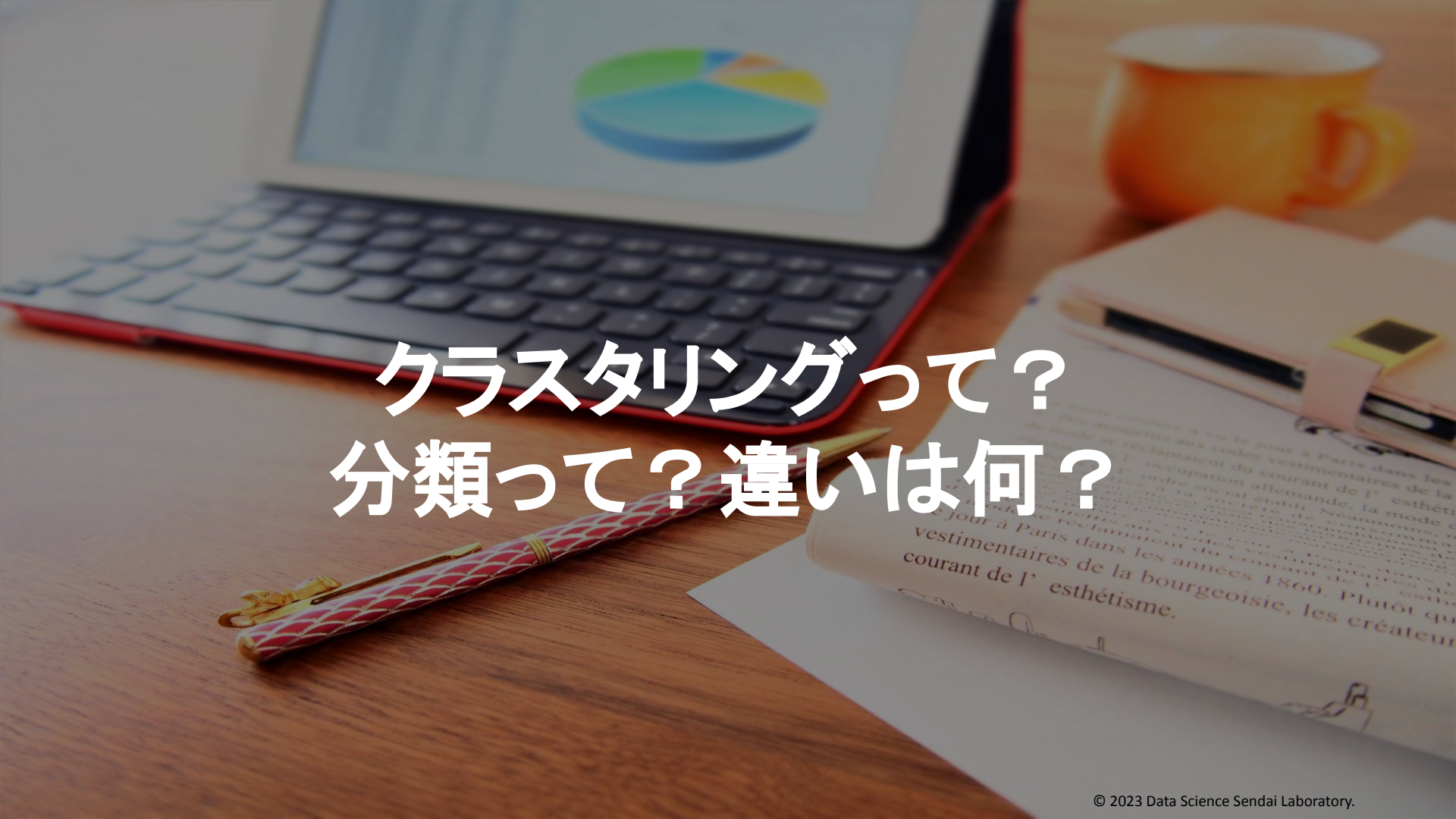
1つの標本以外のN-1個を使って
モデル構築して、残り1つで性能評価する

© 2023 Data Science Sendai Laboratory.



データをパターン・特徴ごとに分類・クラスタリングする
(教師なし学習)

© 2023 Data Science Sendai Laboratory.

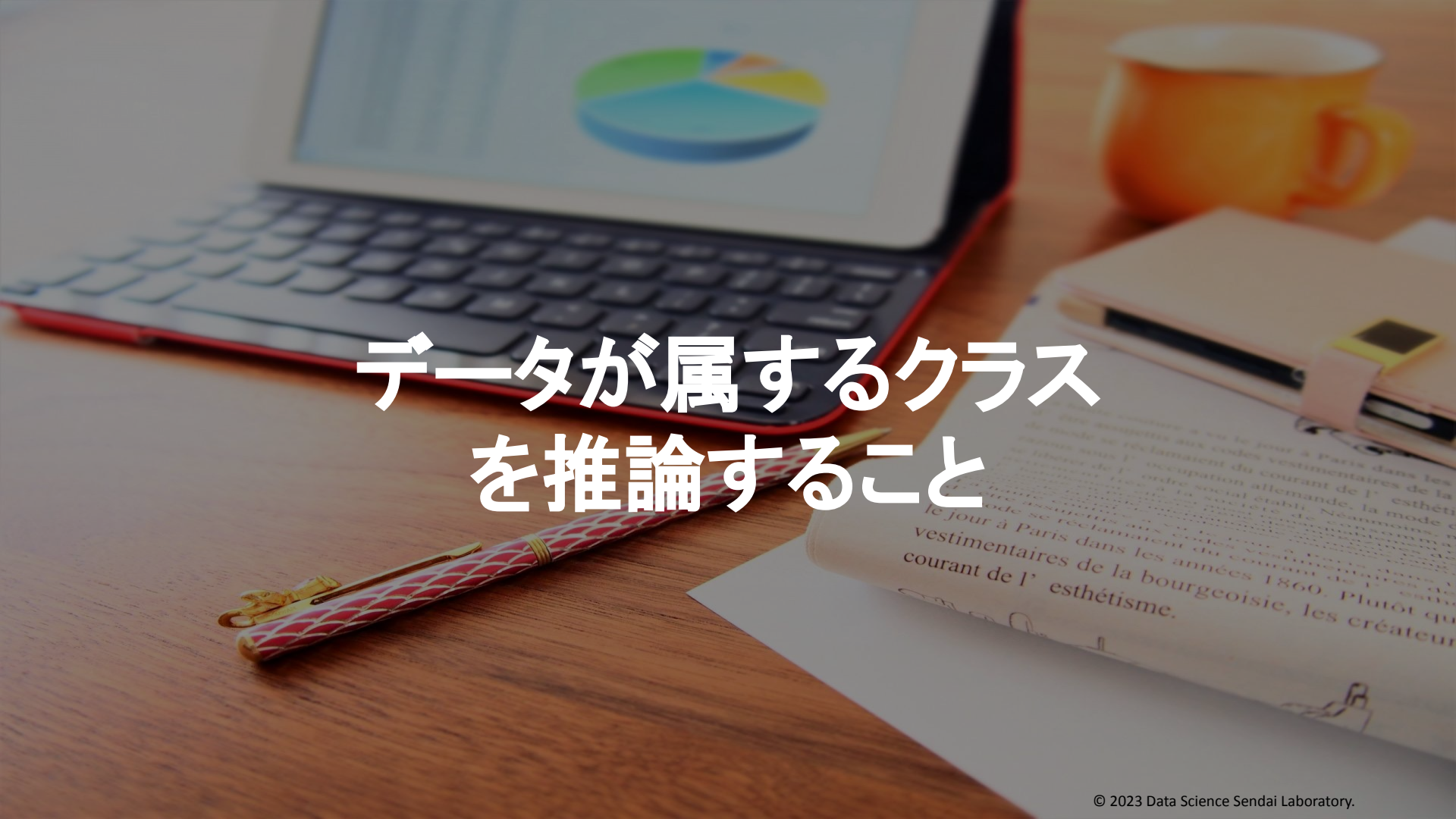


クラスタリングって？
分類って？違いは何？

© 2023 Data Science Sendai Laboratory.

