

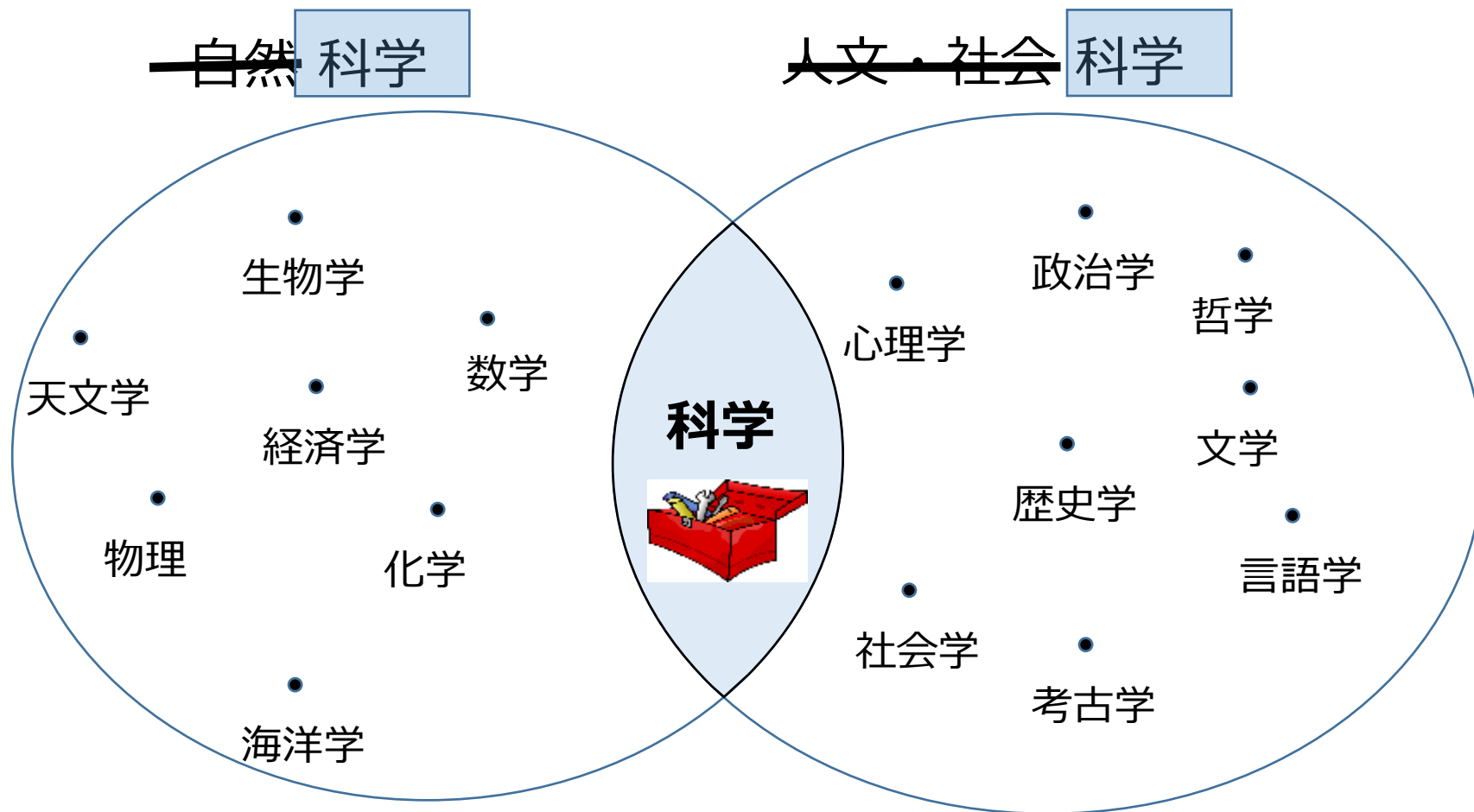
# 機械学習超入門



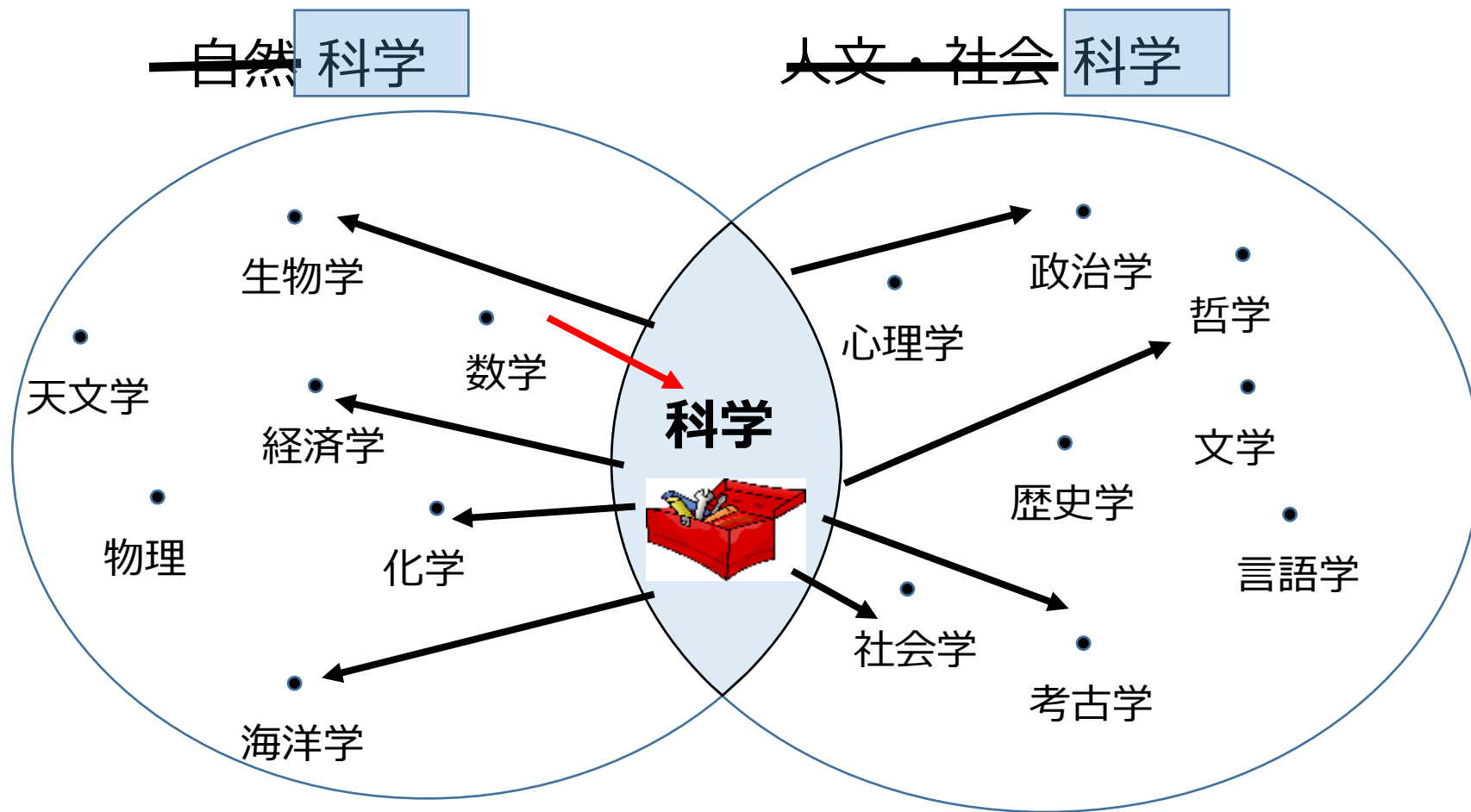
# データサイエンスとは？

---

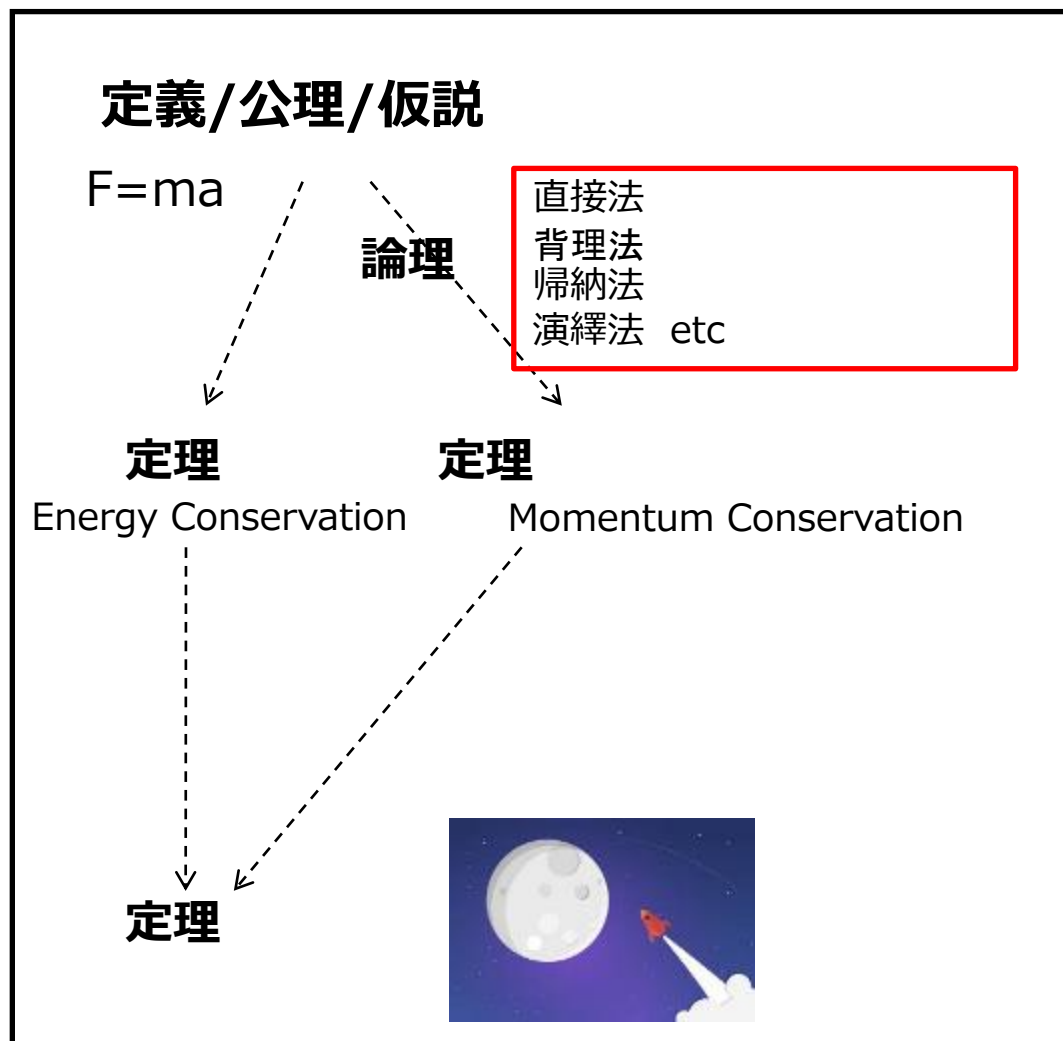
# 科学理論の構造



# 科学理論の構造

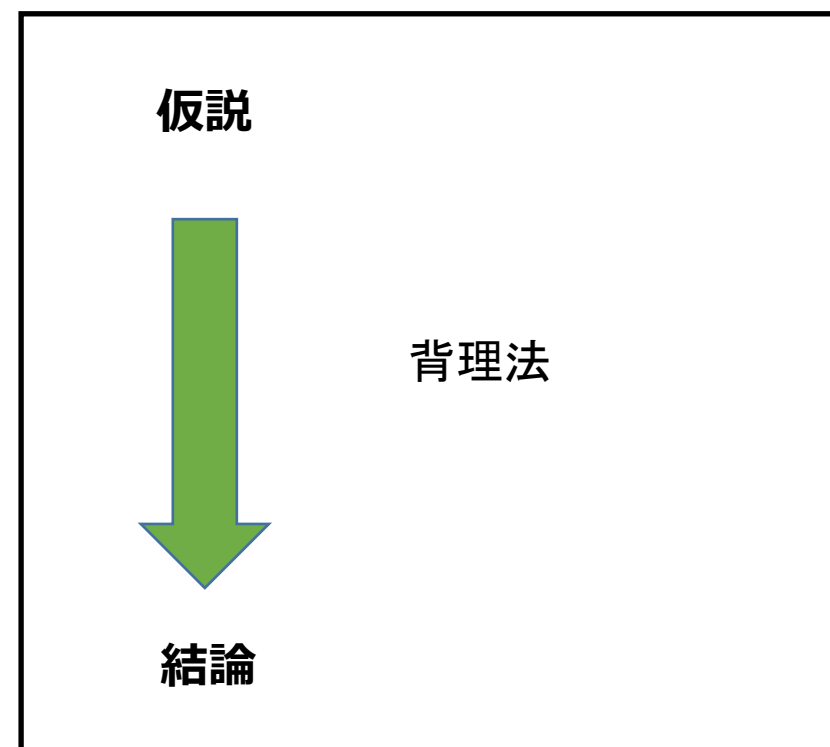


# 科学理論の構造



Theory

## 統計検定の論理構造



# 科学的思考による問題解決

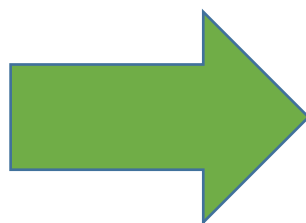
化学

医療

数学

ビジネス

.....



問題解決の為の  
共通アプローチ？

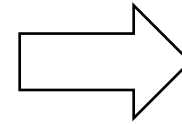
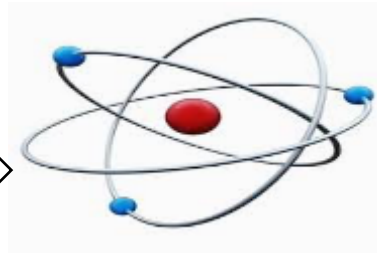
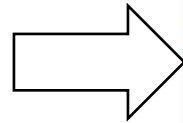
「分解と統合」  
の哲学



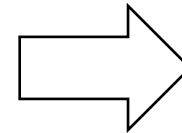
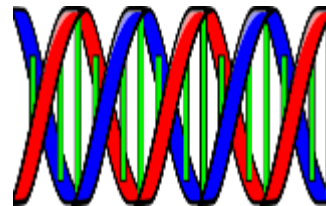
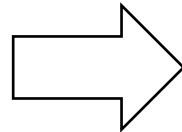
ルネ・デカルト  
(1596-1650)