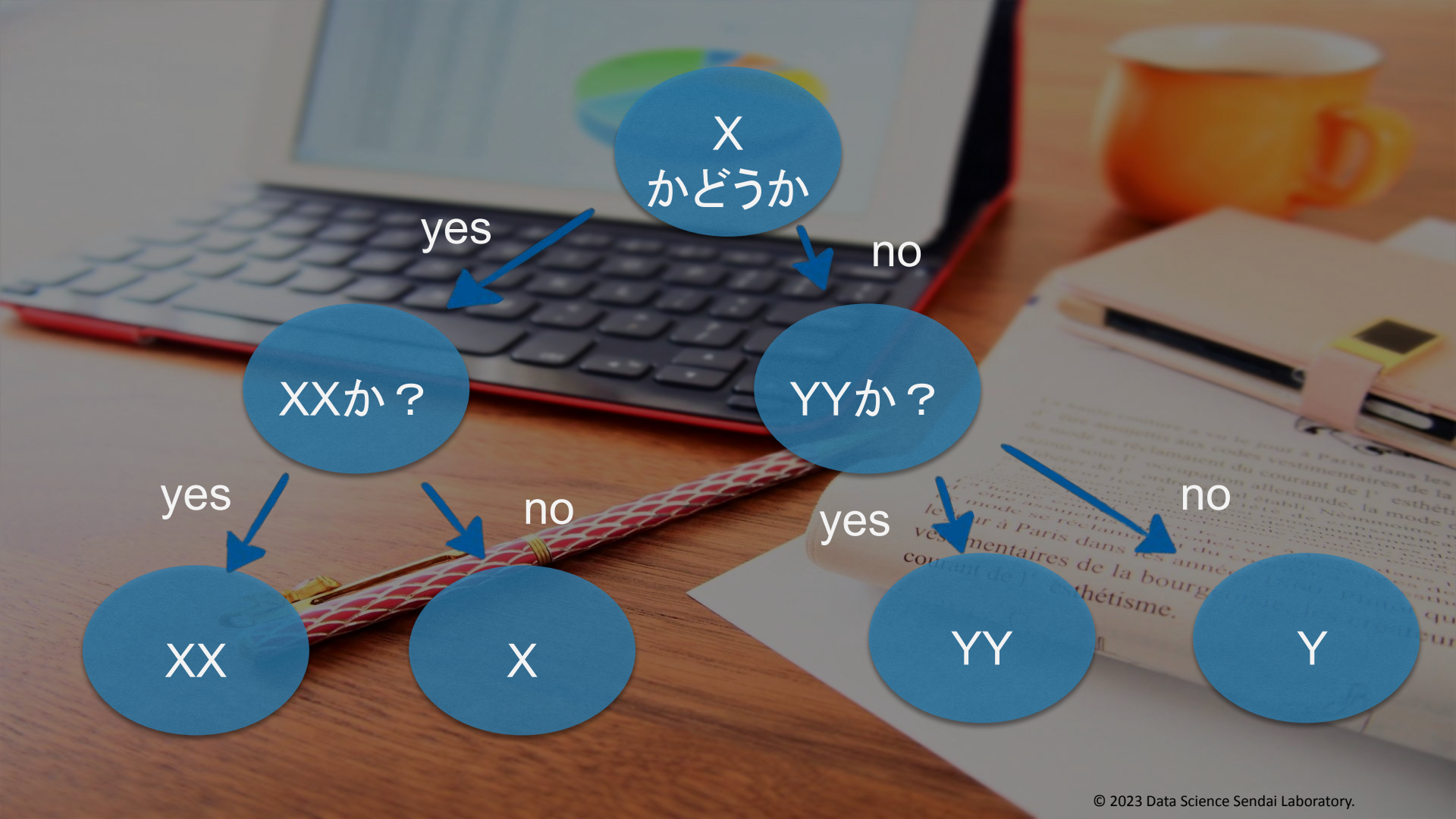
分類の主な目的

© 2023 Data Science Sendai Laboratory.

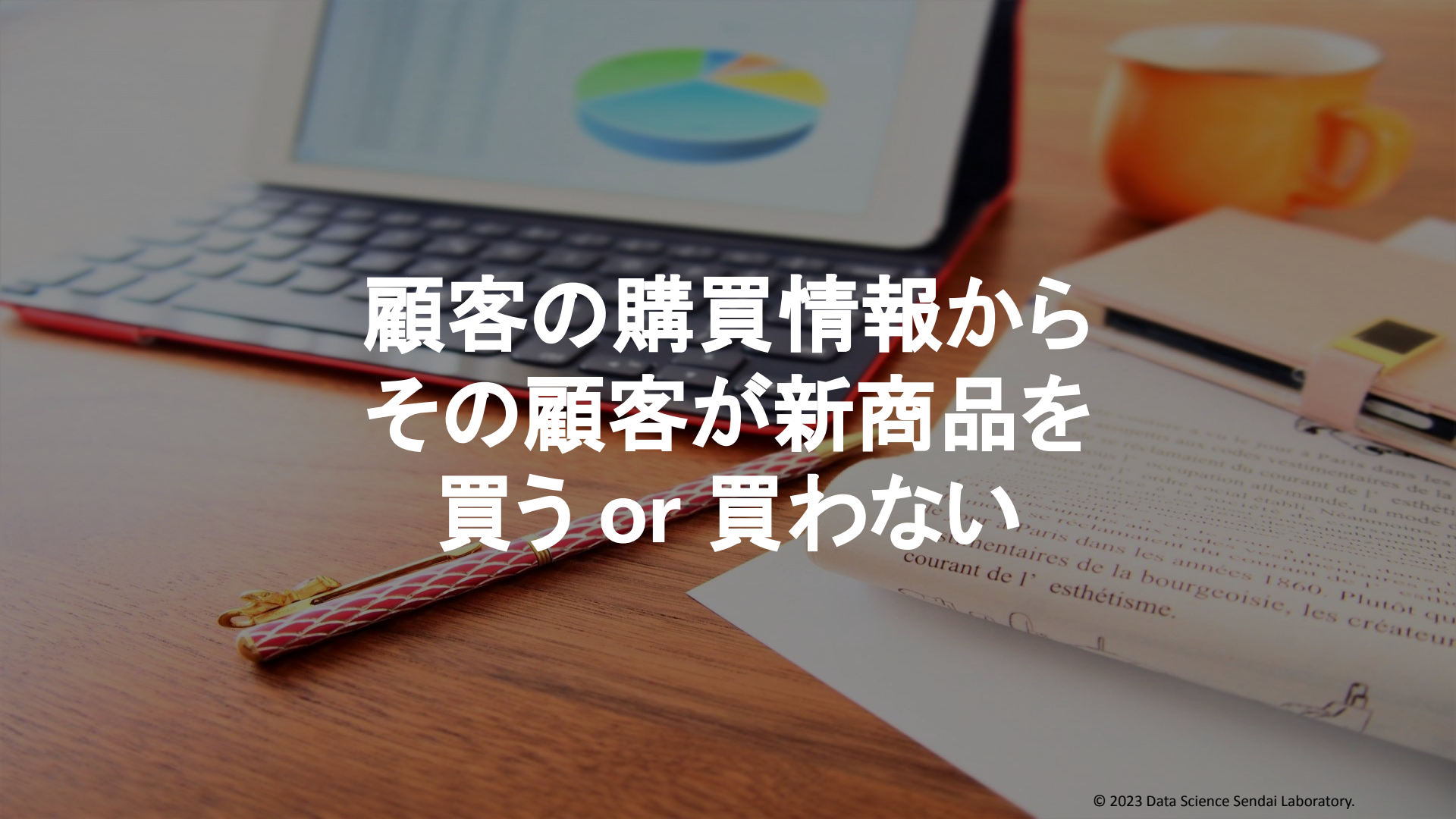


データが属するクラスを推論すること

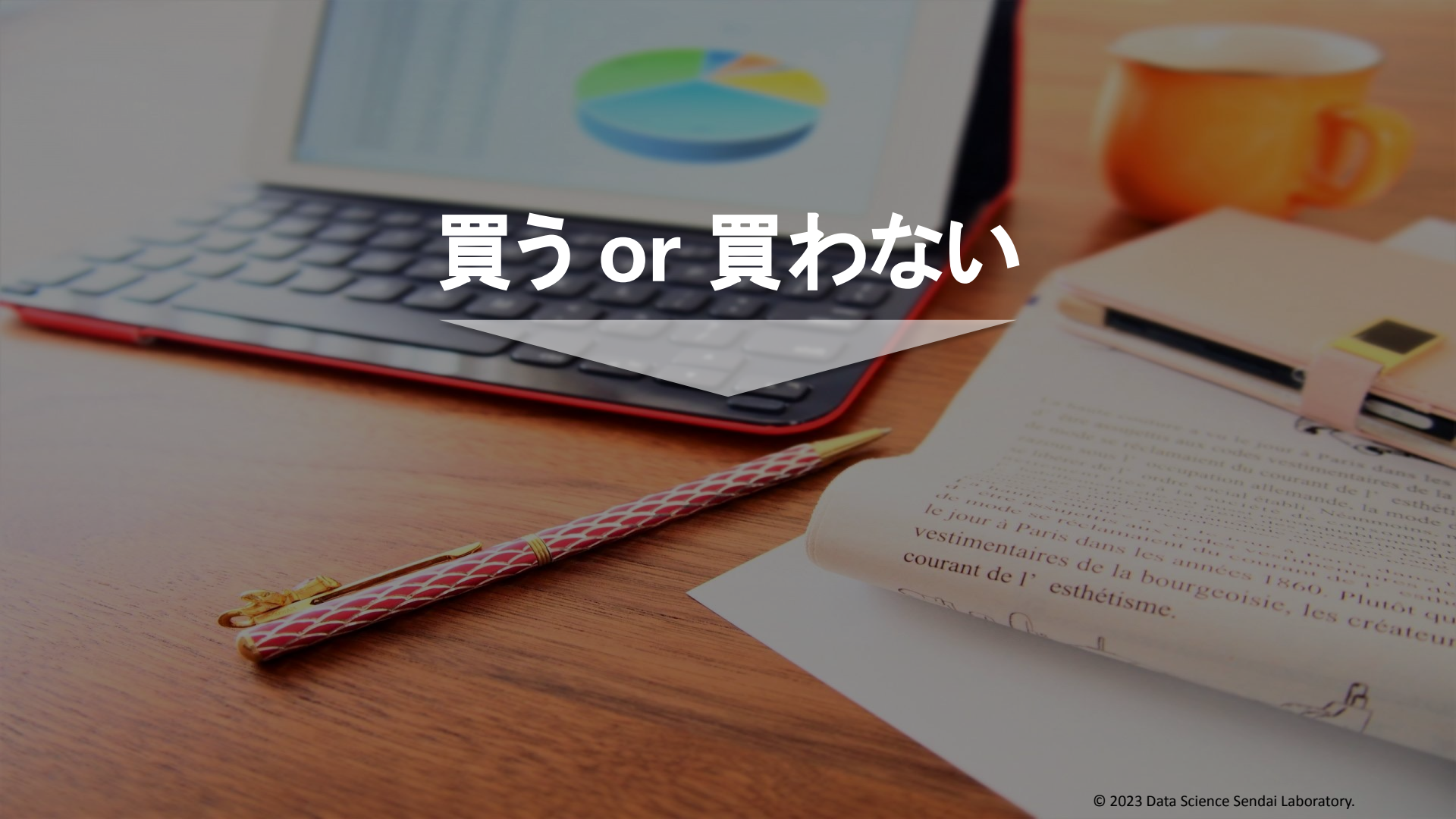
© 2023 Data Science Sendai Laboratory.



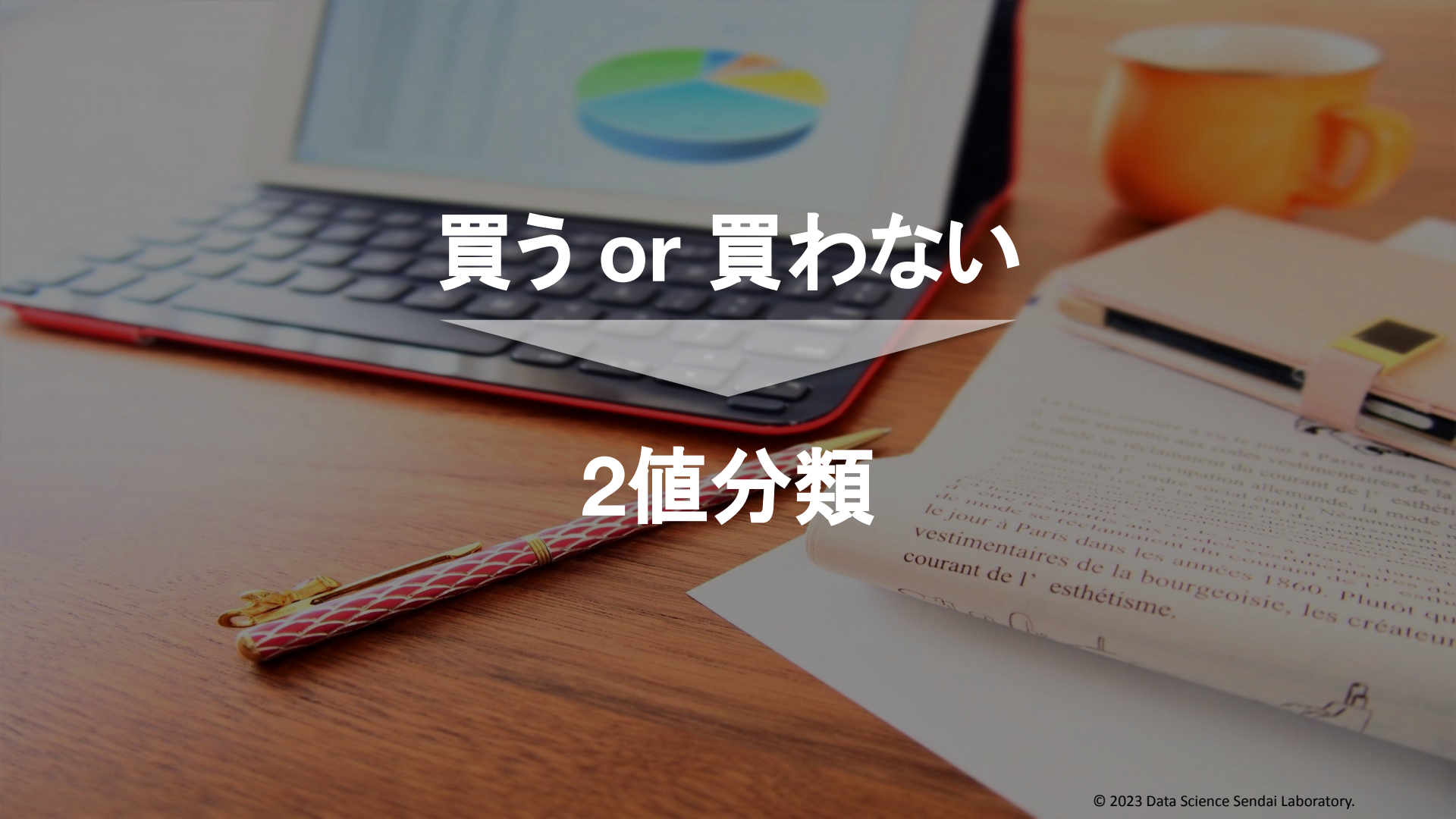
© 2023 Data Science Sendai Laboratory.



© 2023 Data Science Sendai Laboratory.



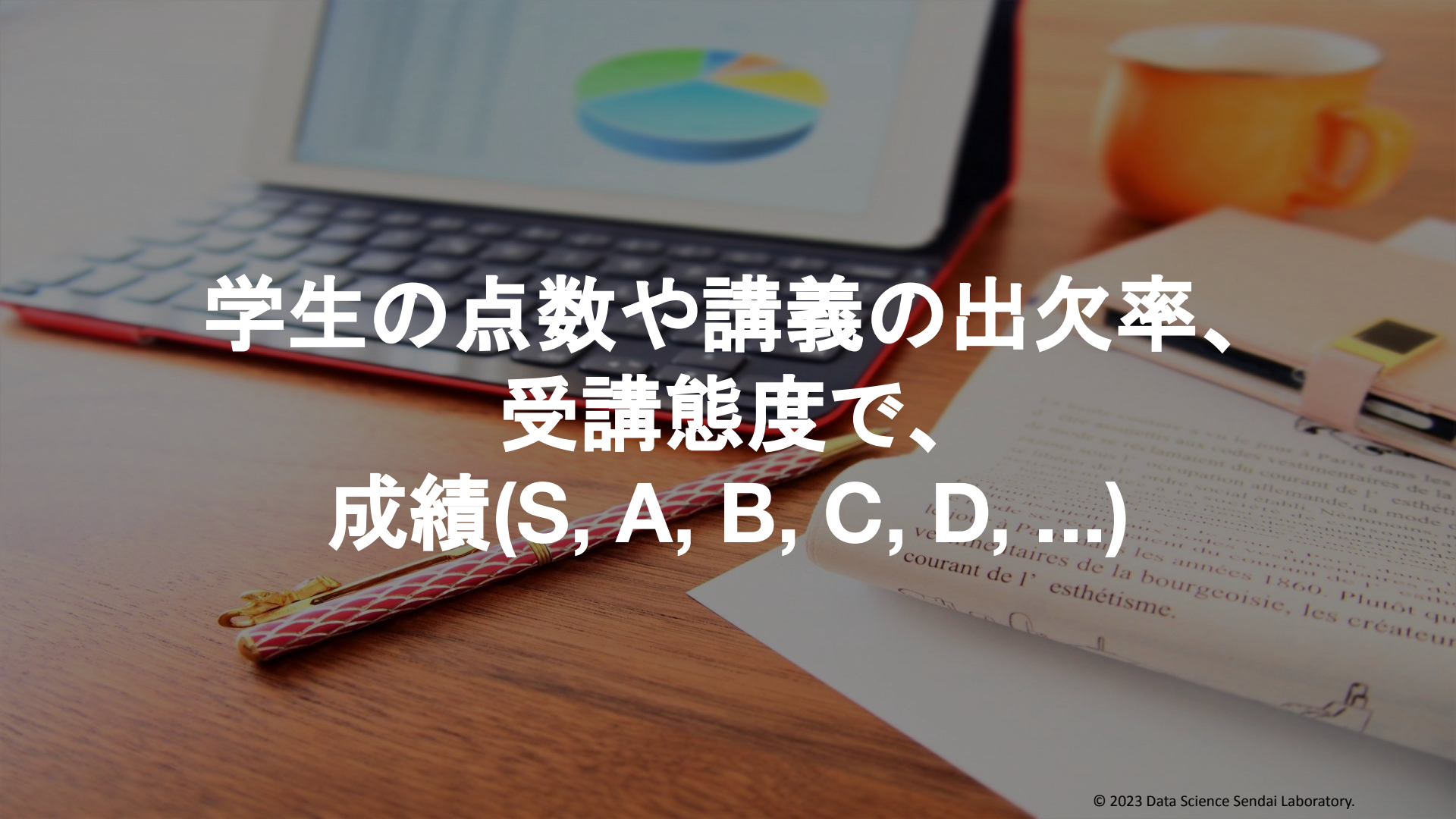
© 2023 Data Science Sendai Laboratory.