# 「分解と統合」の哲学

分子

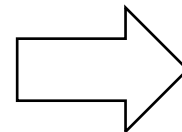


DNA

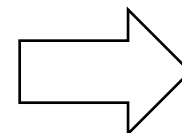


素因数分解

42



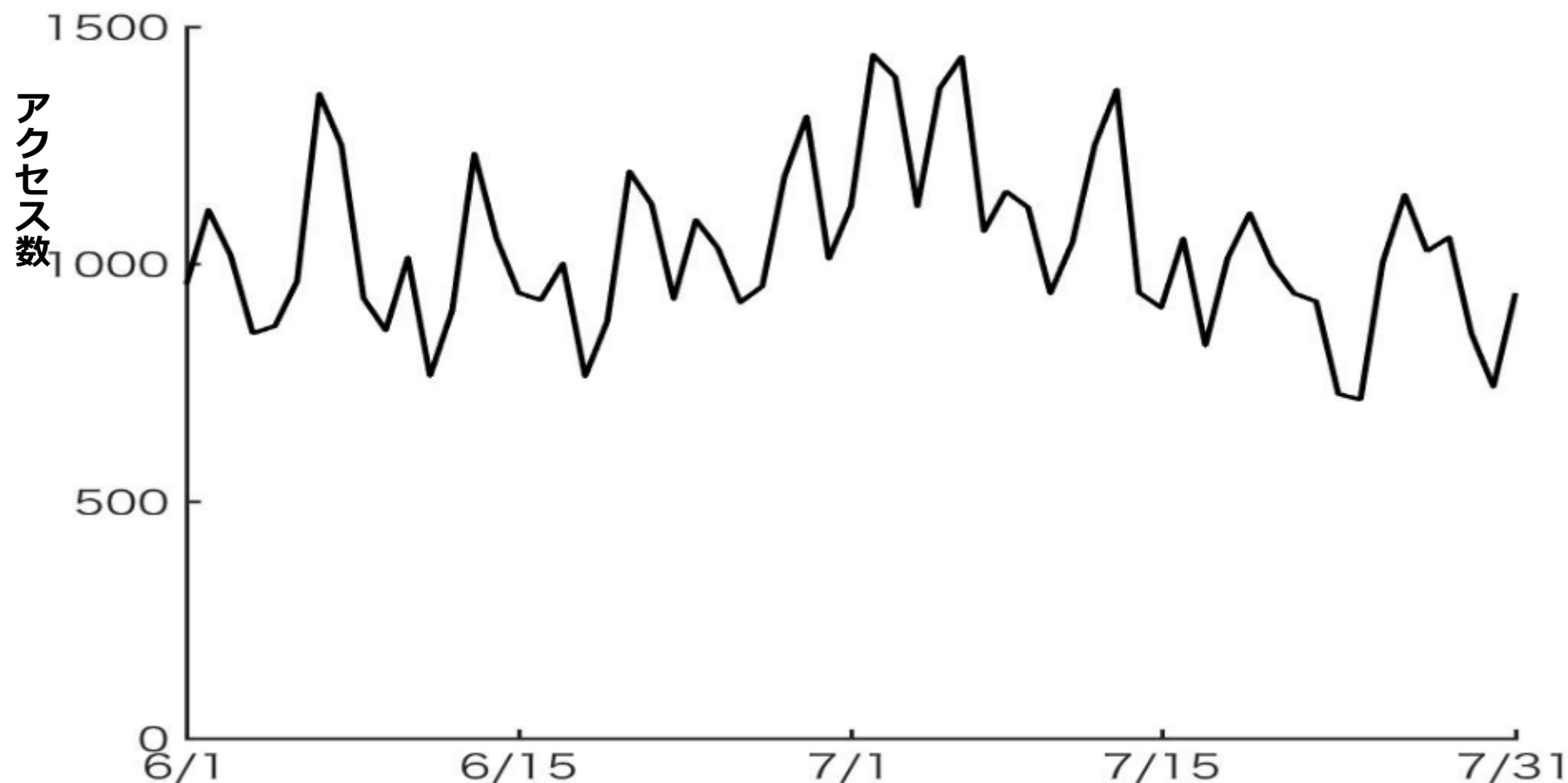
2, 3, 7



$42 = 2 \cdot 3 \cdot 7$

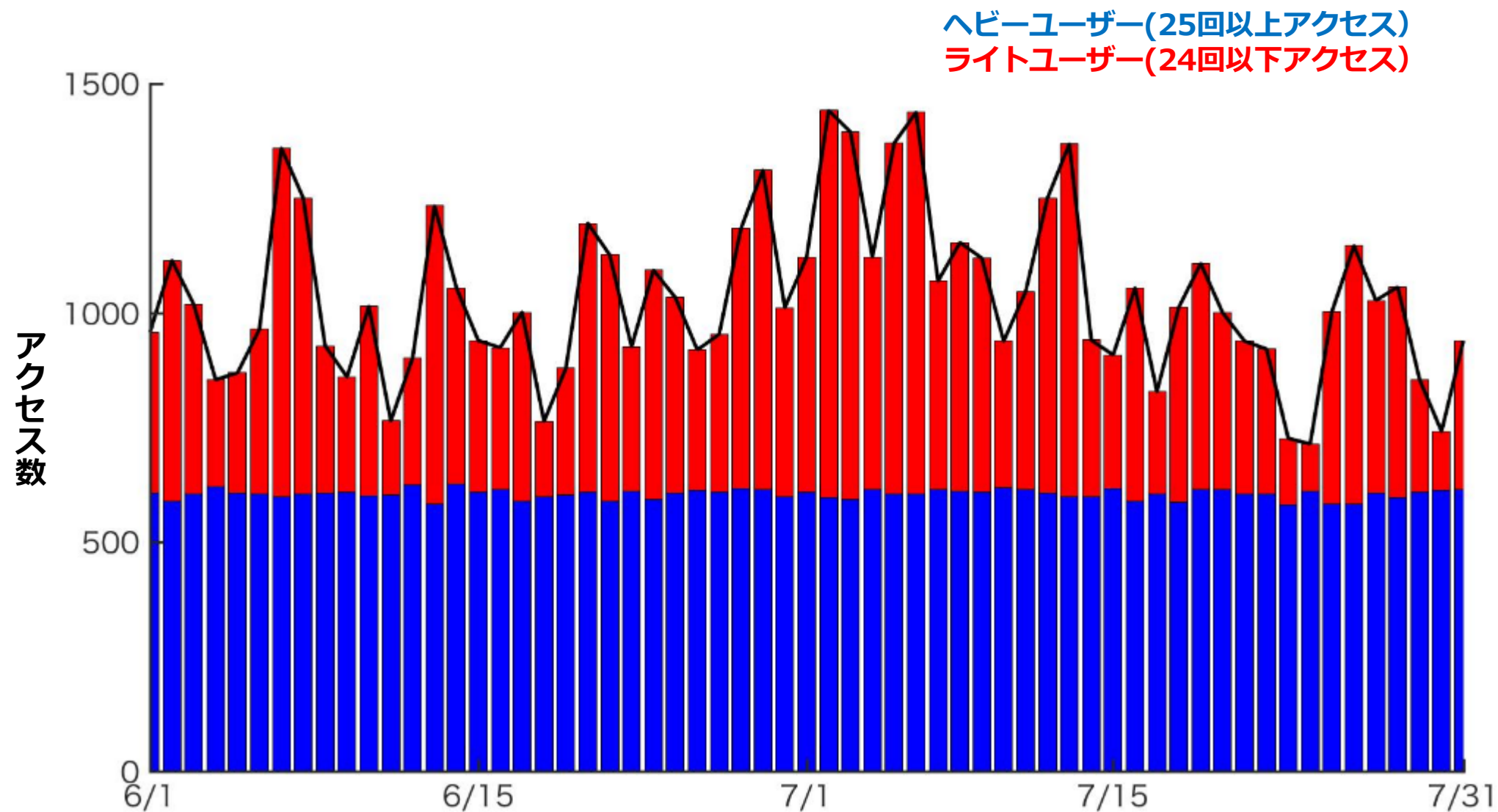
# データ分析の考え方 1

課題：「サイトへの登録者数が減少しているようだが、アクセス数からその原因を調査できないか？」

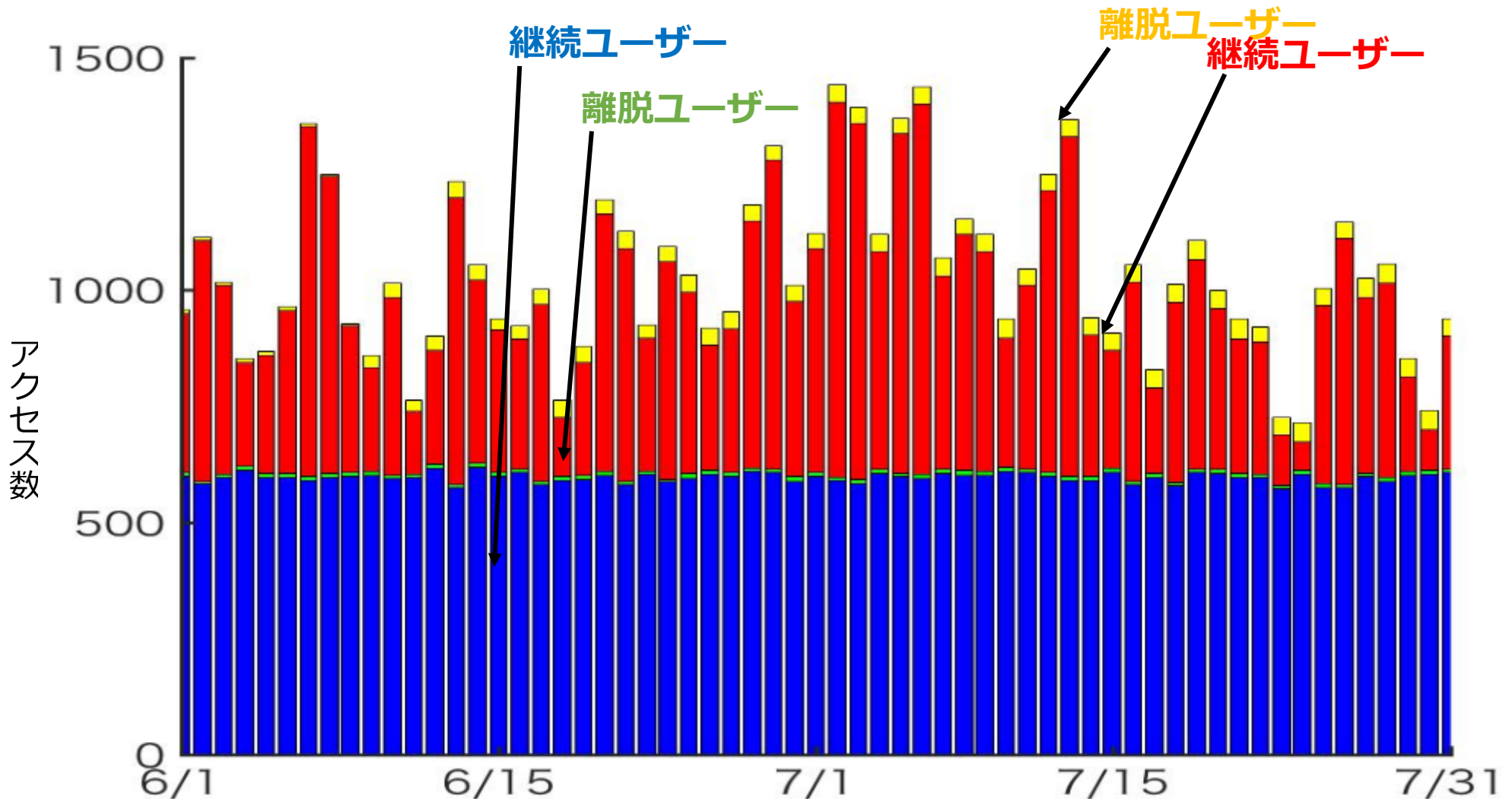




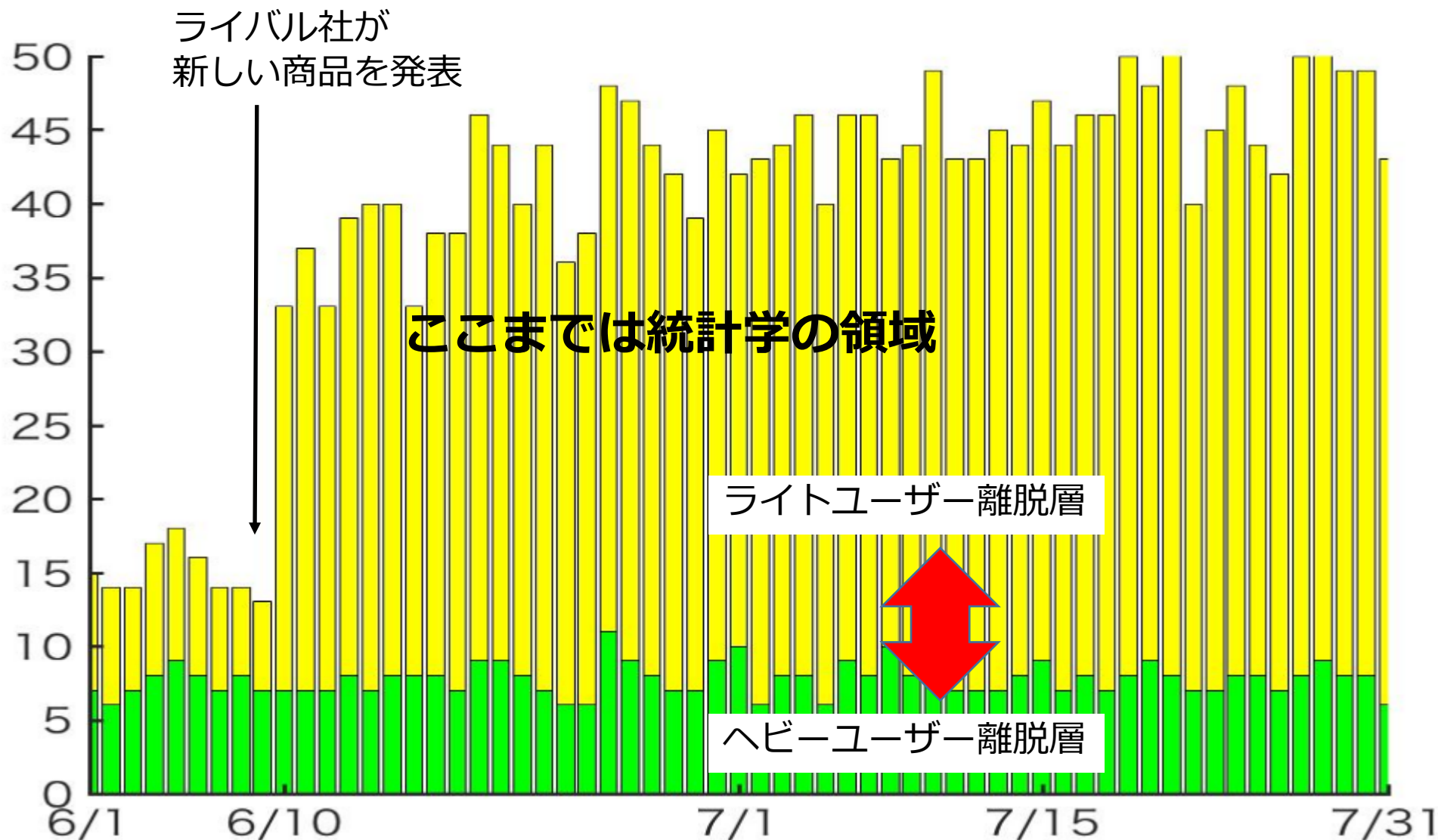
# 分解と統合



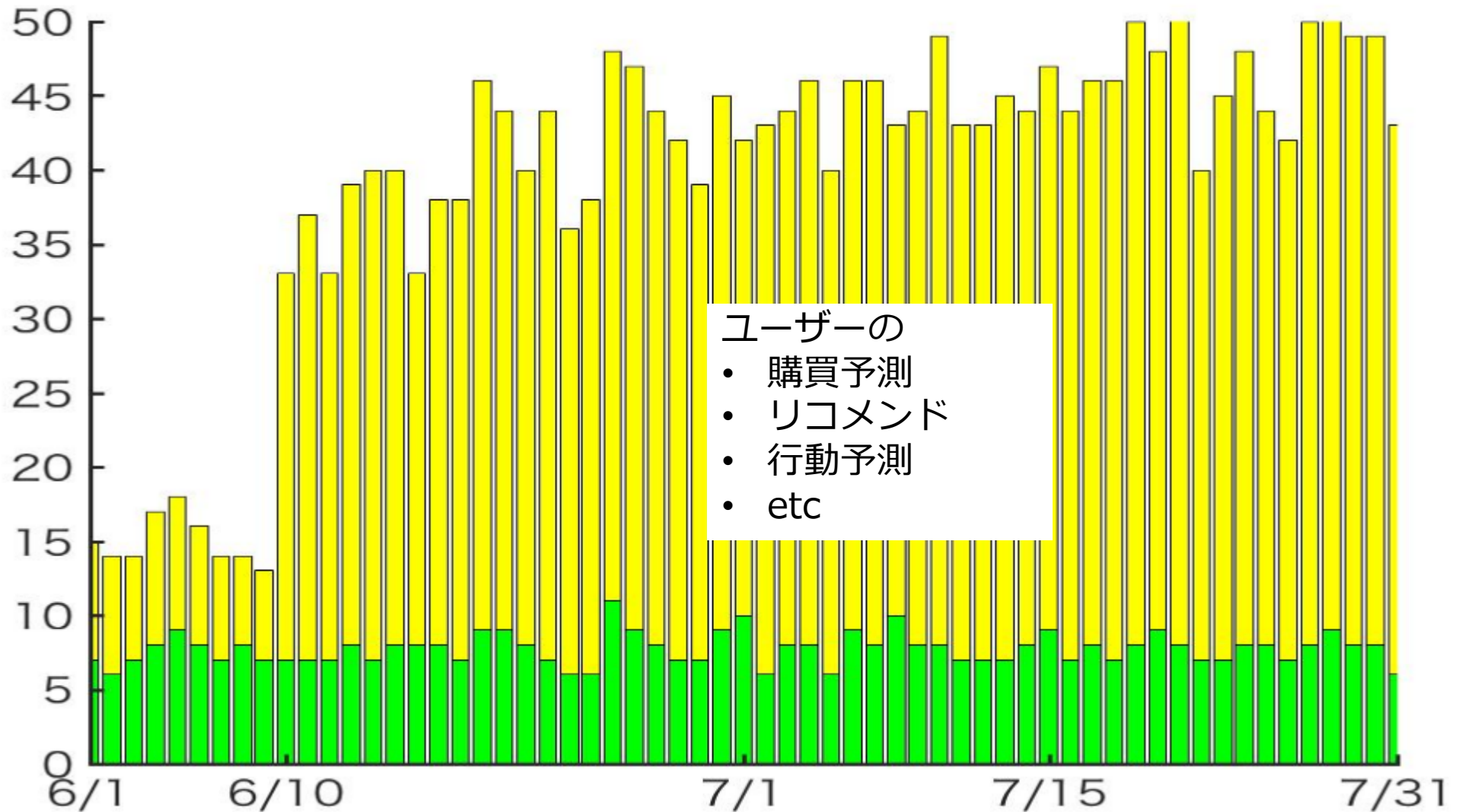
# 分解と統合



# 分解と統合



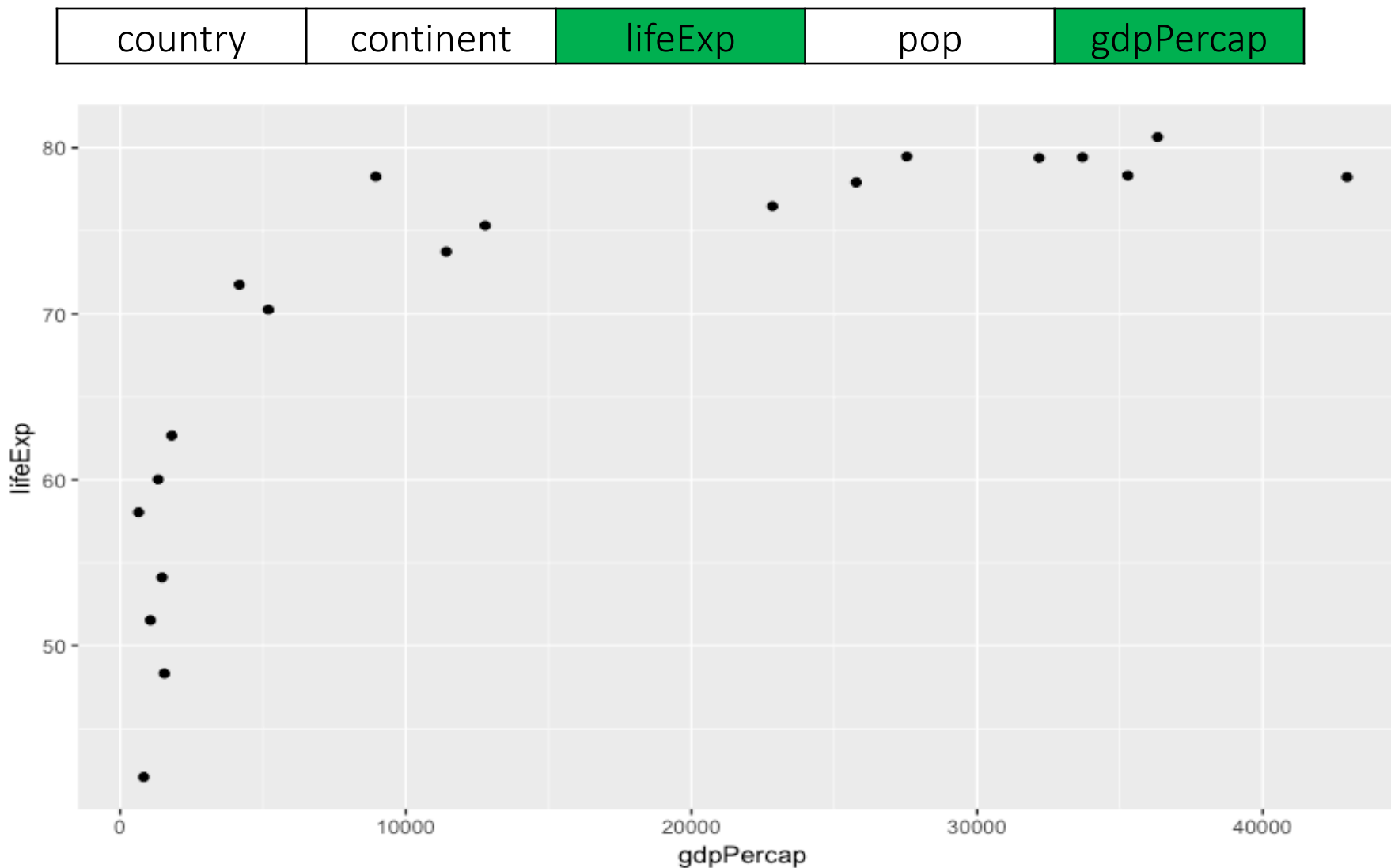
# ここから先が機械学習



# データの可視化

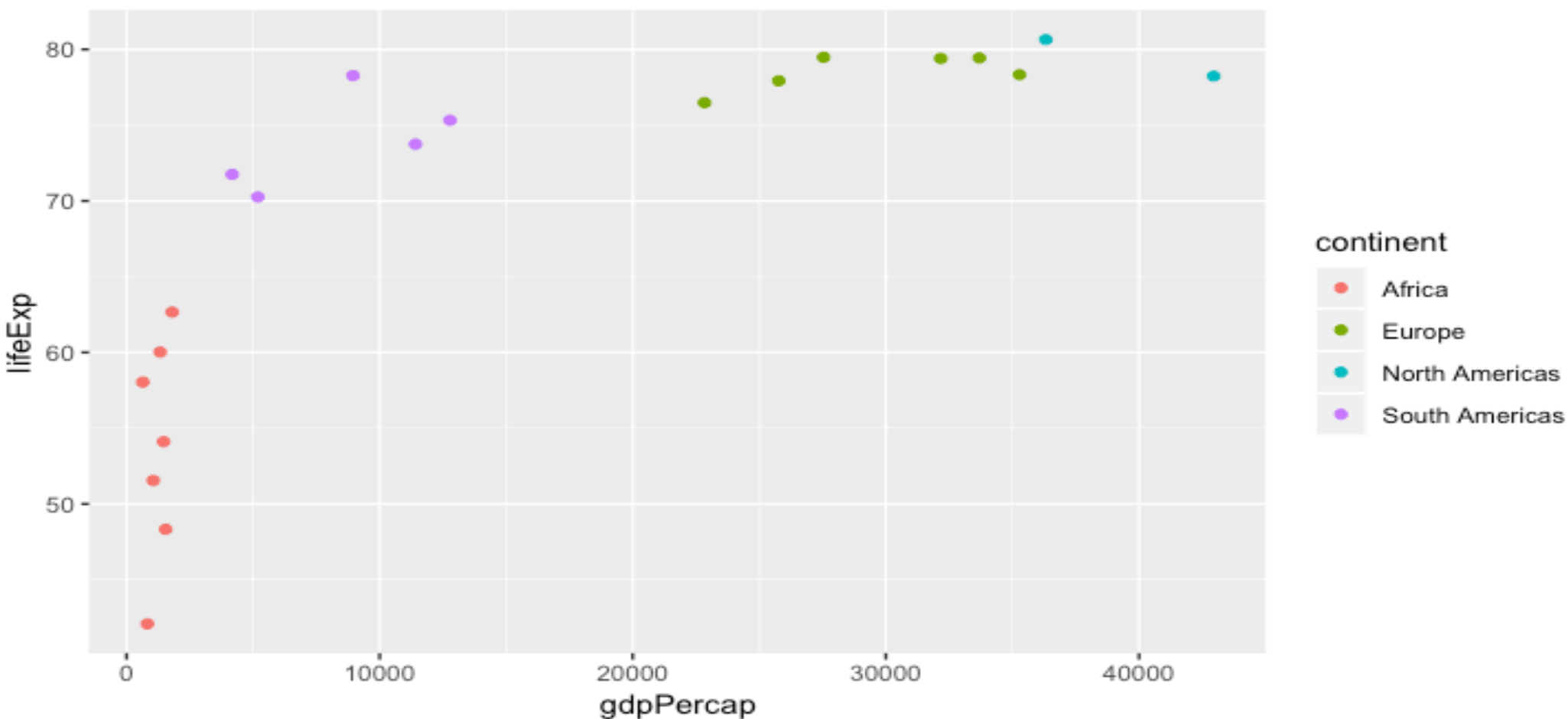
country	continent	lifeExp	pop	gdpPercap
Argentina	South Americas	75.32	40301927	12779.3796
Canada	North Americas	80.653	33390141	36319.235
Cote d'Ivoire	Africa	48.328	18013409	1544.75011
-----	-----	-----	-----	-----
Mauritania	Africa	62.664	3270065	1803.1515
Belgium	Europe	79.441	10392226	33692.6051

# データの可視化

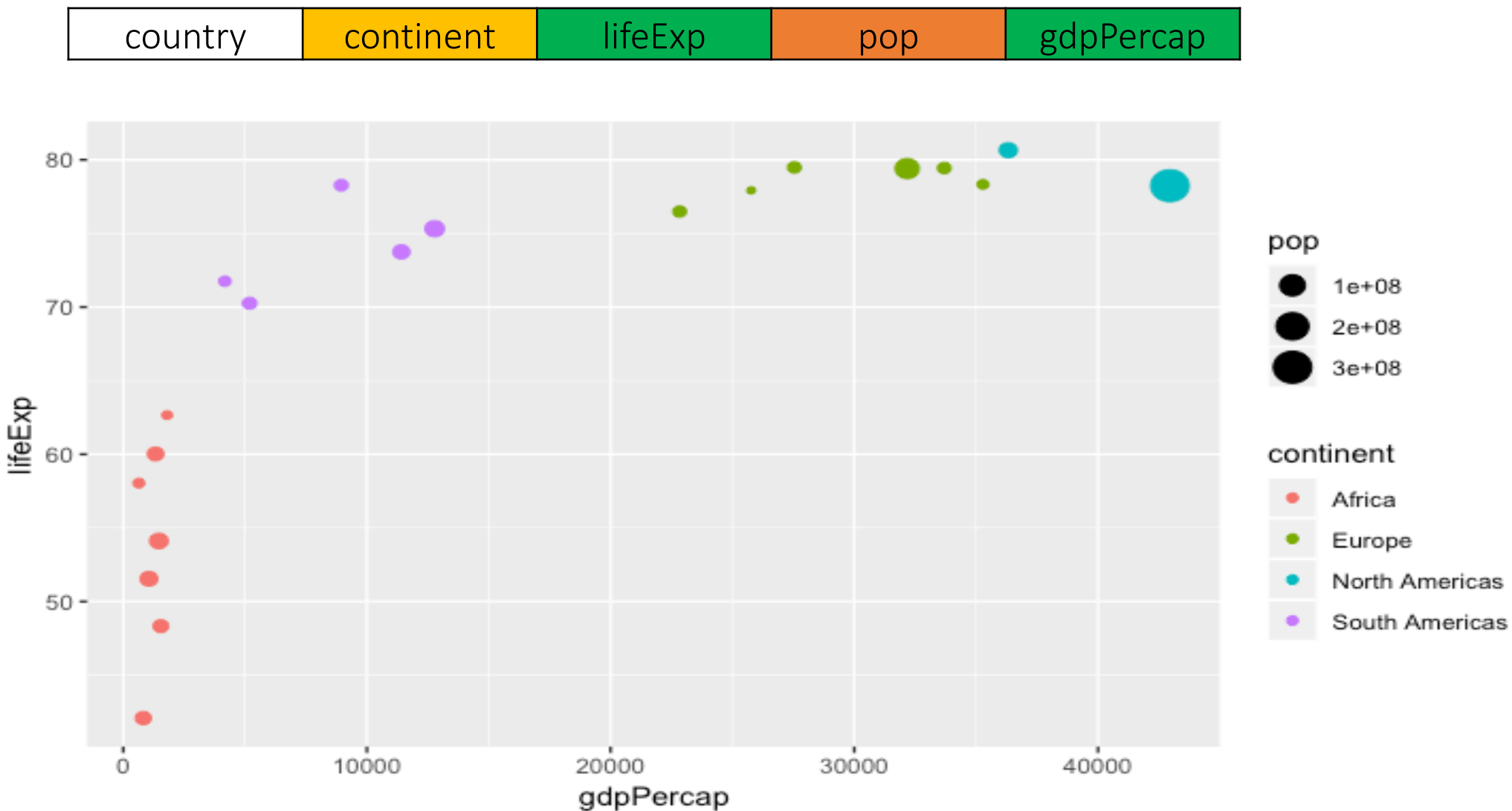


# データの可視化

country	continent	lifeExp	pop	gdpPercap
---------	-----------	---------	-----	-----------

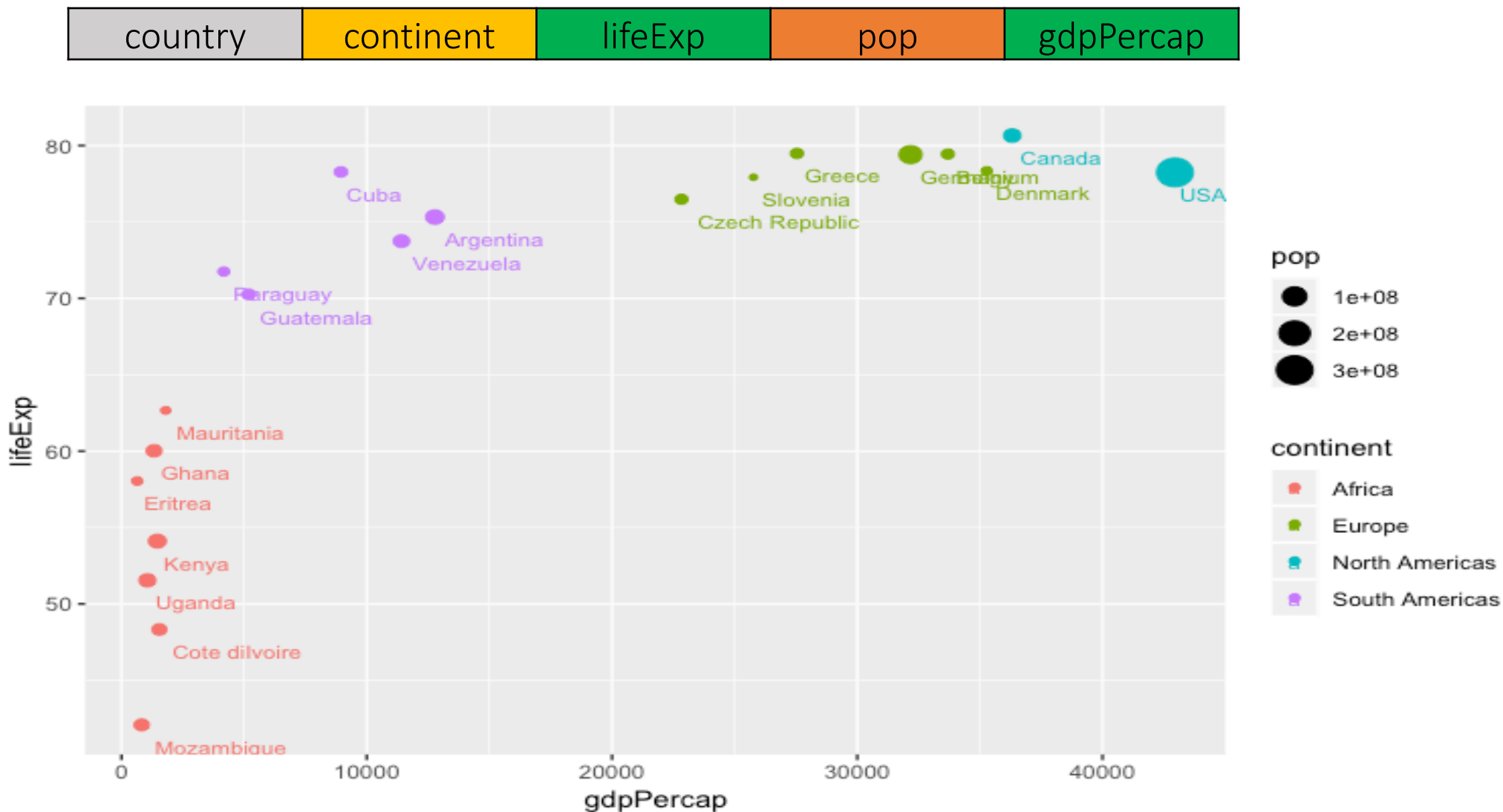


# データの可視化





# データの可視化



# 検定

オバマ大統領が簡単なテストで、6000万ドルもの収益を上げた方法



# 検定の応用

## Join ABCSPORTS

Username:

Email:

Password:

☐ I accept the Terms and Conditions

Sign up +

**Type A**

## Join ABCSPORTS

Username:

Email:

Password:

☐ I accept the Terms and Conditions

100% privacy. We will never spam you !

Sign up +

**Type B**

# 検定の応用

Type A

6/1	6/2	6/3	6/4	----	6/29	6/30
250	333	560	521	----	390	430

**1 日平均  
445**

Type B (100% privacy. We will never spam you ! )