買う or 買わない

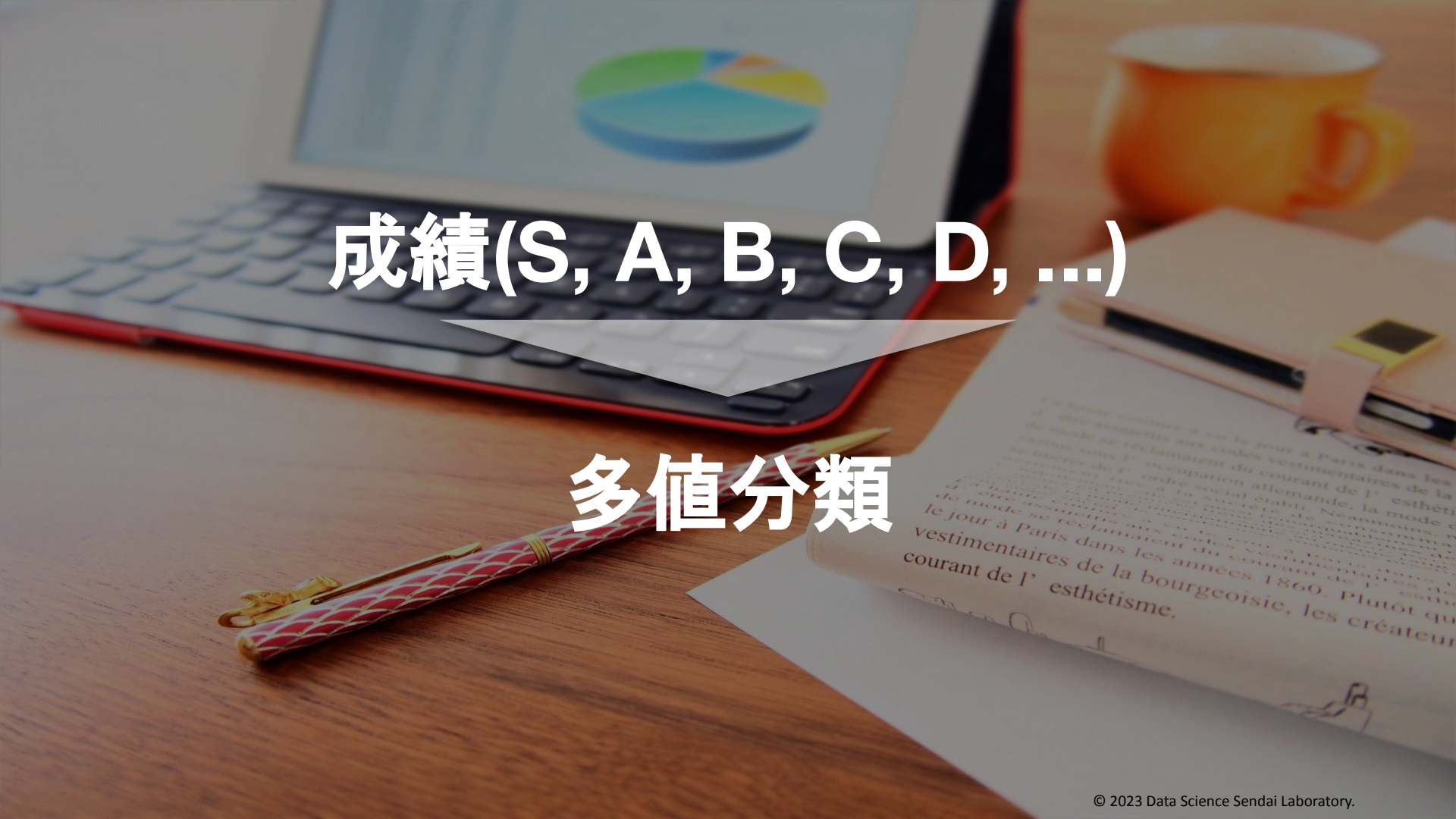
2値分類

© 2023 Data Science Sendai Laboratory.



学生の点数や講義の出欠率、受講態度で、成績(S, A, B, C, D, ...)

© 2023 Data Science Sendai Laboratory.



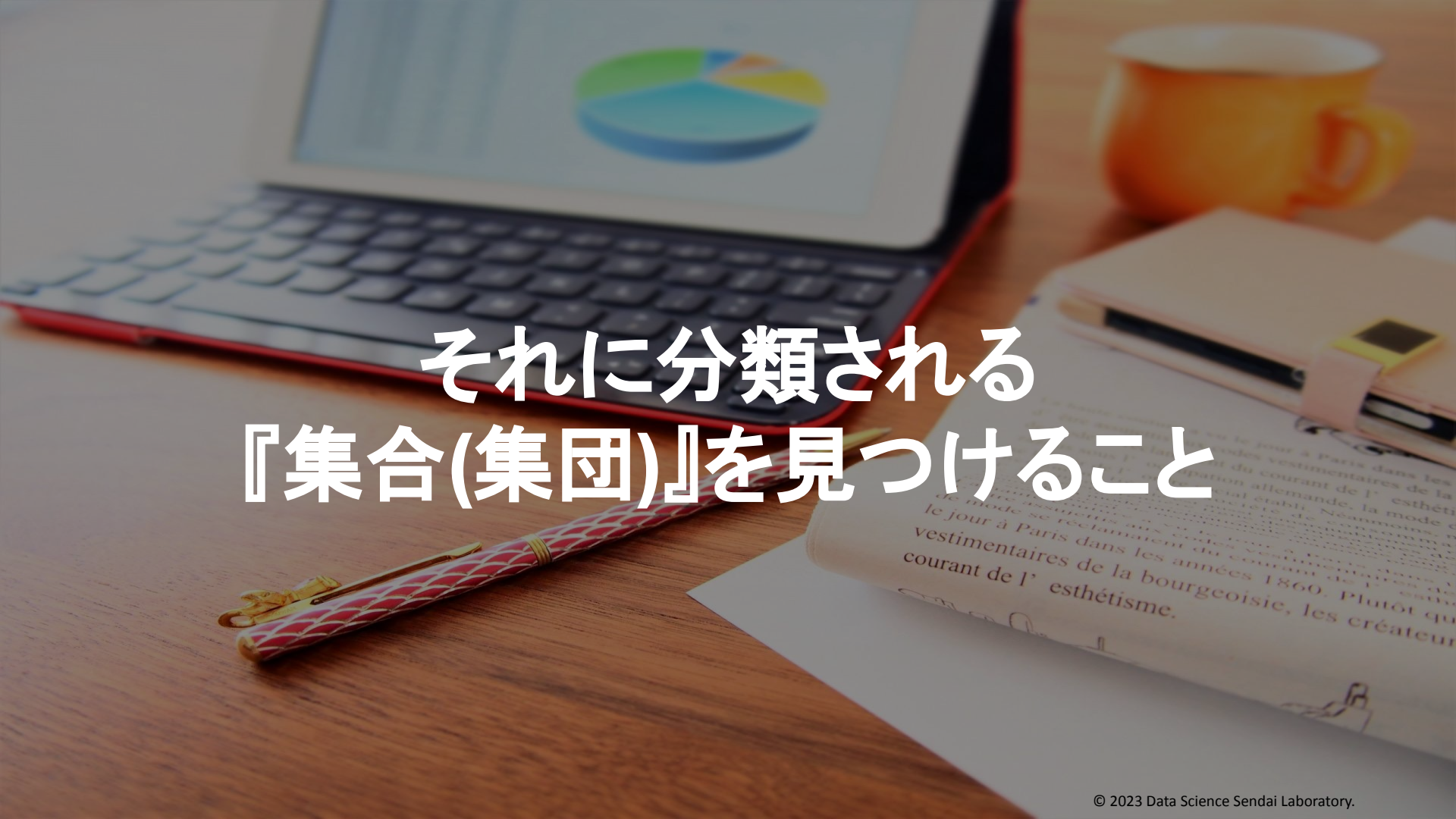
成績(S, A, B, C, D, ...)

多値分類

© 2023 Data Science Sendai Laboratory.

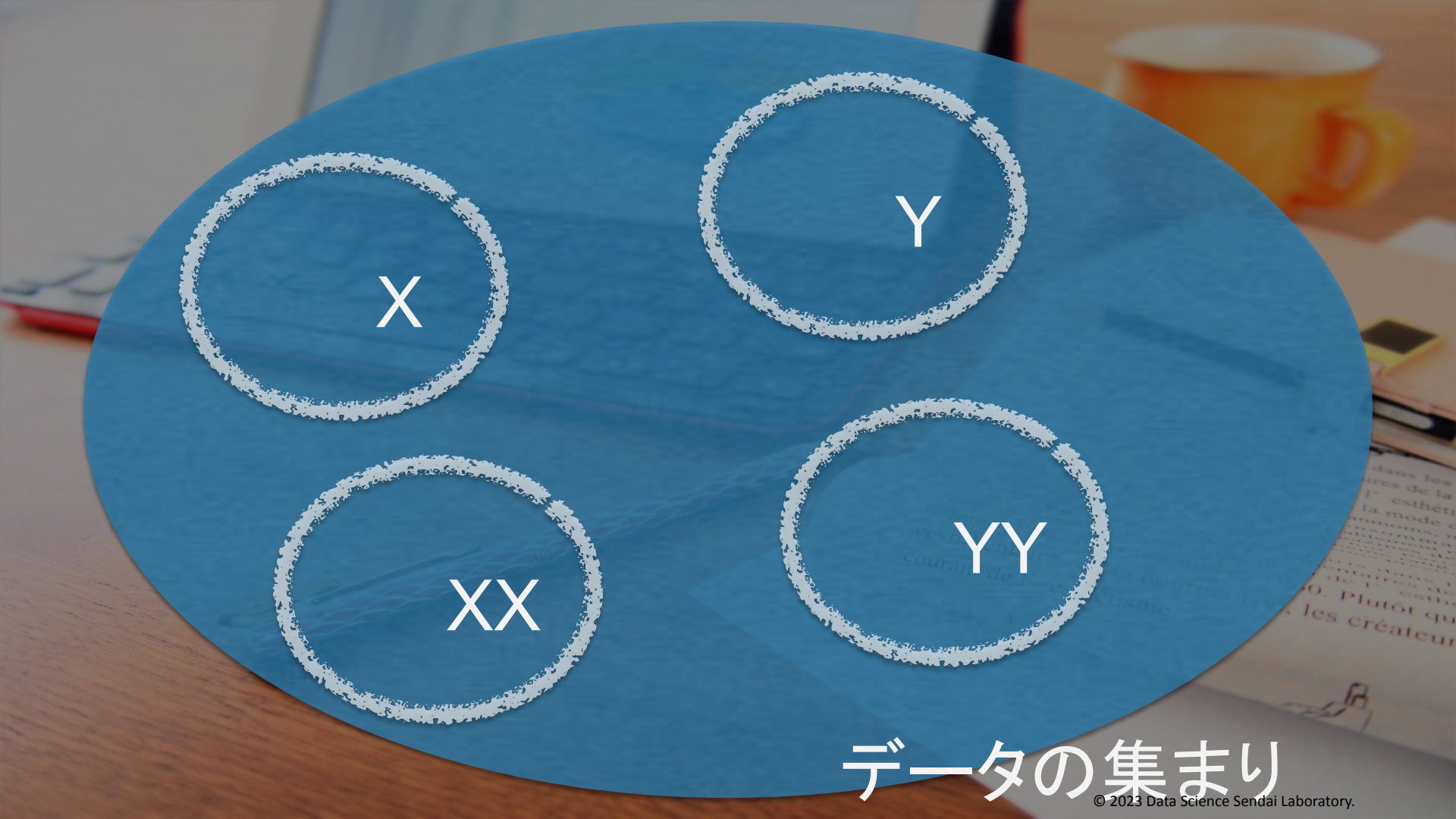
クラスタリングの主な目的

© 2023 Data Science Sendai Laboratory.

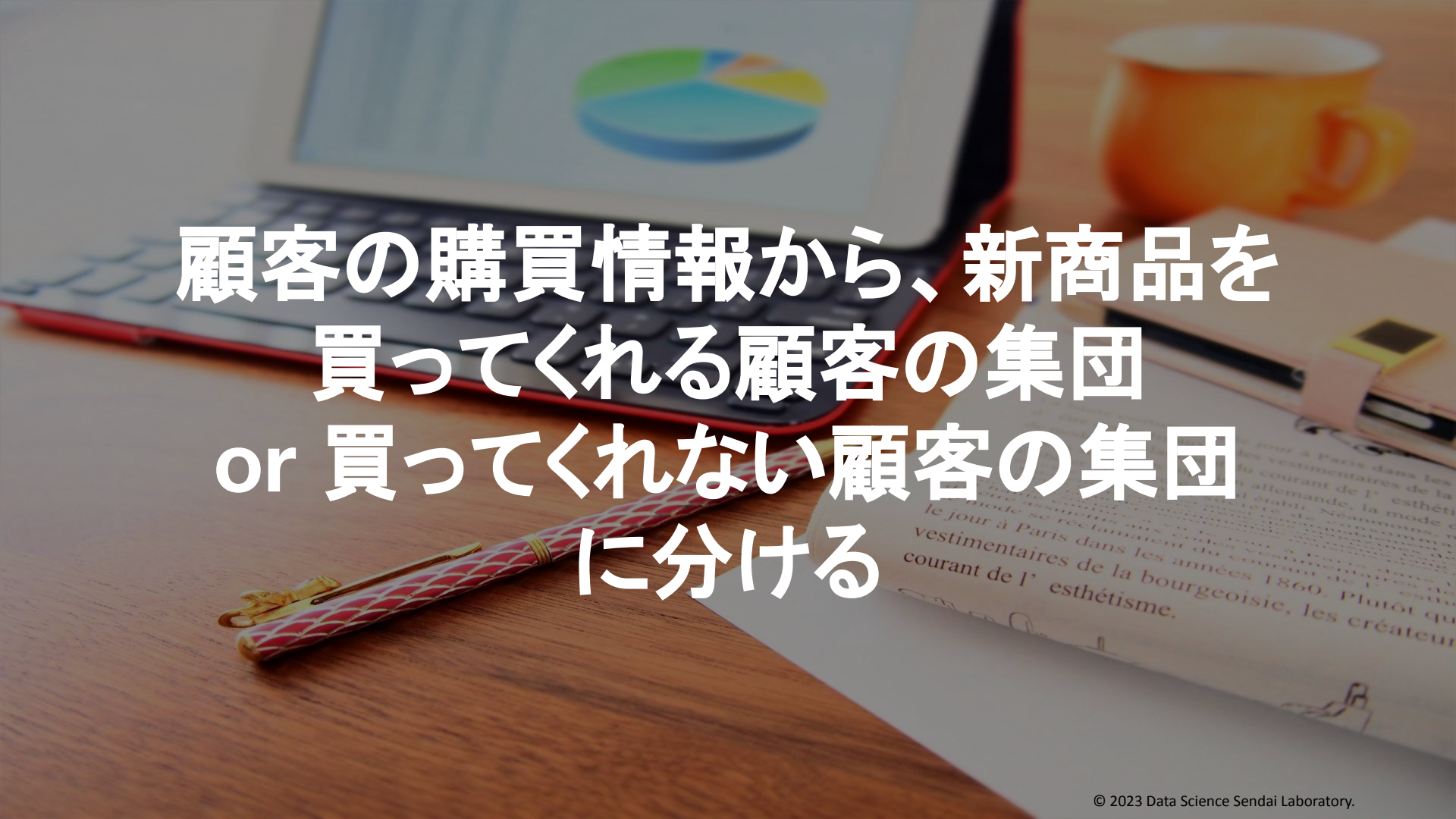


それに分類される
『集合(集団)』を見つけること

© 2023 Data Science Sendai Laboratory.



© 2023 Data Science Sendai Laboratory.