6/1	6/2	6/3	6/4	----	6/29	6/30
159	253	462	412	-----	350	320

**1 日平均  
405**

タイプAとBのサインアップ数の間に違いがあるのか？

# 検定の応用

## Join ABCSPORTS

Username:

Email:

Password:

☐ I accept the Terms and Conditions

Sign up +

## Join ABCSPORTS

Username:

Email:

Password:

☐ I accept the Terms and Conditions

100% privacy. We will never spam you !

Sign up +

## Type B

## Type A



18 % less signups

# 検定の応用

## Join ABCESPORTS

Username:

Email:

Password:

☐ I accept the Terms and Conditions

Sign up +

## Join ABCSPORTS

Username:

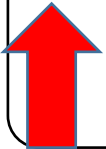
Email:

Password:

☐ I accept the Terms and Conditions

We guarantee 100% privacy. Your information will not be shared.

Sign up +



16 % more signups

# ABテスト

パターン1（画像、オリジナル）：「Obama」の旗に囲まれる柔らかな

パターン2（画像）：家族と一緒に写っている写真

パターン3（画像）：正面からアップで撮影した凛々しい表情の写真



# ABテスト

パターン1（オリジナル）：SIGN UP「会員登録」

パターン2：SIGN UP NOW「今すぐ会員登録」

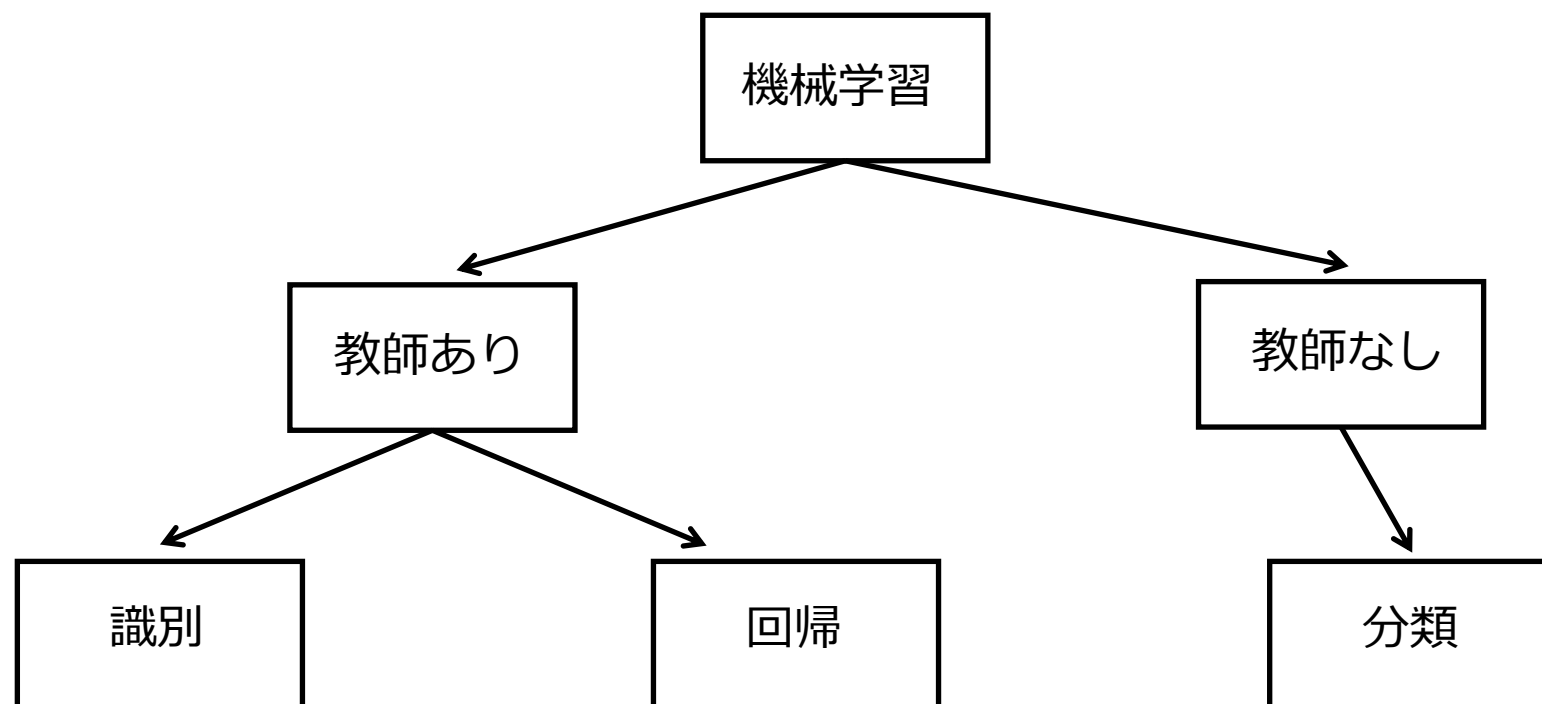
パターン3：JOIN US NOW「今すぐ参加する」

パターン4：LEARN MORE「もっと詳しく」

A red rectangular button with rounded corners and a black border, containing the text "SIGN UP" in white, bold, uppercase letters.A red rectangular button with rounded corners and a black border, containing the text "SIGN UP NOW" in white, bold, uppercase letters.A red rectangular button with rounded corners and a black border, containing the text "JOIN US NOW" in white, bold, uppercase letters.A red rectangular button with rounded corners and a black border, containing the text "LEARN MORE" in white, bold, uppercase letters.