顧客の購買情報から、新商品を買ってくれる顧客の集団 or 買ってくれない顧客の集団に分ける

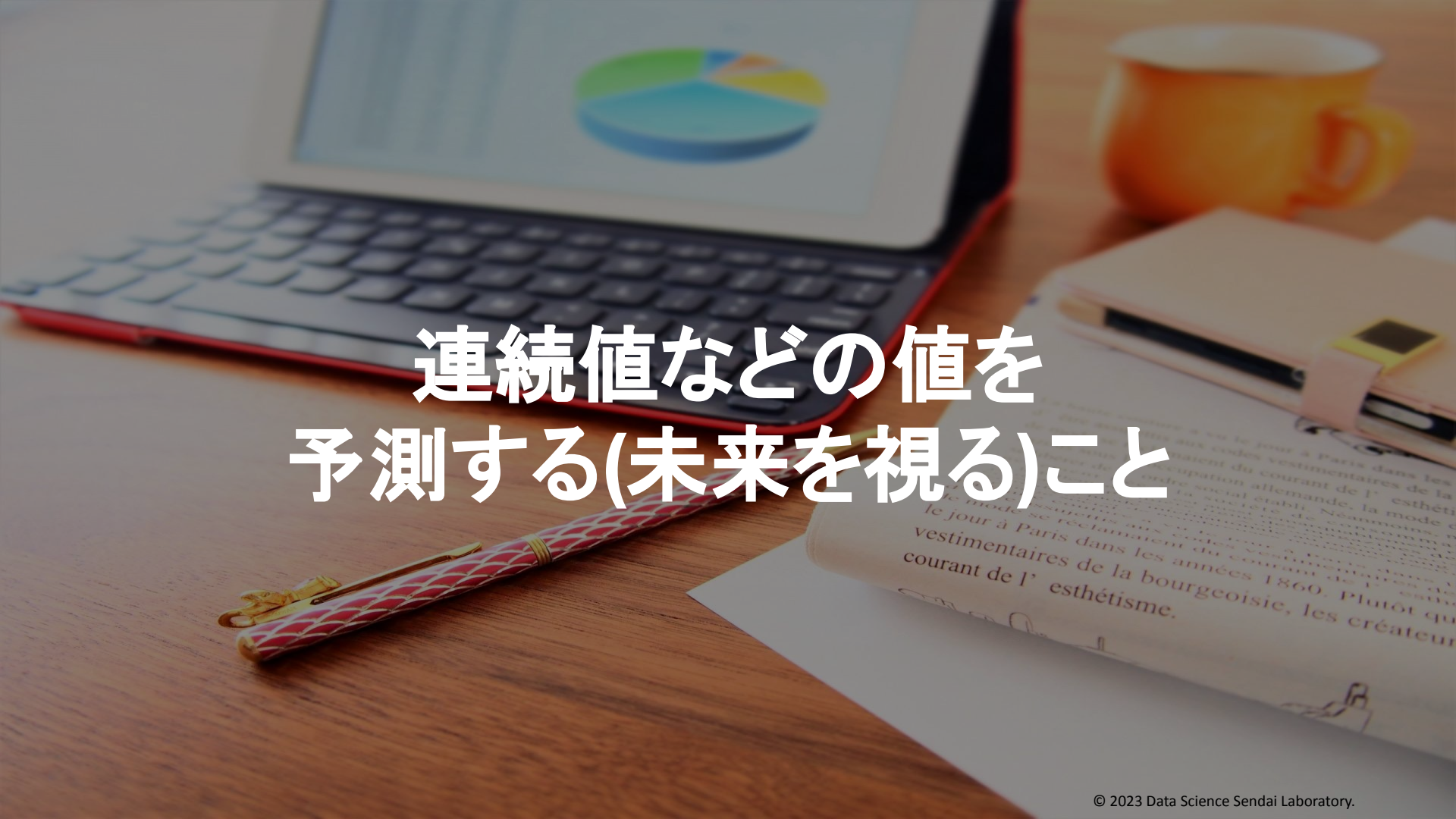
© 2023 Data Science Sendai Laboratory.

学生の点数や講義の出欠率、
受講態度で、
成績Sをとる学生の集団、
成績Aをとる学生の集団、…
に分ける

© 2023 Data Science Sendai Laboratory.

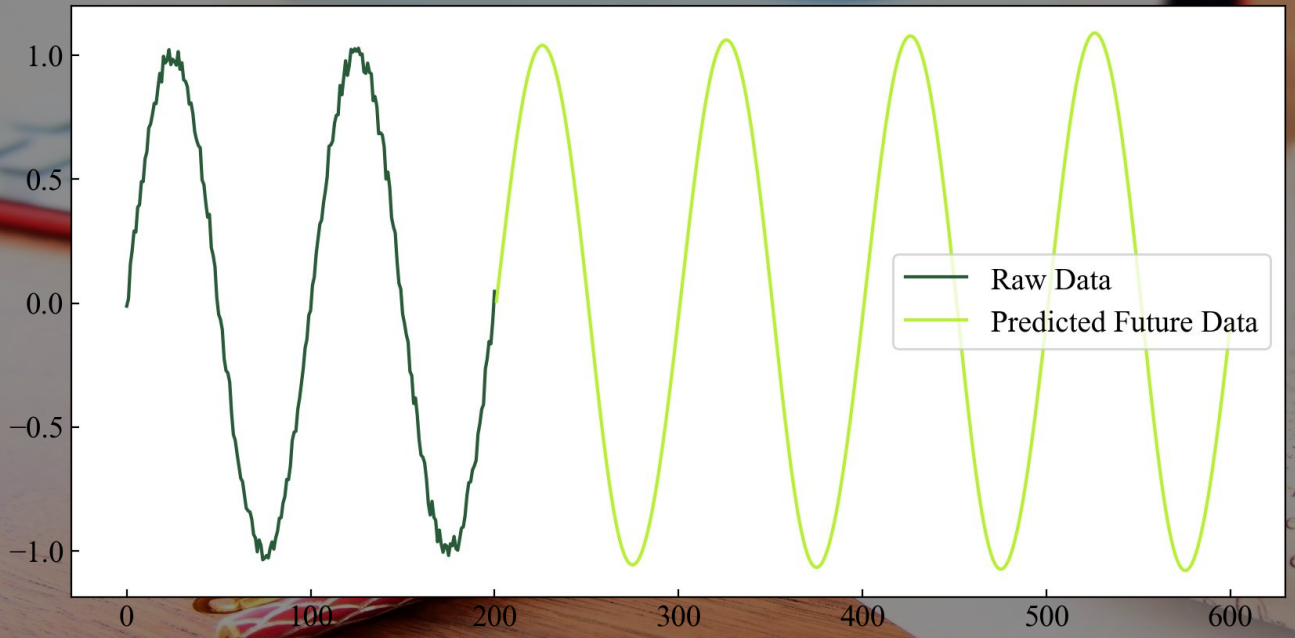
予測の主な目的

© 2023 Data Science Sendai Laboratory.

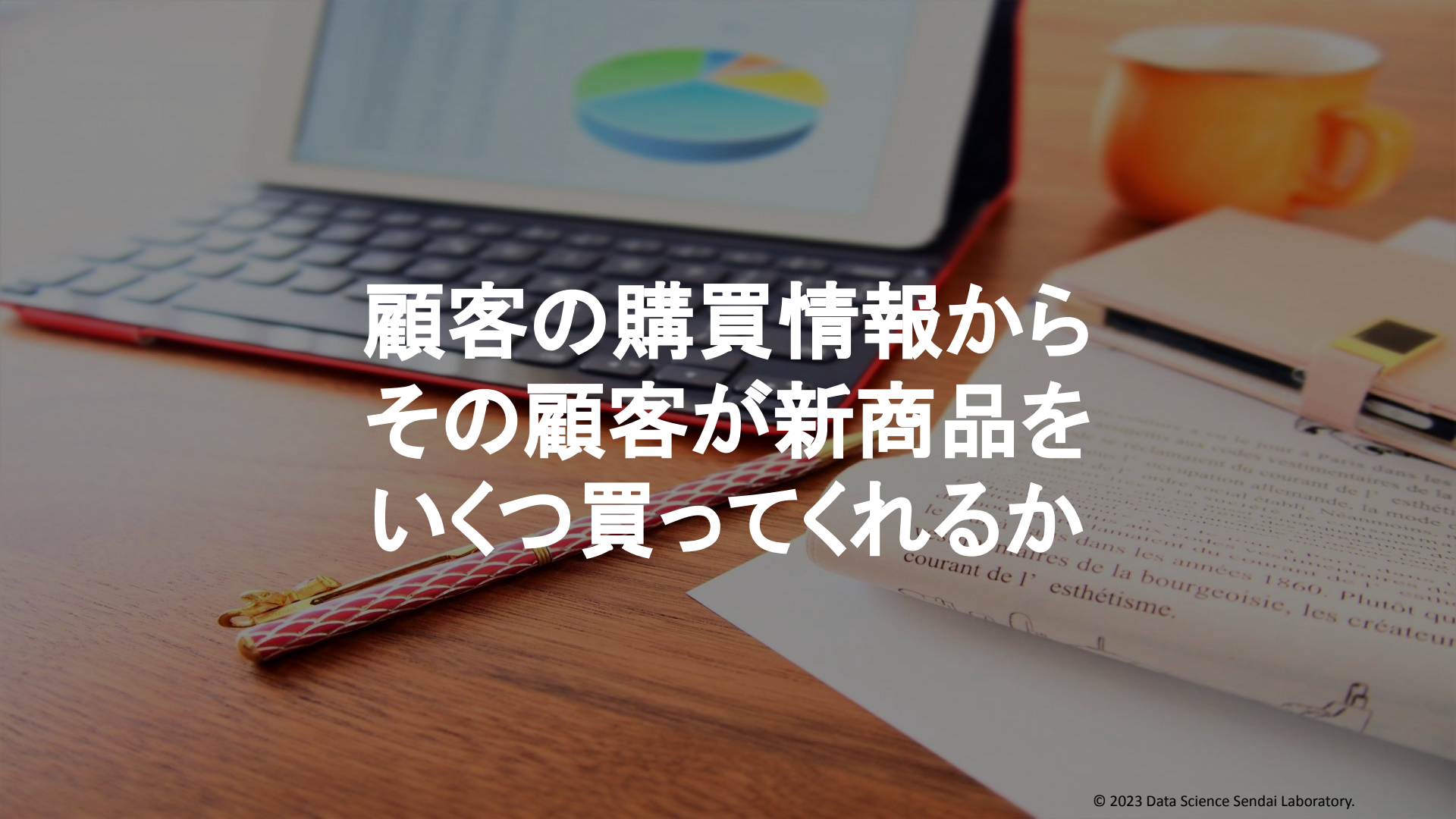


連続値などの値を予測する(未来を視る)こと

© 2023 Data Science Sendai Laboratory.