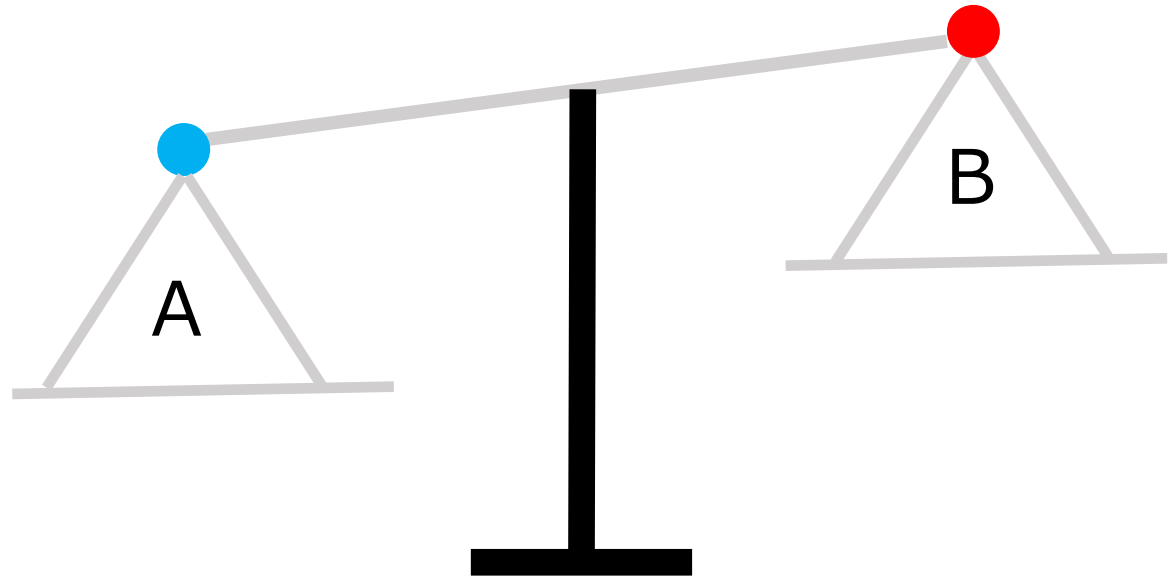
# 機械学習



# 機械学習で問われる 3つの質問

- 質問 1  
「AかBか？」

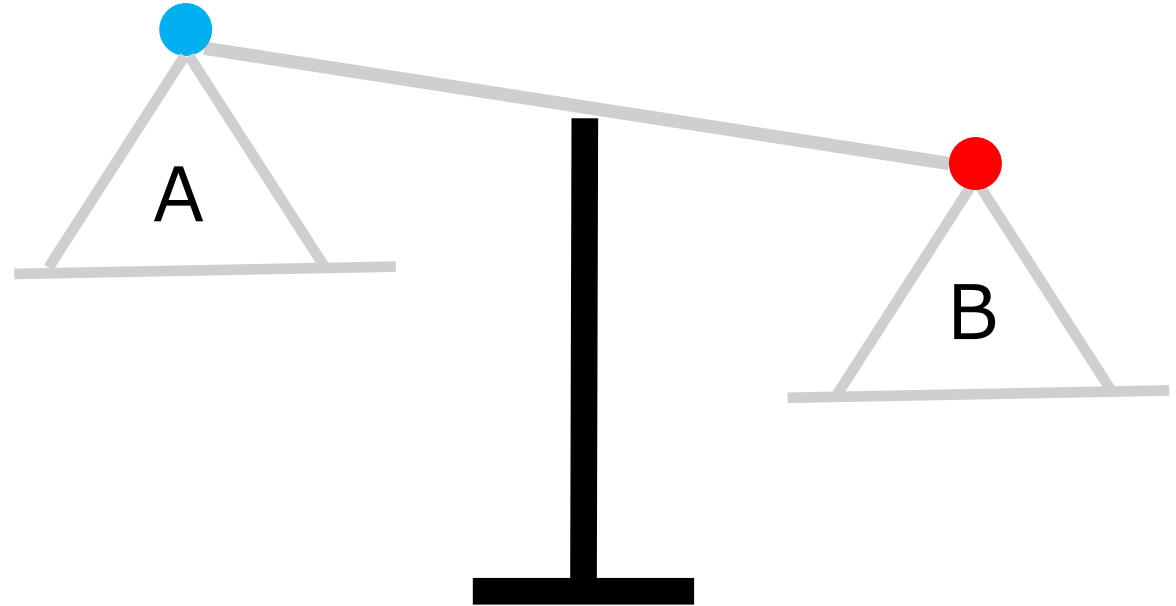
識別アルゴリズム



# 機械学習で問われる 3つの質問

- 質問 1  
「AかBか？」

識別アルゴリズム

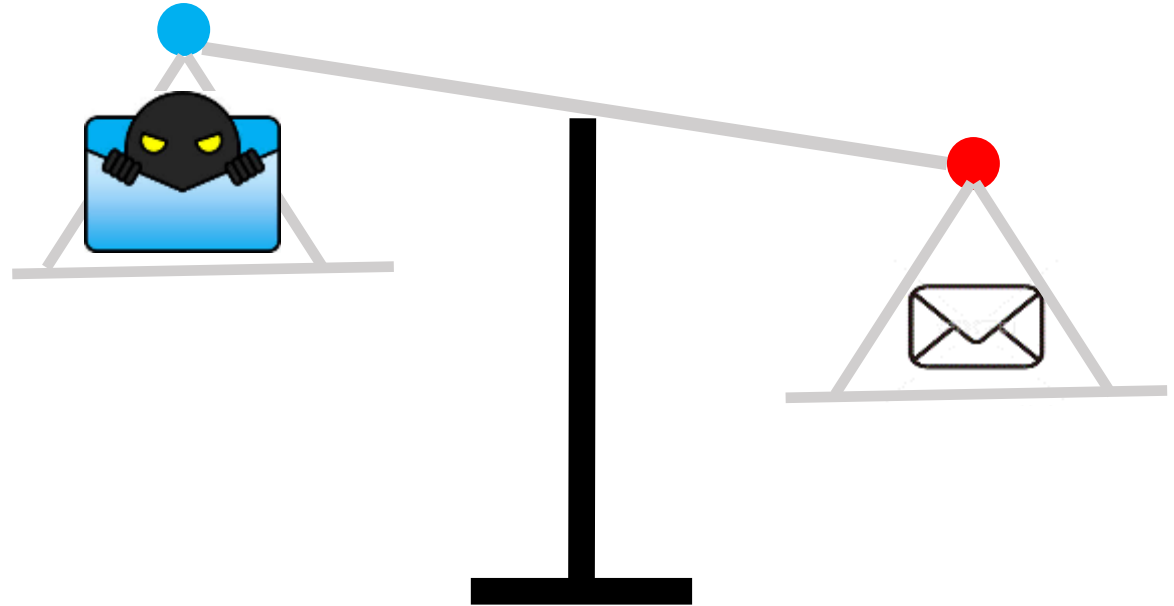


# 機械学習で問われる 3つの質問

- 質問 1  
「AかBか？」

識別アルゴリズム

ナイーブベイズ

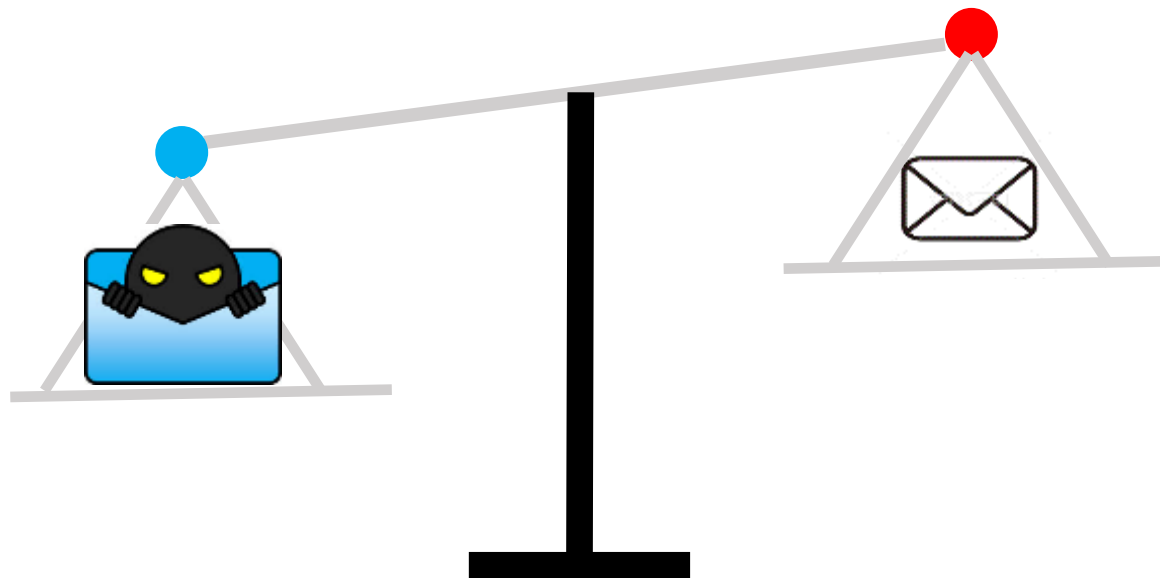


# 機械学習で問われる 3つの質問

- 質問 1  
「AかBか？」

識別アルゴリズム

ナイーブベイズ

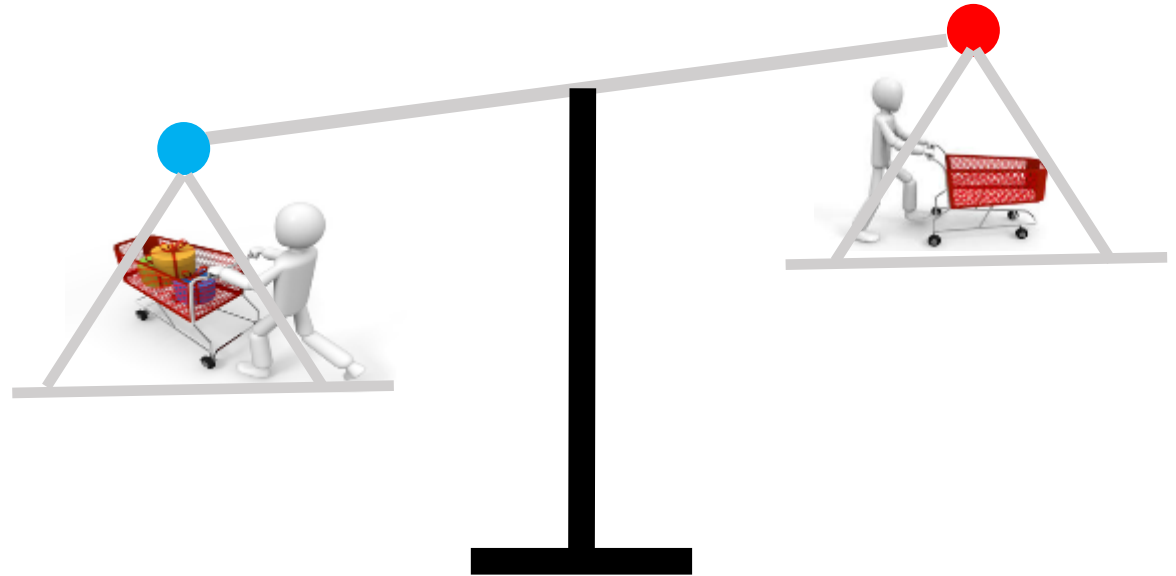


# 機械学習で問われる 3つの質問

- 質問 1  
「AかBか？」

識別アルゴリズム

決定木

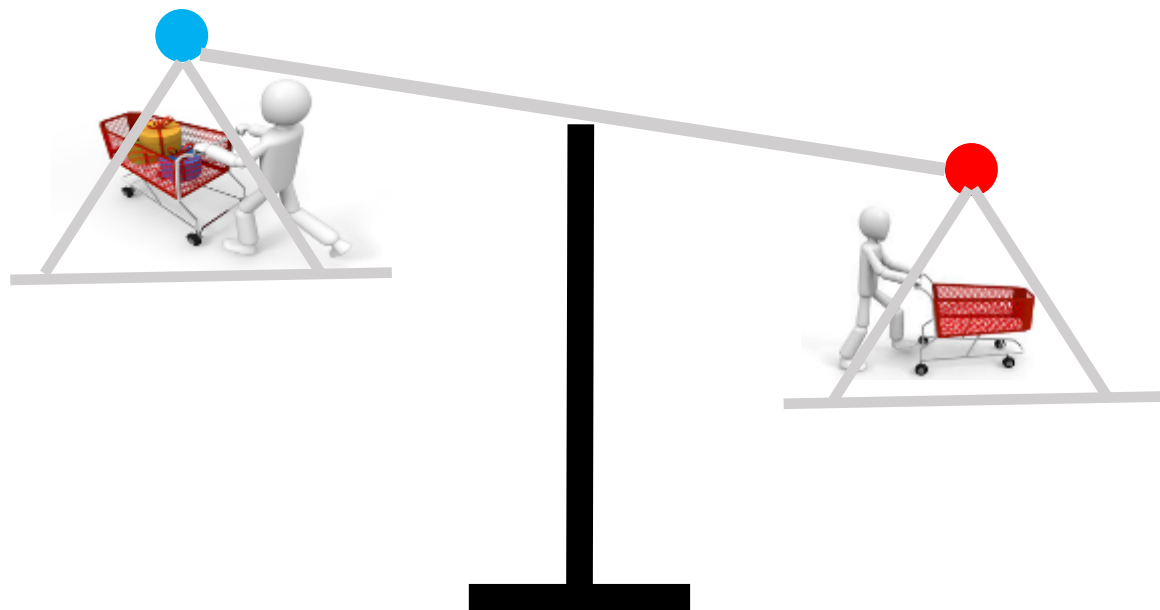


# 機械学習で問われる 3つの質問

- 質問 1  
「AかBか？」

識別アルゴリズム

決定木



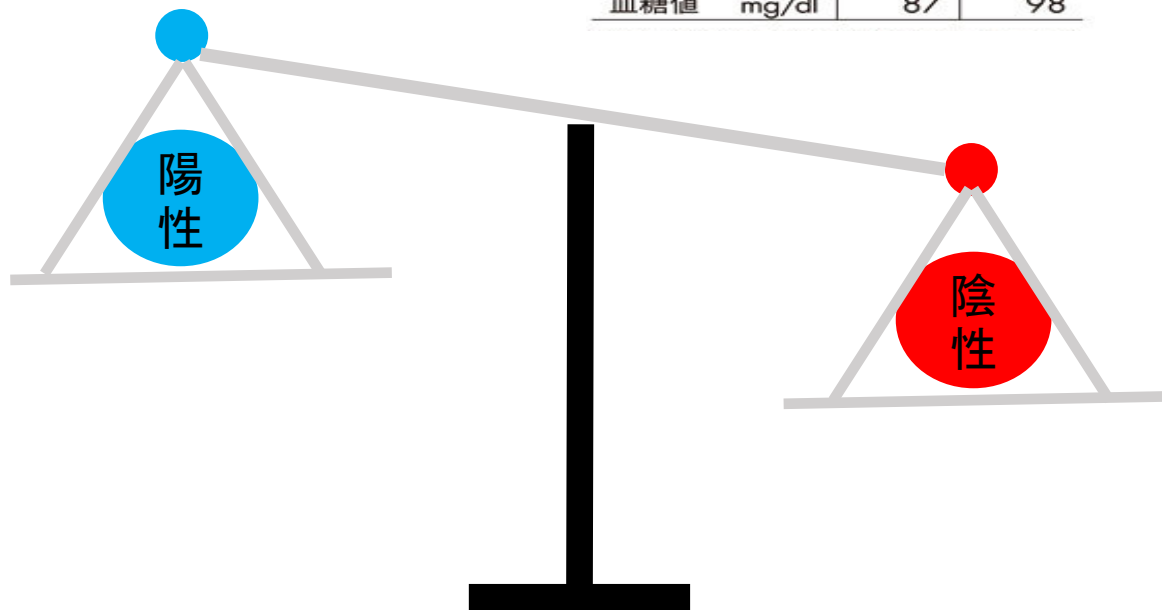
# 機械学習で問われる3つの質問

- 質問 1  
「AかBか？」

識別アルゴリズム

ロジスティック回帰分析

		09年	→12年
身長	cm	170.5	170.5
体重	kg	68.9	65.5
腹囲	cm	92.5	83.5
中性脂肪	mg/dl	296	226
LDL	mg/dl	159	165
HDL	mg/dl	43	43
γ-GTP	IU/l	41	28
血糖値	mg/dl	87	98





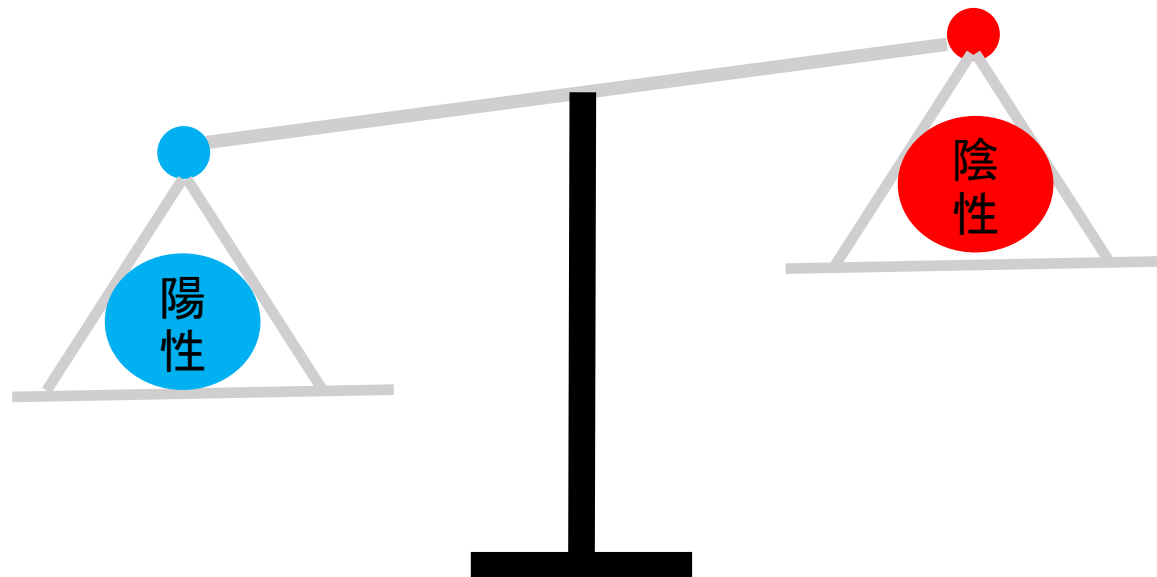
# 機械学習で問われる3つの質問

- 質問 1  
「AかBか？」

識別アルゴリズム

ロジスティック回帰分析

		09年	→12年
身長	cm	170.5	170.5
体重	kg	68.9	65.5
腹囲	cm	92.5	83.5
中性脂肪	mg/dl	296	226
LDL	mg/dl	159	165
HDL	mg/dl	43	43
γ-GTP	IU/l	41	28
血糖値	mg/dl	87	98

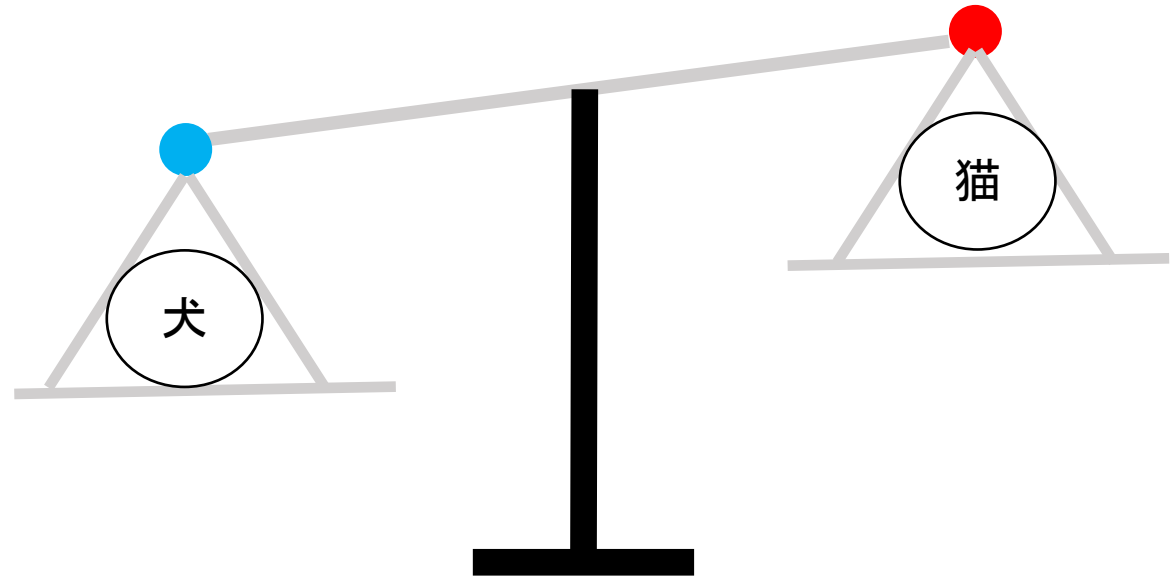


# 機械学習で問われる 3つの質問

- 質問 1  
「AかBか？」

識別アルゴリズム

Deep learning

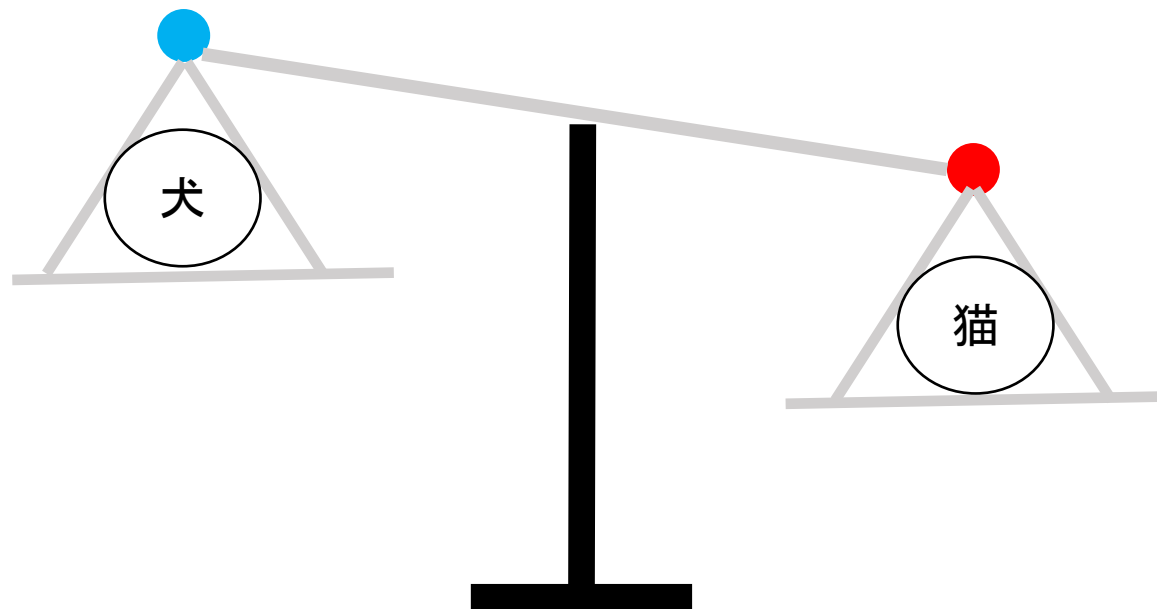


# 機械学習で問われる3つの質問

- 質問 1  
「AかBか？」

識別アルゴリズム

Deep learning



# 機械学習で問われる 3つの質問

- 質問 2

「どのくらいの量または数か？」

回帰アルゴリズム

次の火曜日の気温は何度か？

月曜日



32度

火曜日

何度？

# 機械学習で問われる 3つの質問

## ・質問 2

「どのくらいの量または数か？」

### 回帰アルゴリズム



新しい物件価格を予測



800万円



2億5千万円

過去のデータ

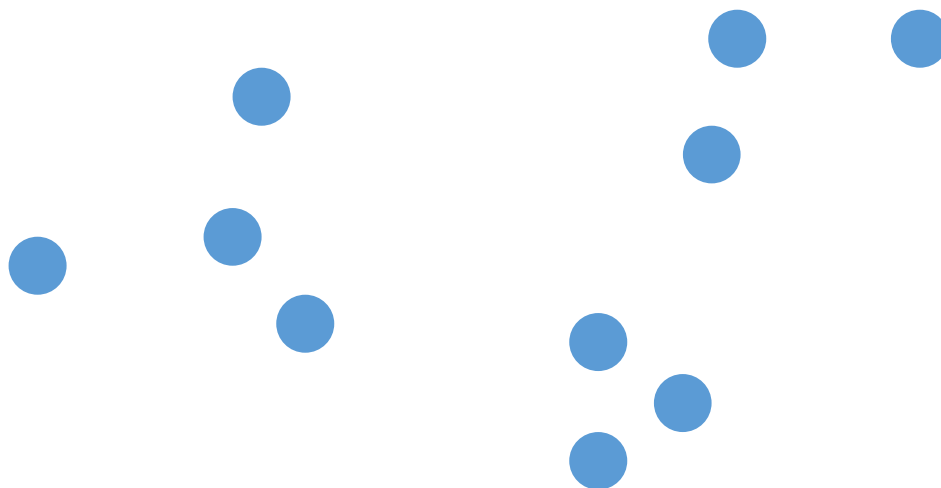
# 機械学習で問われる 3つの質問

- 質問 3

「どのような編成になっているのか？」

分類アルゴリズム

どの視聴者が同じ種類の  
映画を好むか？



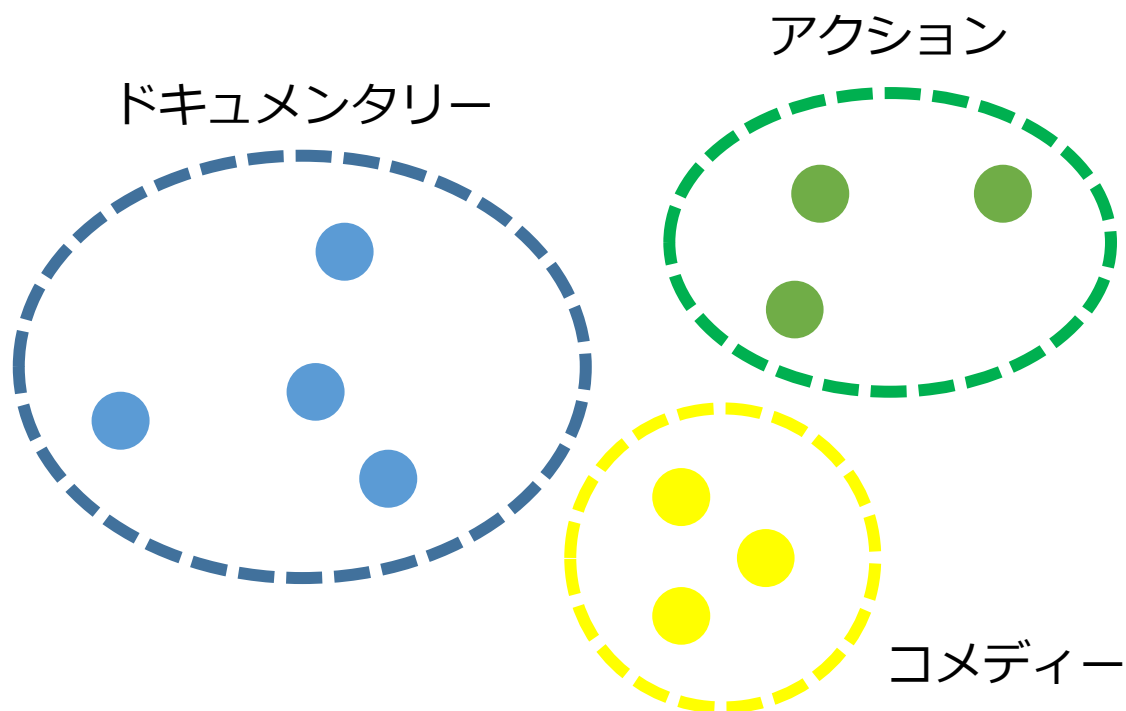
# 機械学習で問われる3つの質問

- 質問3

「どのような編成になっているのか？」

## 分類アルゴリズム

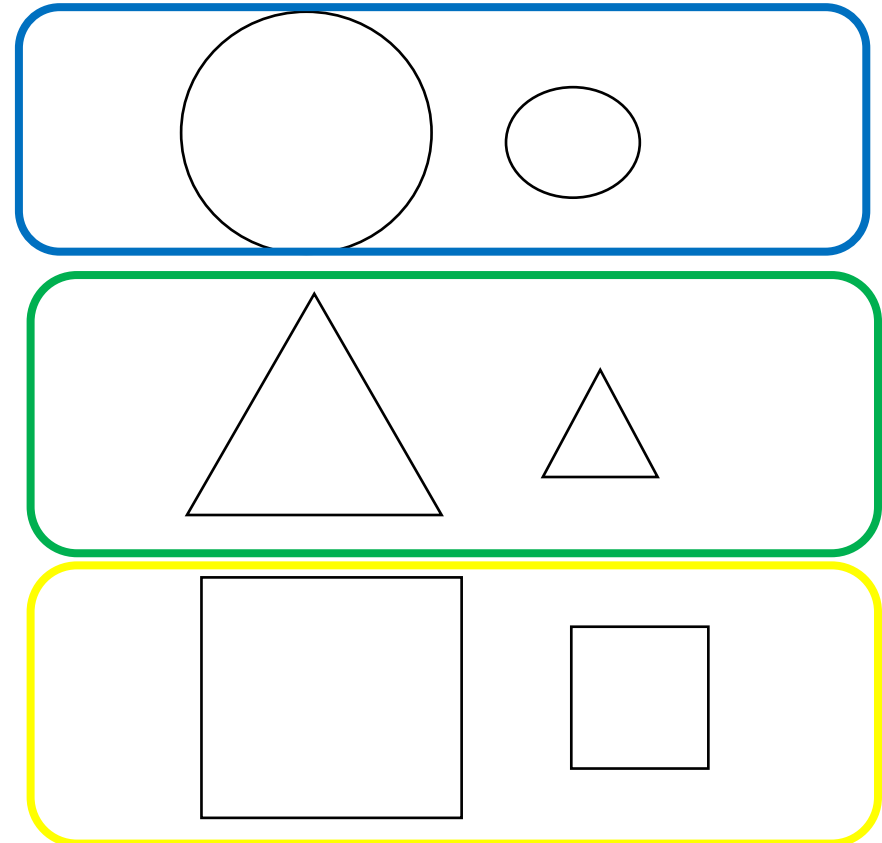
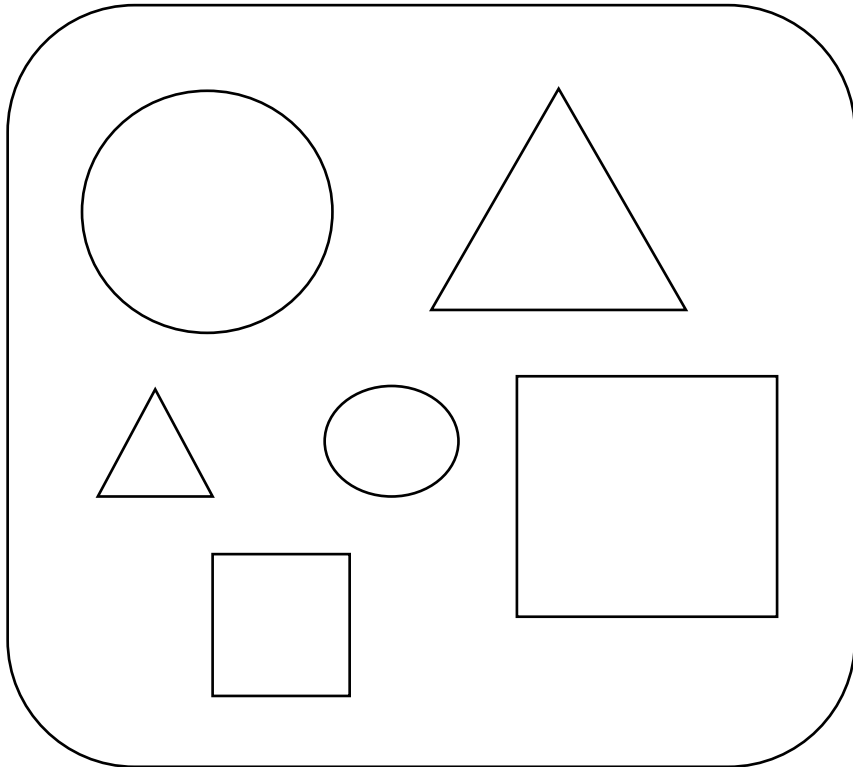
どの視聴者が同じ種類の映画を好むか？