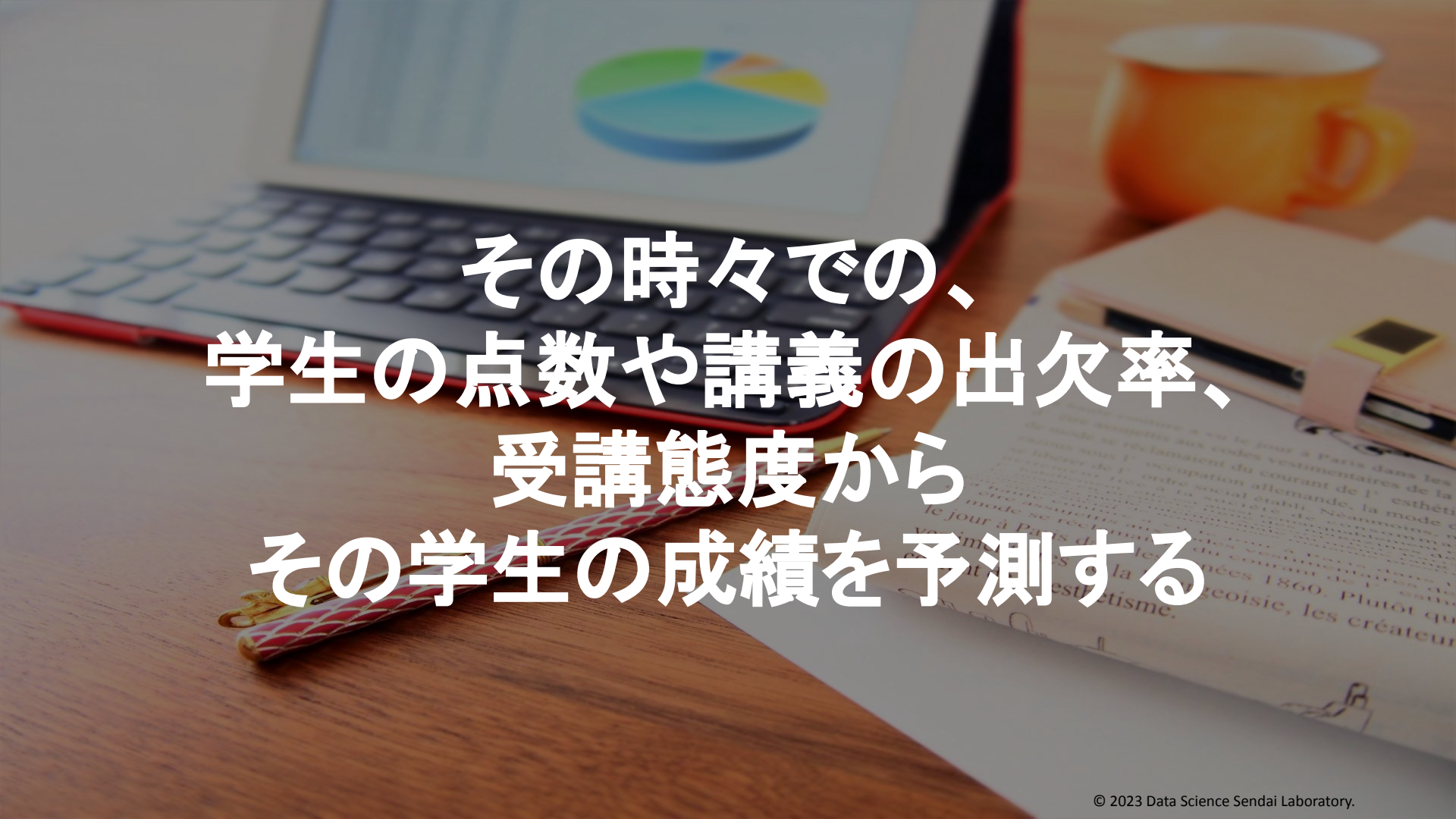


© 2023 Data Science Sendai Laboratory.



顧客の購買情報から
その顧客が新商品を
いくつ買ってくれるか

© 2023 Data Science Sendai Laboratory.



その時々での、
学生の点数や講義の出欠率、
受講態度から
その学生の成績を予測する

© 2023 Data Science Sendai Laboratory.

教師あり学習
代表的なアルゴリズム
（回帰・分類）

© 2023 Data Science Sendai Laboratory.

機械学習モデル

統計的分布を仮定しない

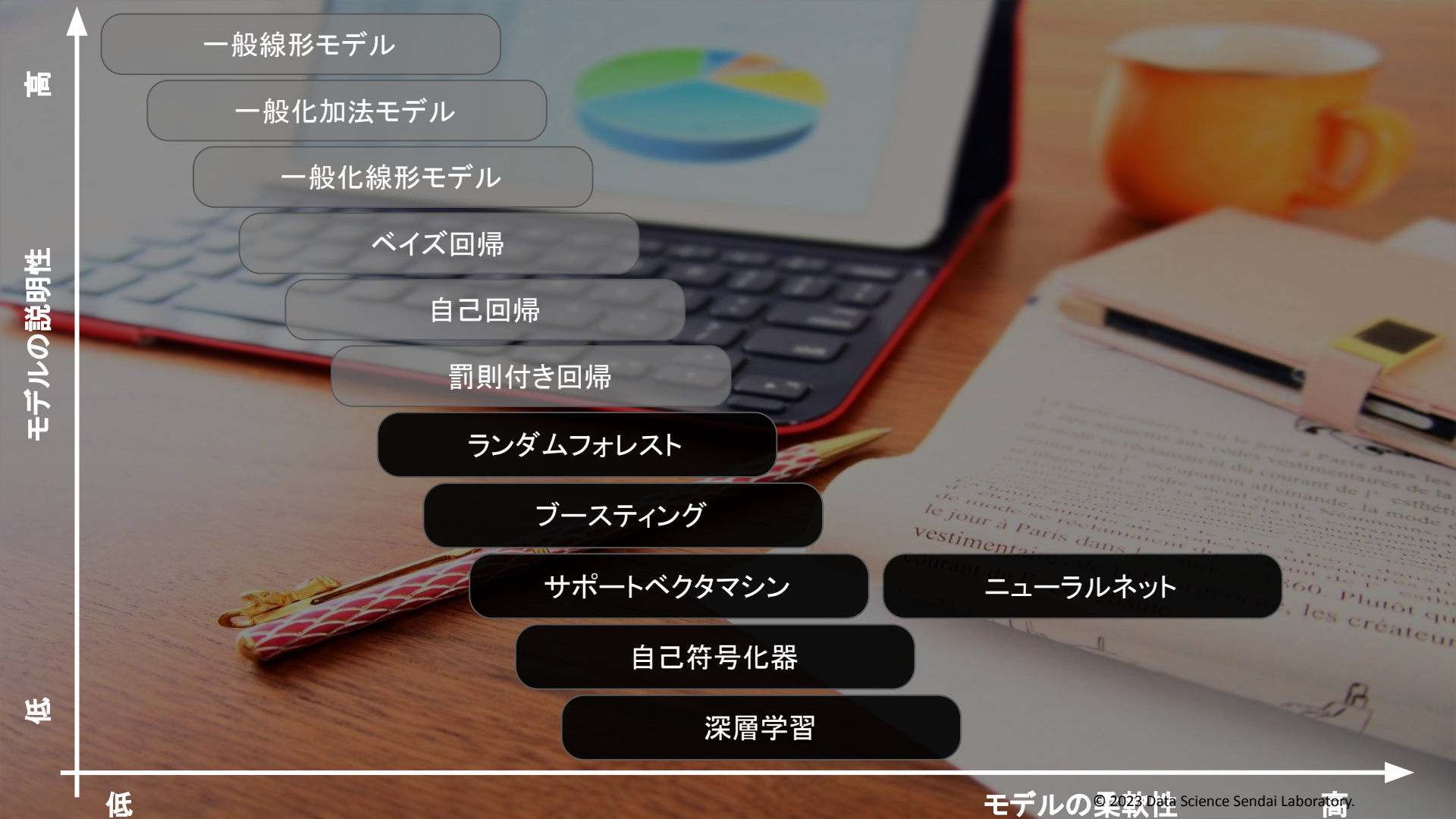
- ランダムフォレスト
- ブースティング
- サポートベクタマシン
- ニューラルネット
- 自己符号化器
- 深層学習

統計モデル

統計的分布を仮定する

- 一般線形モデル
- 一般化加法モデル
- 一般化線形モデル
- ベイズ回帰
- 自己回帰
- 罰則付き回帰

© 2023 Data Science Sendai Laboratory.



© 2023 Data Science Sendai Laboratory.

教師なし学習
代表的なアルゴリズム
（特徴抽出・クラスタリング・次元削減）

© 2023 Data Science Sendai Laboratory.