# 機械学習で問われる 3つの質問

- 質問 3

「どのような編成になっているのか？」

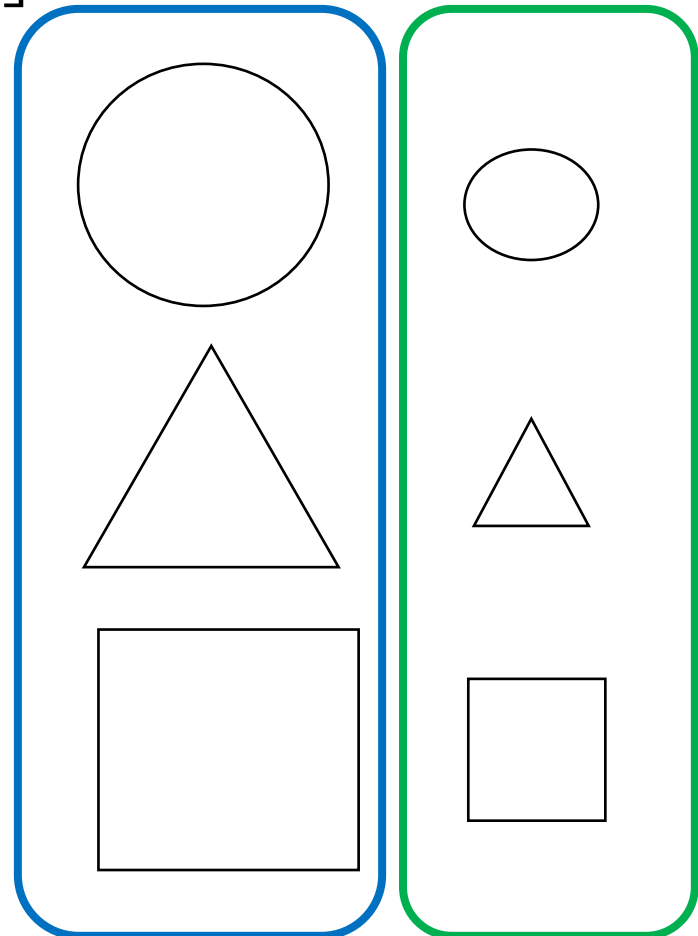
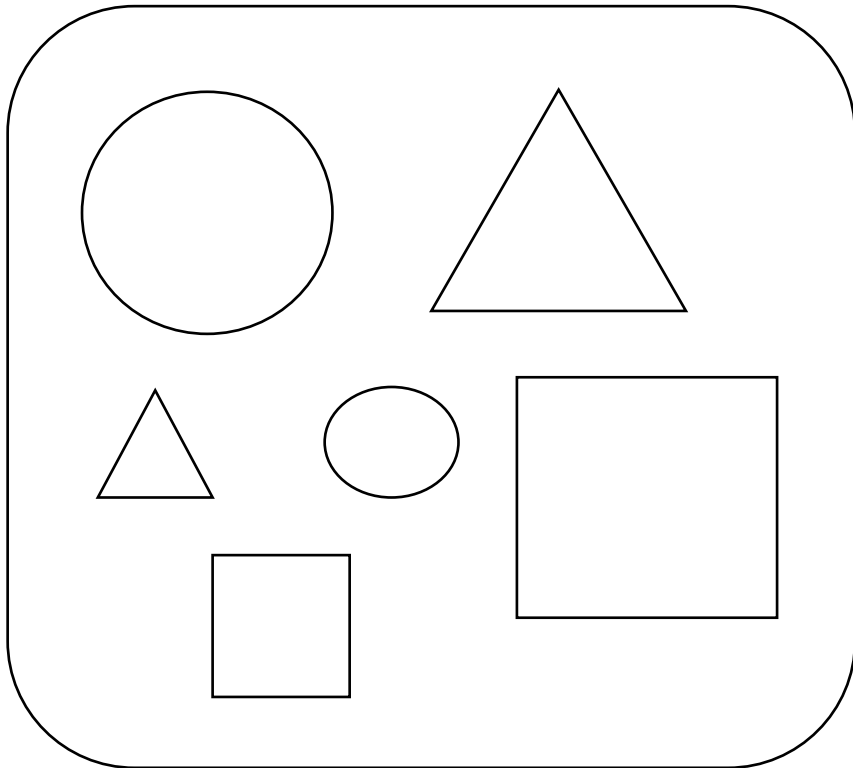




# 機械学習で問われる 3つの質問

- 質問 3

「どのような編成になっているのか？」



# 機械学習の区分

## 機械学習

教師あり

### 識別

AかBか

決定木



ナイーブベイズ



ニューラル  
ネットワーク



SVM



ロジスティック回帰



教師あり

### 回帰

どのくらいの量か

回帰分析



教師なし

### 分類

どう分けるか

k-means法



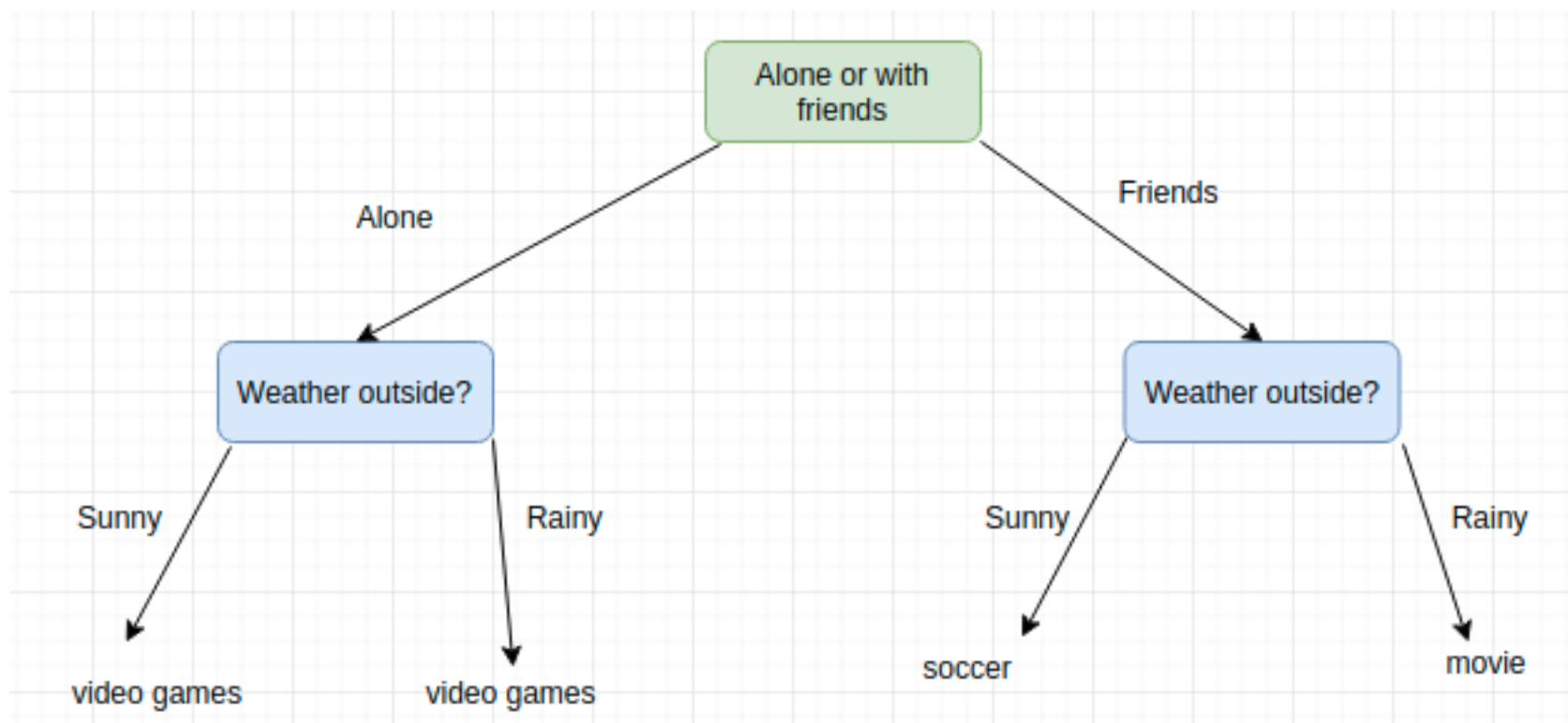
主成分分析



## ・ 教師あり ・ 機械学習 ・ 識別

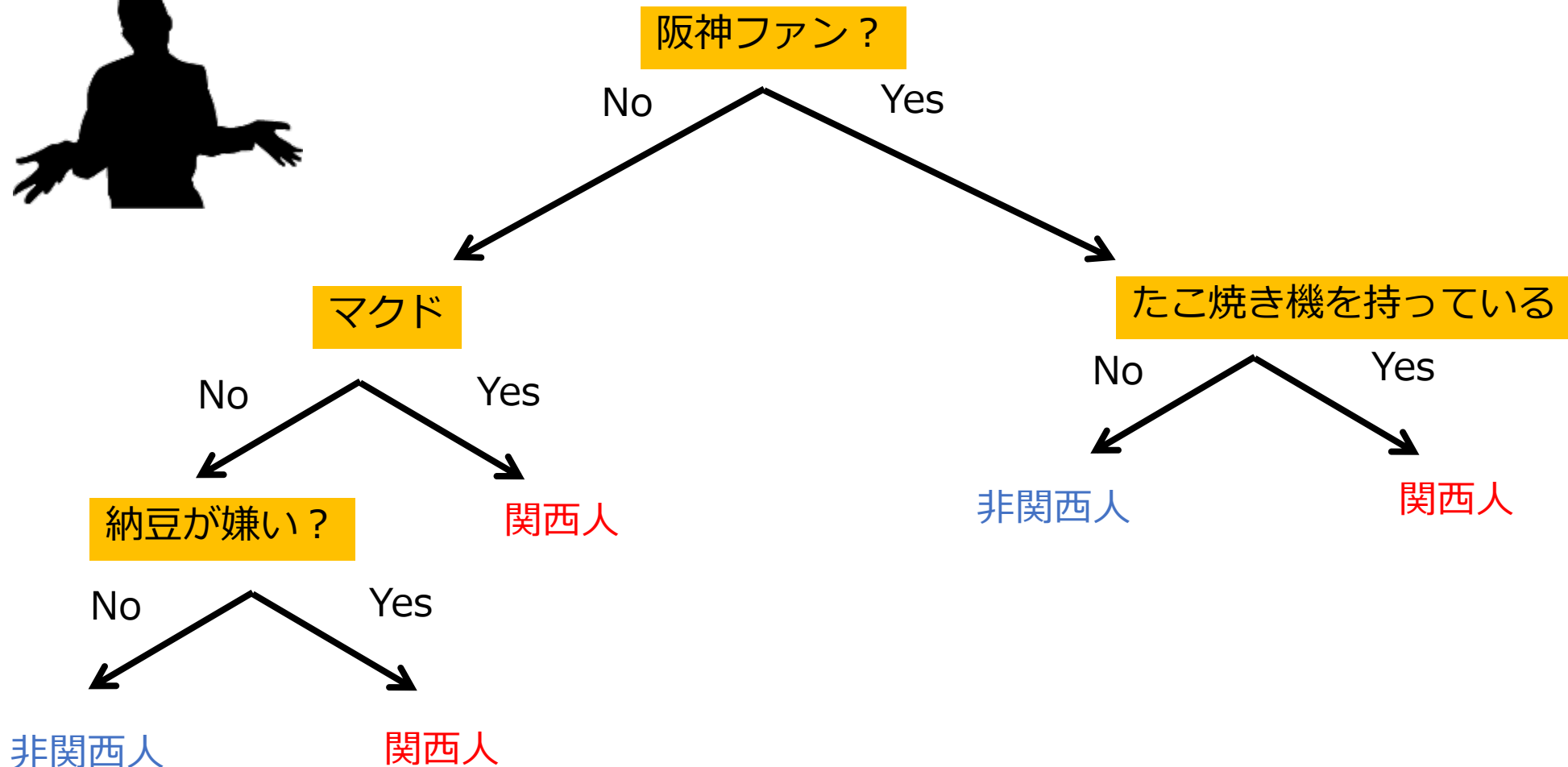
# 決定木

## 樹形図



# 決定木：識別能力の高い質問による分類

関西人なのか？



# 問題「大陸の識別」

country	continent	lifeExp	pop	gdpPercap
Argentina	Americas	75.32	40301927	12779.3796
Canada	Americas	80.653	33390141	36319.235
Cote d'Ivoire	Africa	48.328	18013409	1544.75011
Cuba	Americas	78.273	11416987	8948.10292
-----	-----	-----	-----	-----
Mauritania	Africa	62.664	3270065	1803.1515
Belgium	Europe	79.441	10392226	33692.6051

# ざっくり分けるなら

## 機械学習

### 識別

AかBか

決定木



ナイーブベイズ



ニューラル  
ネットワーク



SVM



ロジスティック回帰



### 回帰

どのくらいの量か

重回帰分析



### 分類

どう分けるか

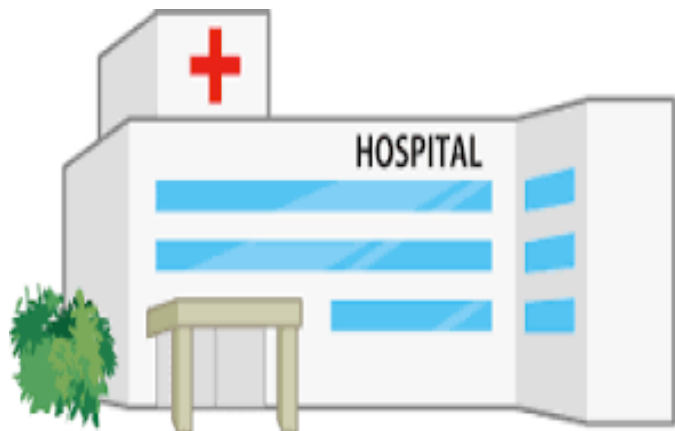
k-means法



主成分分析



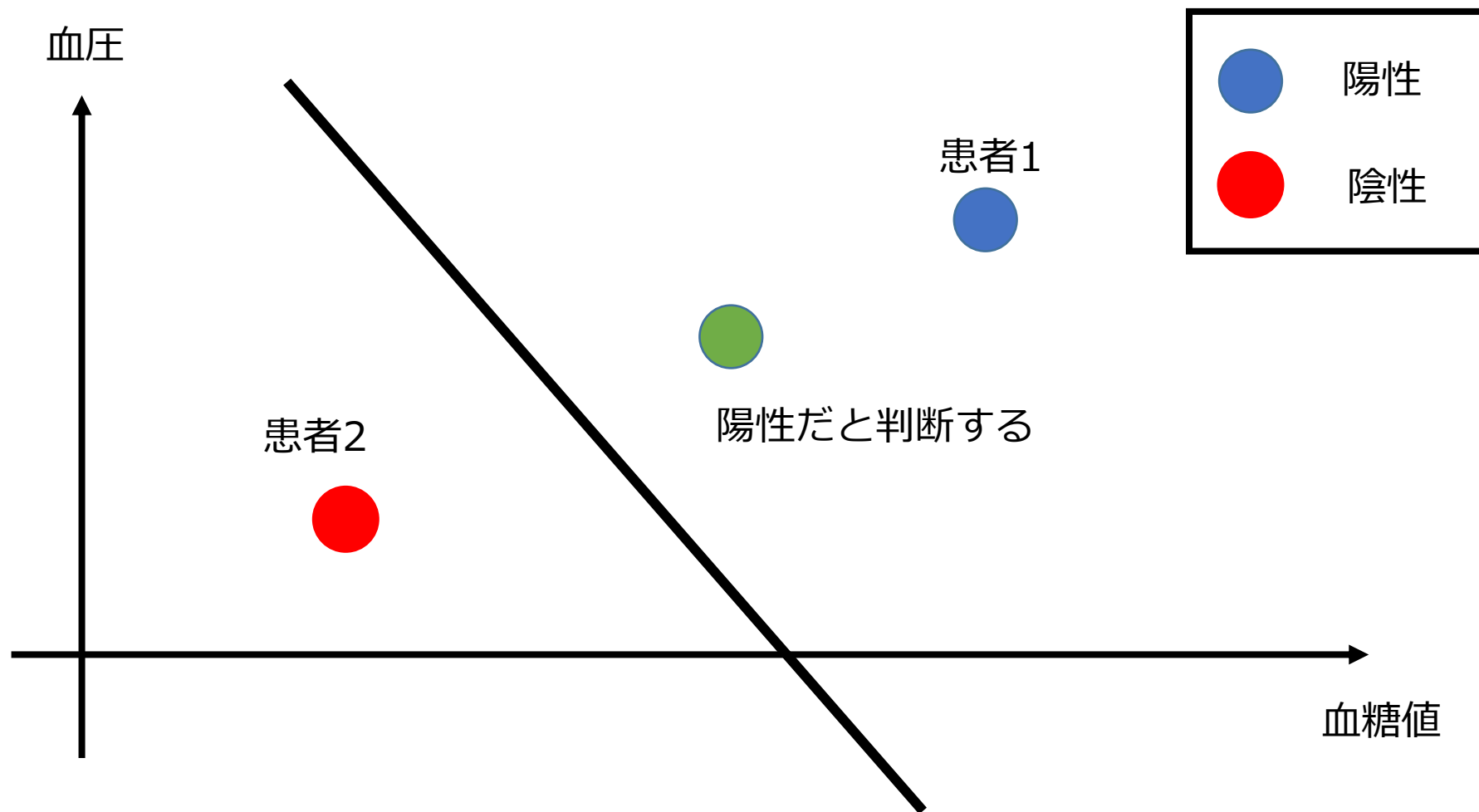
# 陽性・陰性？



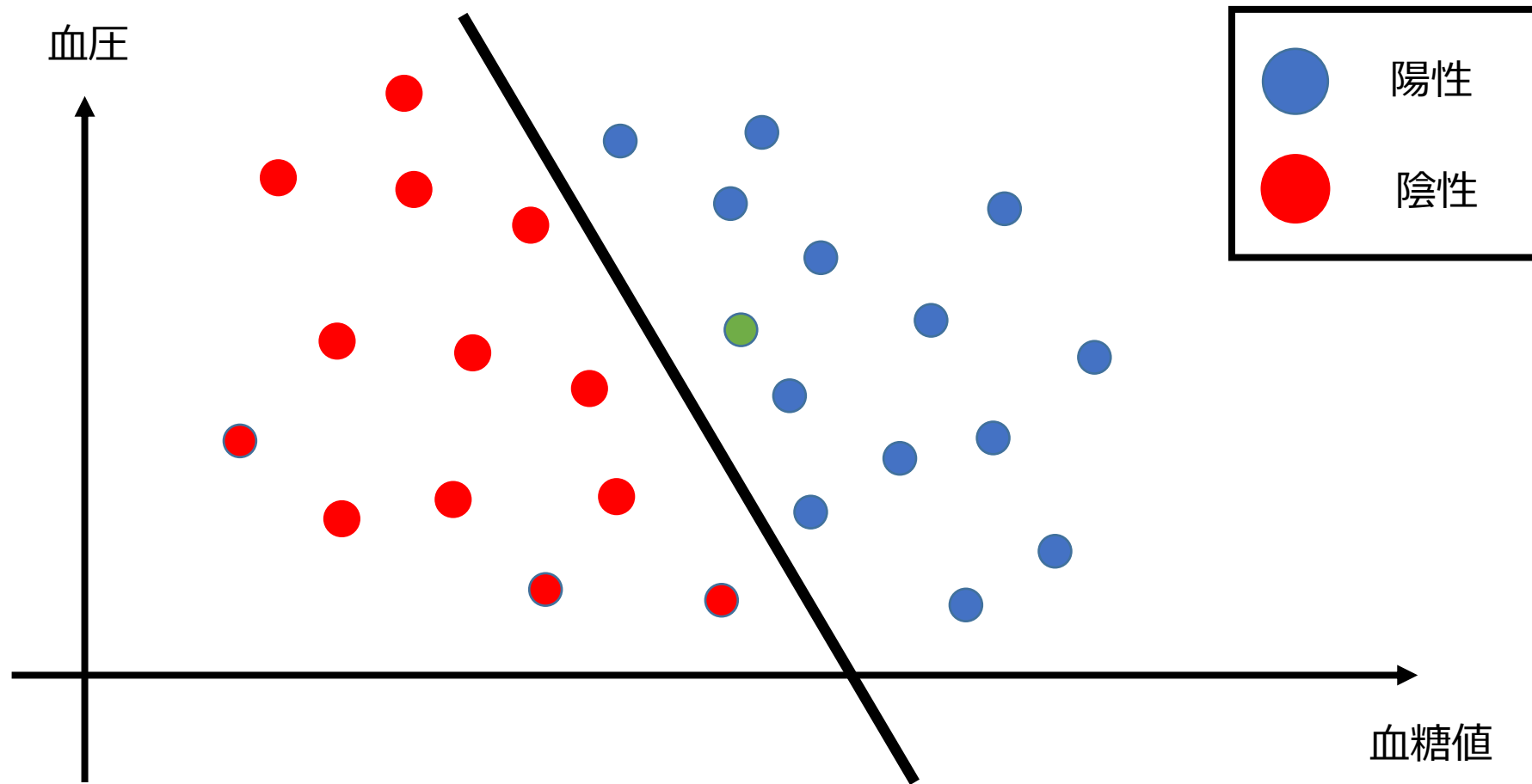
患者ID	血圧	血糖値	
1	120	200	陽性
2	80	130	陰性
3	110	190	?



# 陽性・陰性を識別する

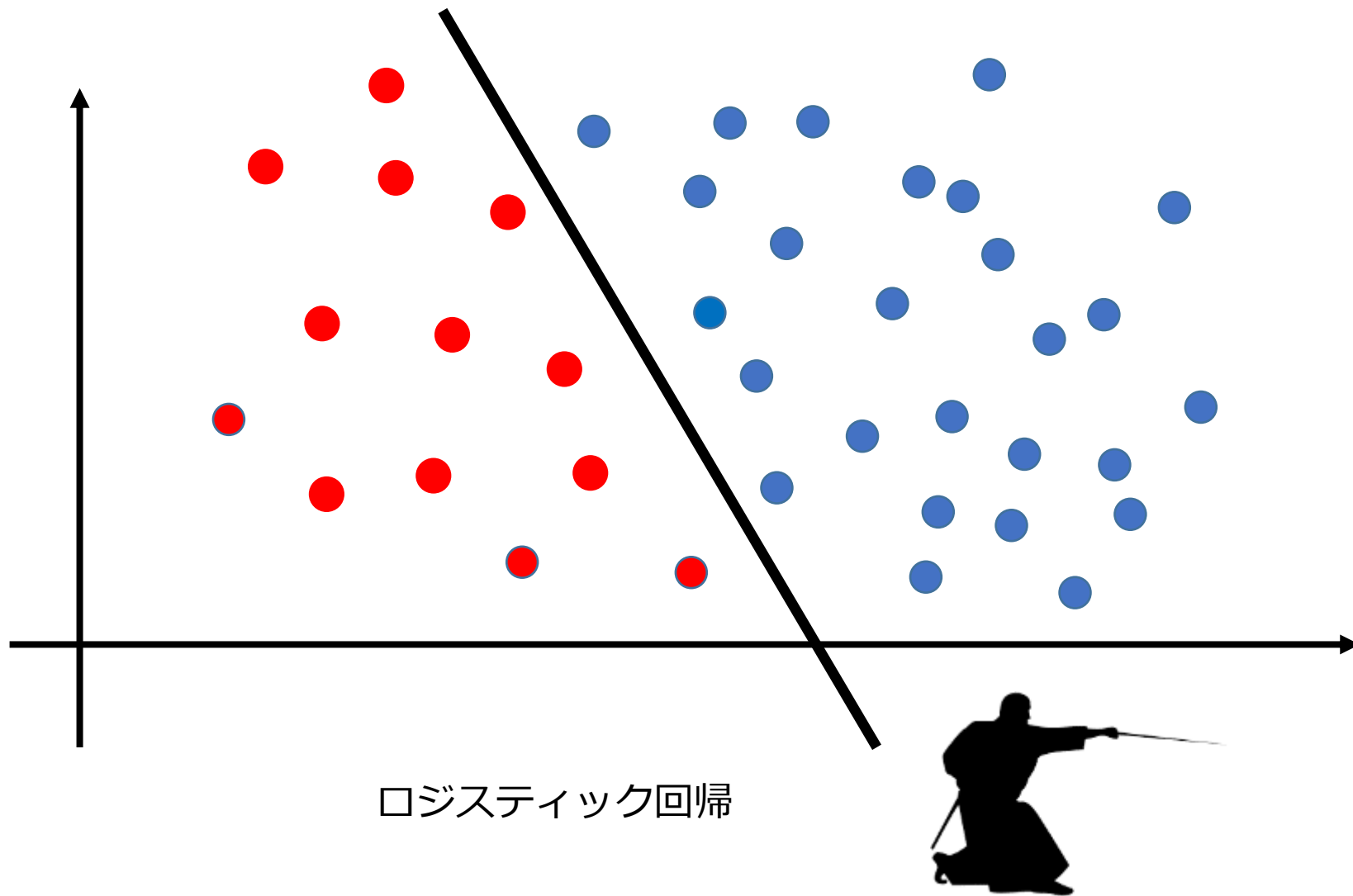


# 陽性・陰性を識別する

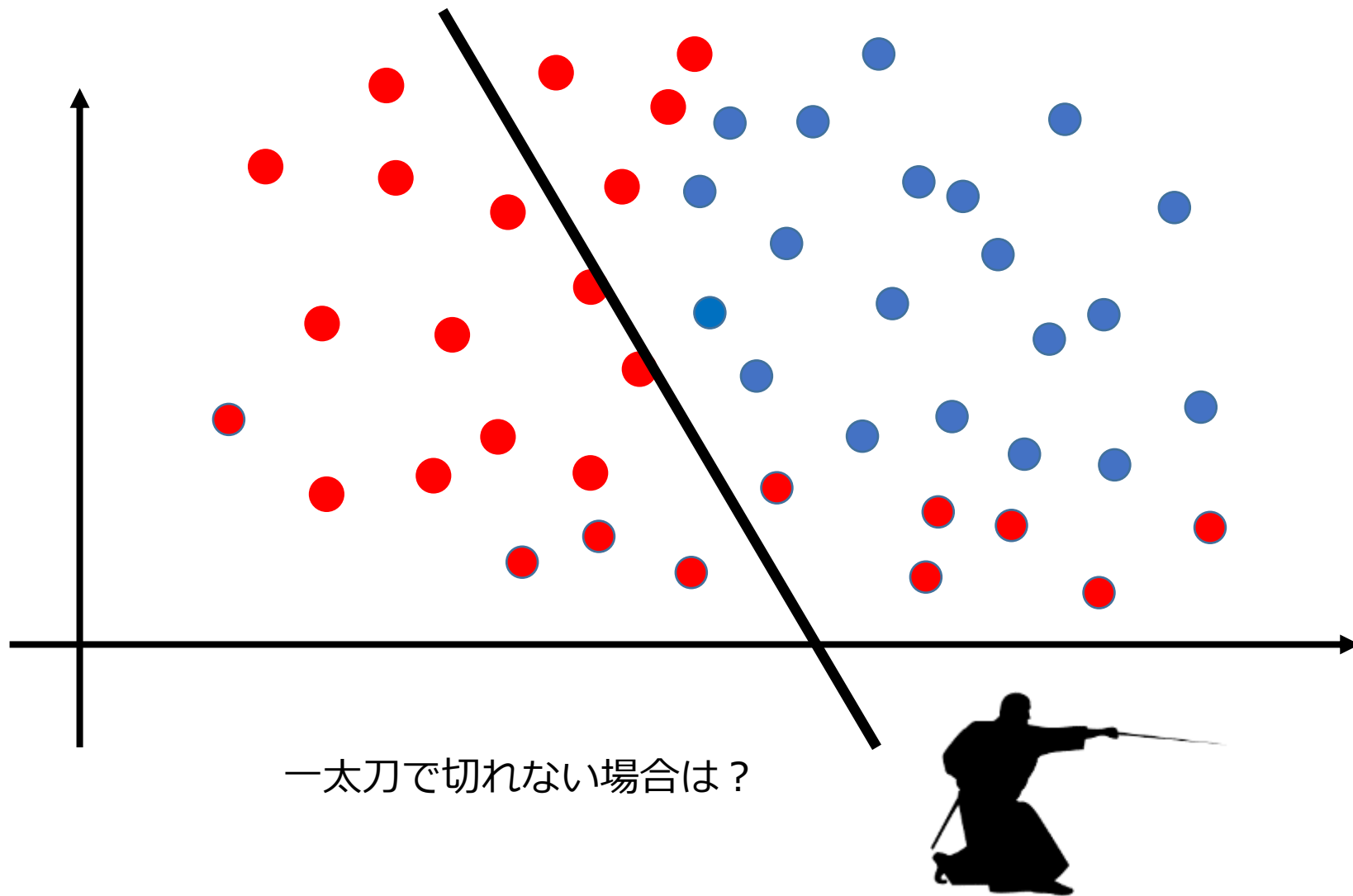


ロジスティック回帰

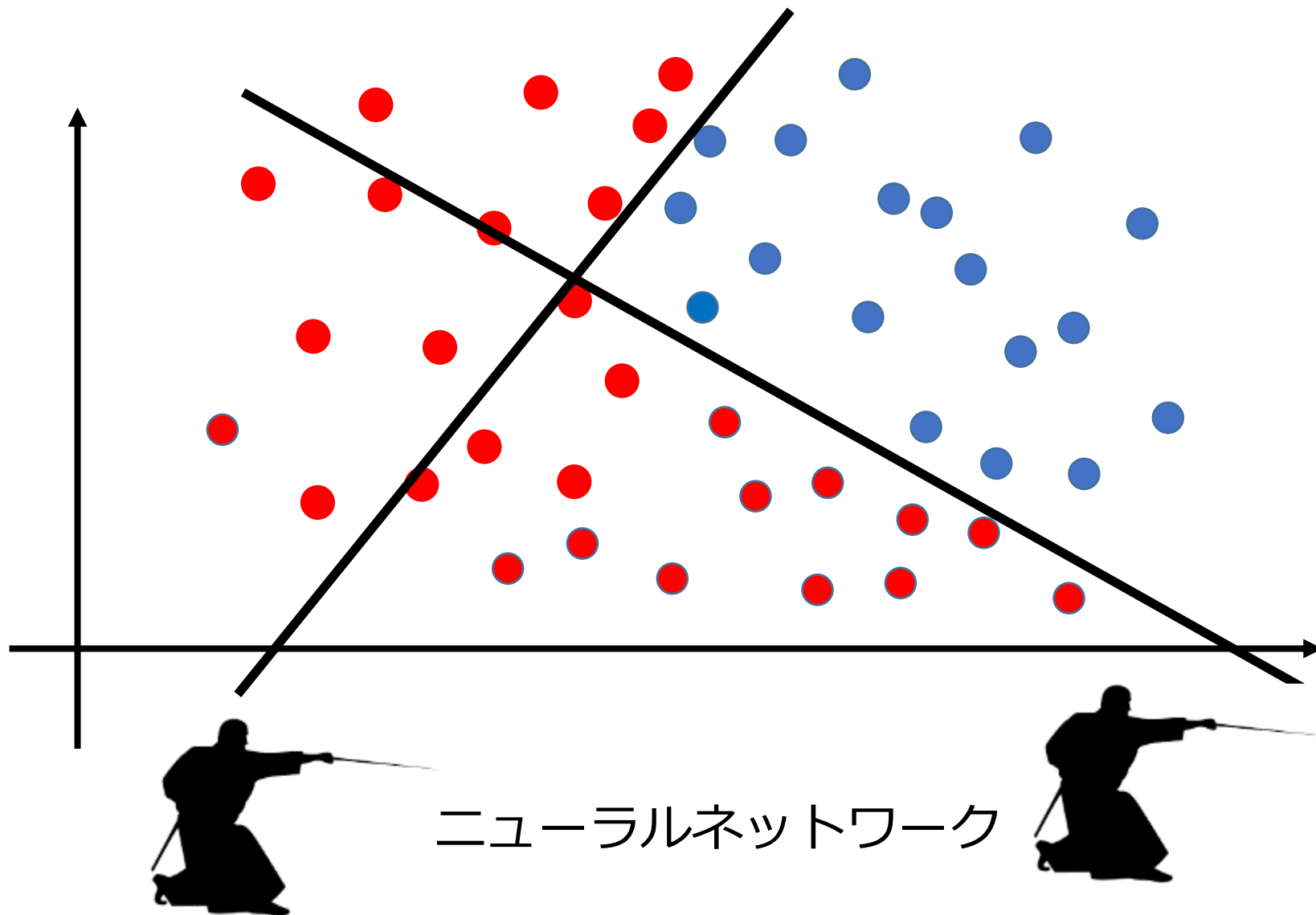
# ロジスティック回帰分析



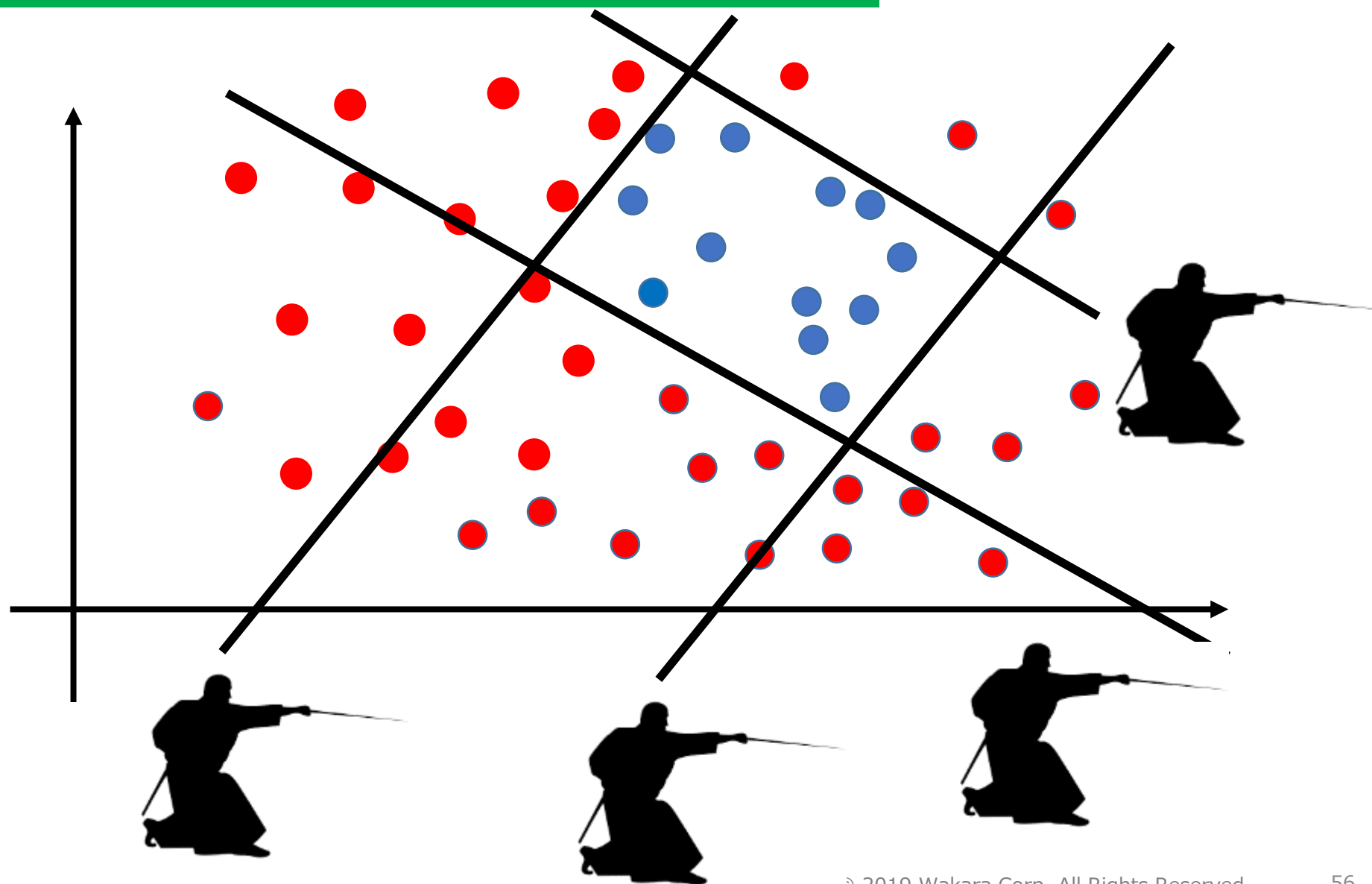
# いろいろな識別方法



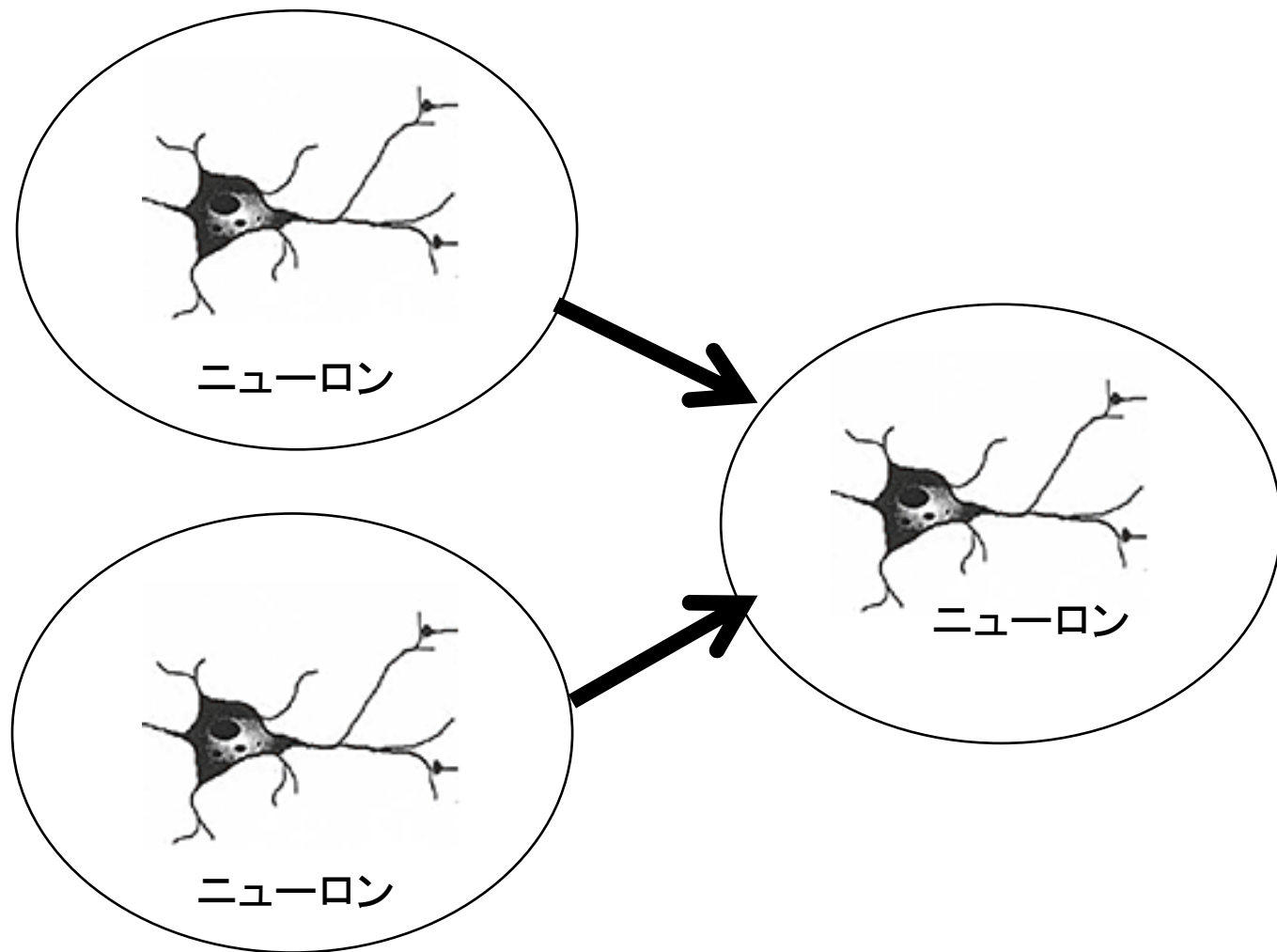
# いろいろな識別方法



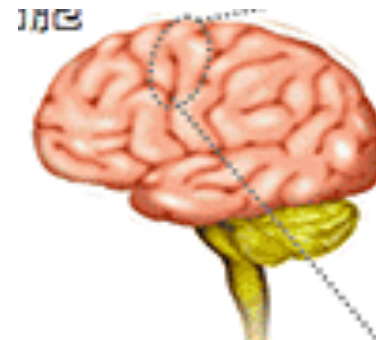
# いろいろな識別方法

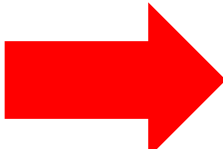


# ニューラルネットワーク

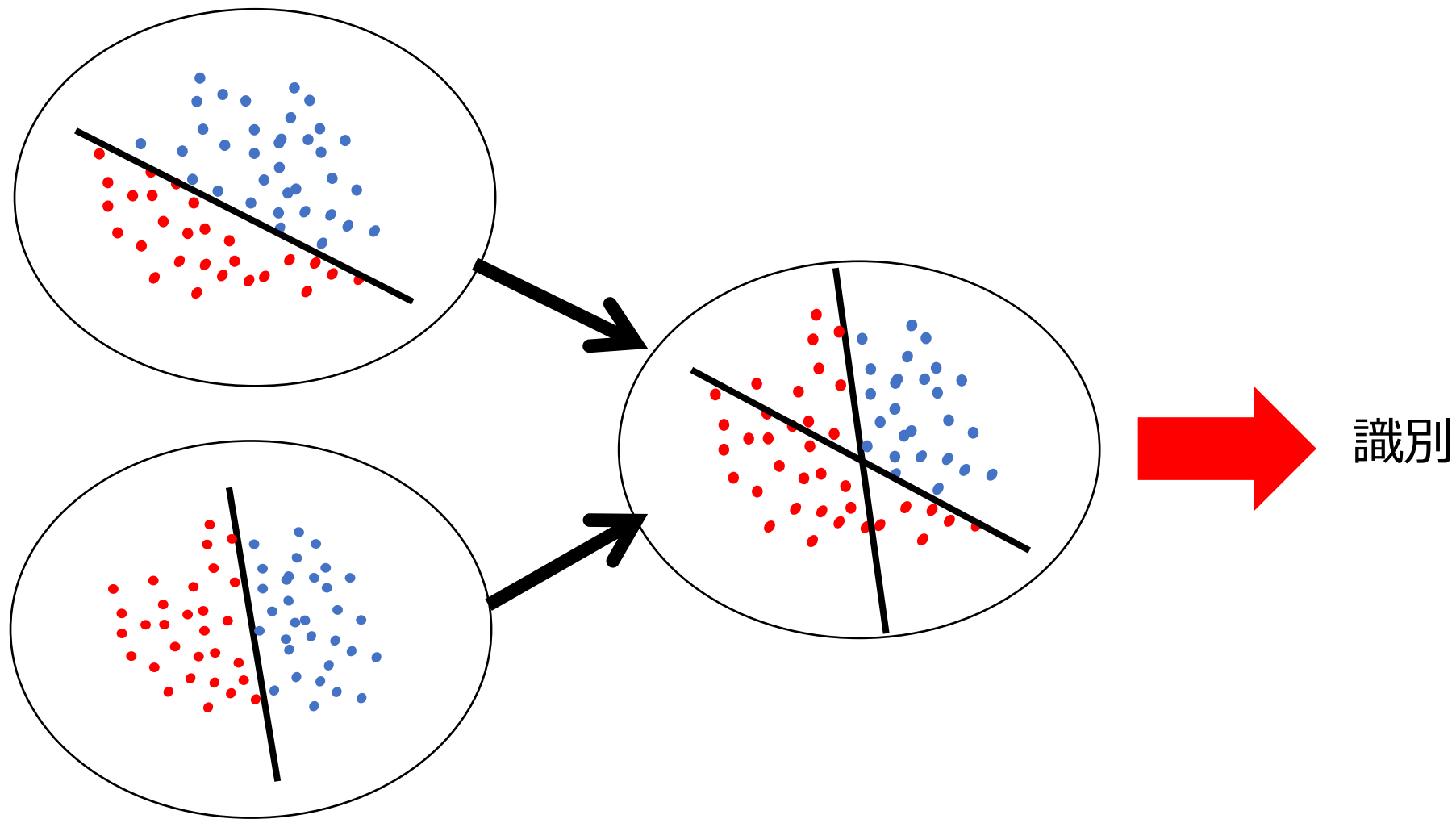


人間の脳はどやって識別  