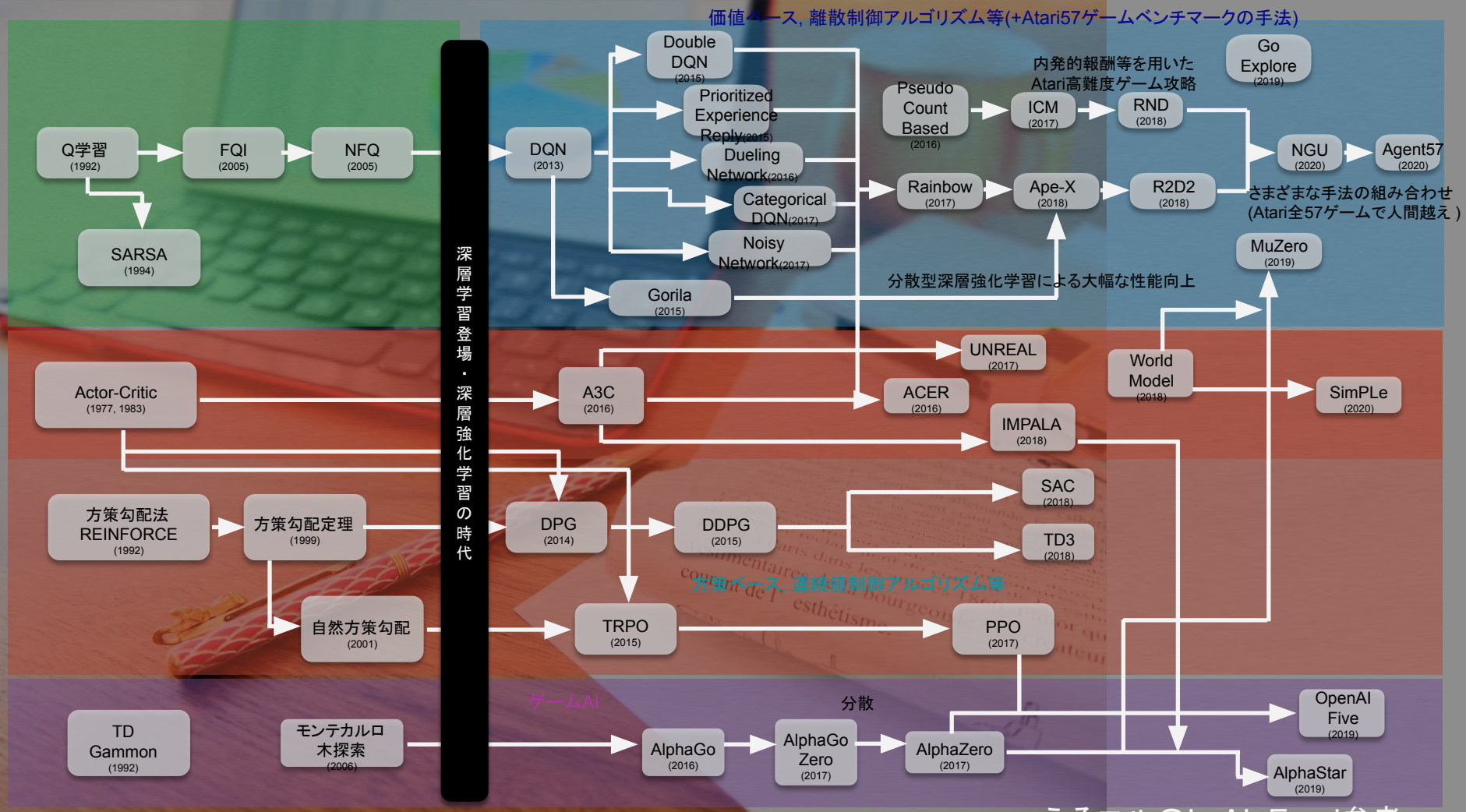
代表的な次元削減手法

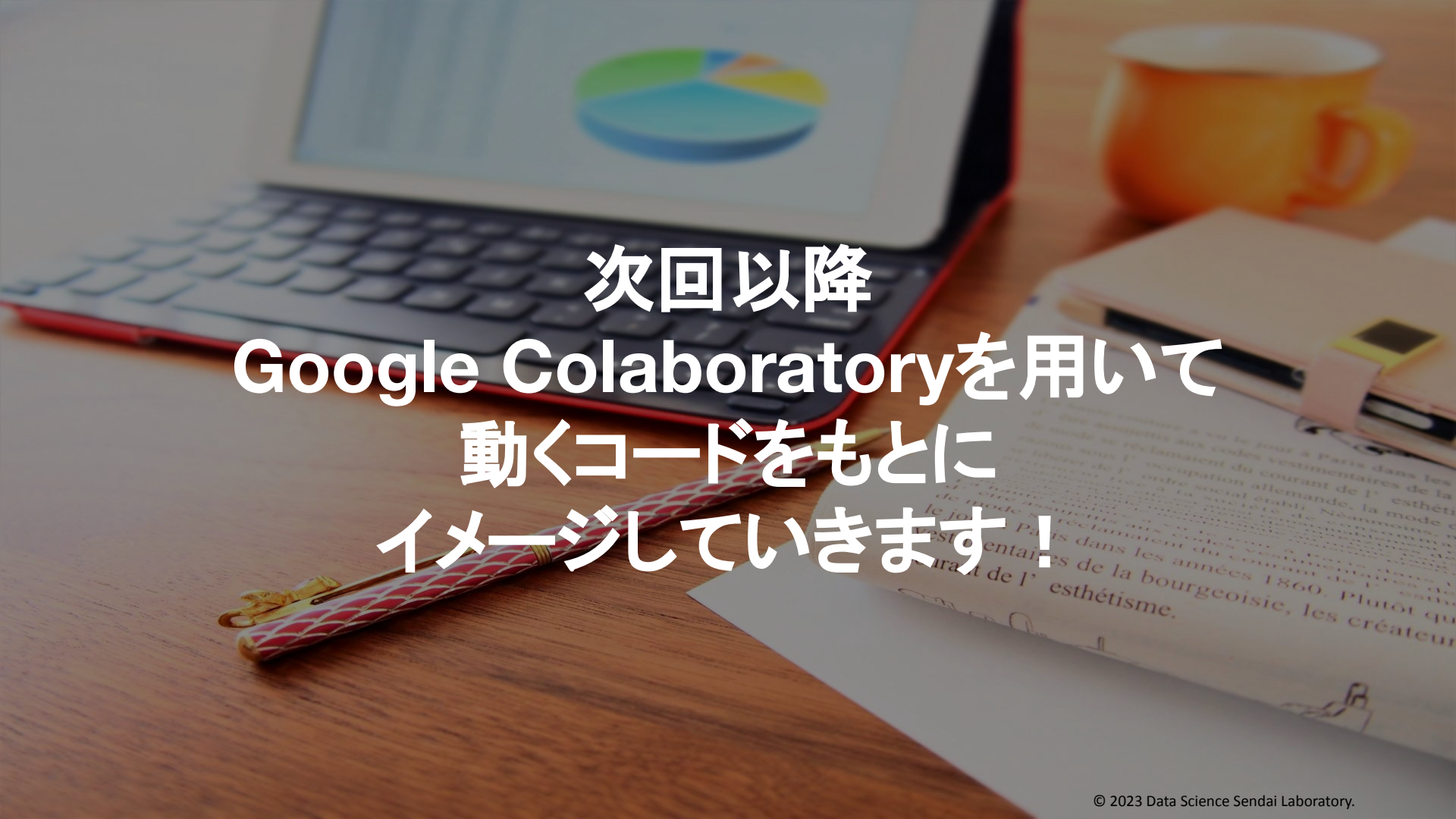
- 主成分分析
- 自己符号化器
- 制約付きボルツマンマシン
- t-SNE
- 自己組織化マップ
- k平均法
- 競合学習
- モンテカルロ法

© 2023 Data Science Sendai Laboratory.

異常検知アルゴリズム（変遷）

© 2023 Data Science Sendai Laboratory.


価値ベース, 離散制御アルゴリズム等(+Atari57ゲームベンチマークの手法)
Q学習 (1992)
FQI (2005)
NFQ (2005)
SARSA (1994)
深層学習登場・深層強化学習の時代
DQN (2013)
Double DQN (2015)
Prioritized Experience Reply (2015)
Dueling Network (2016)
Categorical DQN (2017)
Noisy Network (2017)
Gorila (2015)
Pseudo Count Based (2016)
内発的報酬等を用いた Atari高難度ゲーム攻略
ICM (2017)
RND (2018)
Go Explore (2019)
NGU (2020)
Agent57 (2020)
Rainbow (2017)
Ape-X (2018)
R2D2 (2018)
さまざまな手法の組み合わせ (Atari全57ゲームで人間越え)
分散型深層強化学習による大幅な性能向上
MuZero (2019)
Actor-Critic (1977, 1983)
A3C (2016)
UNREAL (2017)
ACER (2016)
IMPALA (2018)
World Model (2018)
SimPLe (2020)
方策勾配法 REINFORCE (1992)
方策勾配定理 (1999)
自然方策勾配 (2001)
DPG (2014)
DDPG (2015)
SAC (2018)
TD3 (2018)
方策ベース, 連続値制御アルゴリズム等
TRPO (2015)
PPO (2017)
TD Gammon (1992)
モンテカルロ木探索 (2006)
ゲームAI
AlphaGo (2016)
AlphaGo Zero (2017)
分散
AlphaZero (2017)
OpenAI Five (2019)
AlphaStar (2019)
えるエル@ImAI_Eruel参考



次回以降
Google Colaboratoryを用いて
動くコードをもとに
イメージしていきます！

© 2023 Data Science Sendai Laboratory.