を行なっているのか？



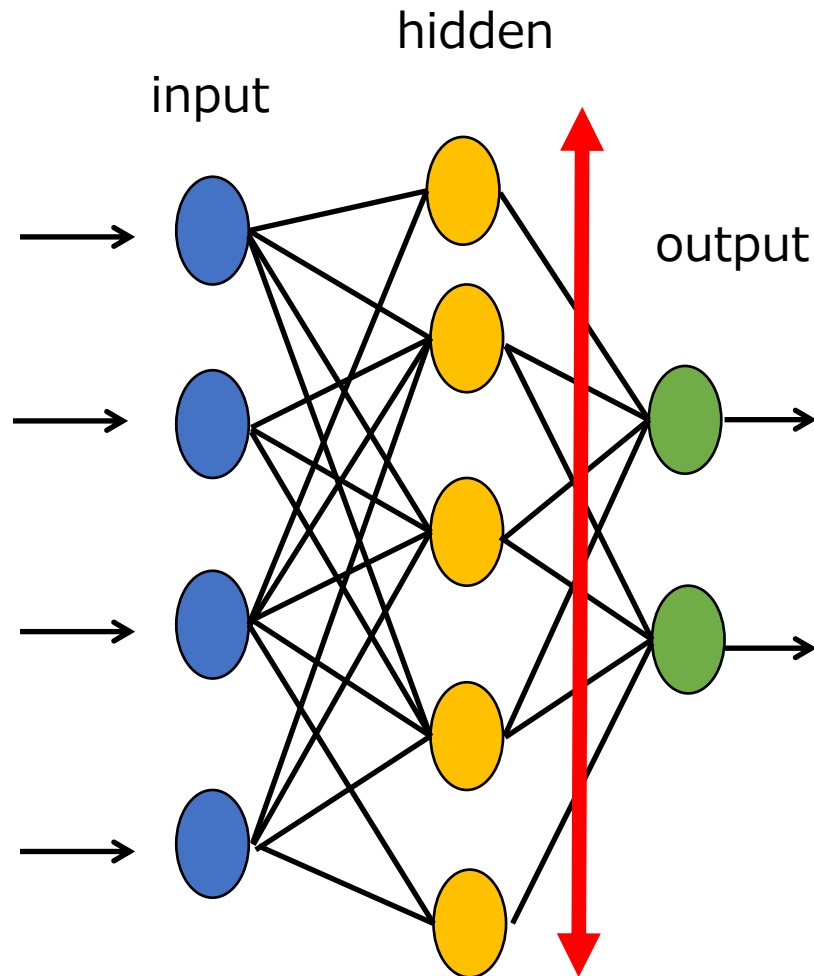
 識別

# ニューラルネットワーク

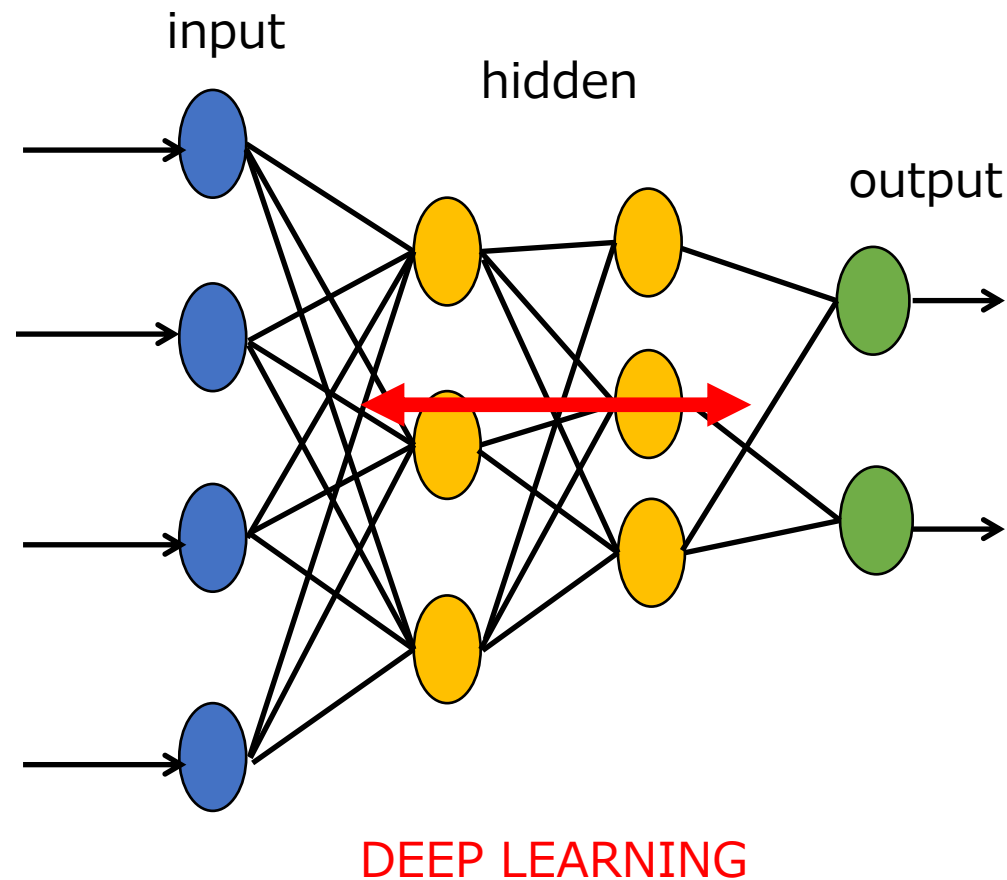




# ニューラルネットワークとディープラーニング



ニューラルネットワーク



## ・ 教師あり ・ 機械学習 ・ 回帰

# 機械学習の区分

## 機械学習

### 識別

AかBか

決定木



ナイーブベイズ



ニューラル  
ネットワーク



SVM



ロジスティック回帰



### 回帰

どのくらいの量か

重回帰分析



### 分類

どう分けるか

k-means法



主成分分析



# 回帰分析

- ・ 予測したい変数を様々な要因から予測する方法

⇔ 住宅の価格を築年数・坪数から予測する方法



800万円

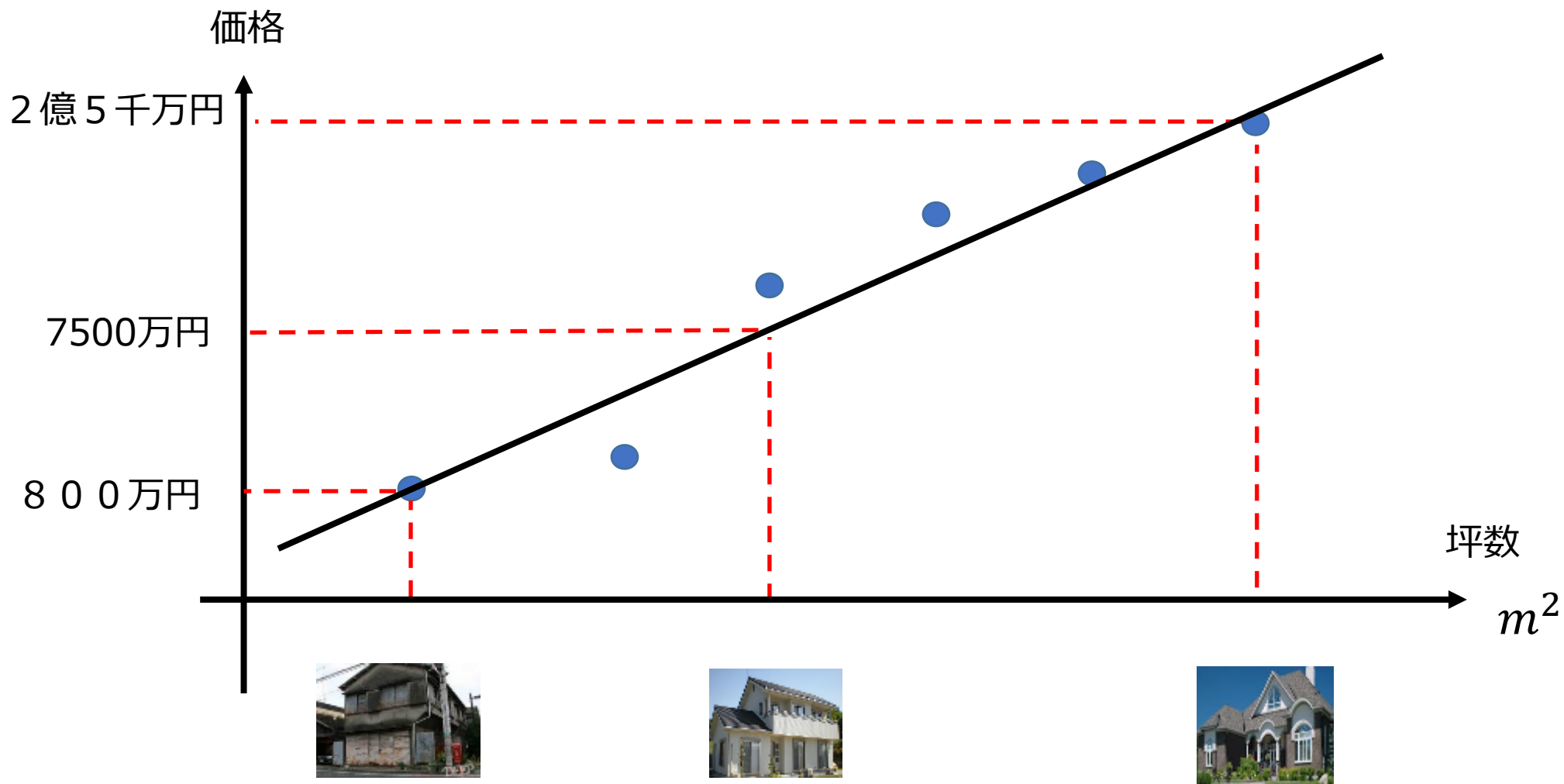


価格を予測する

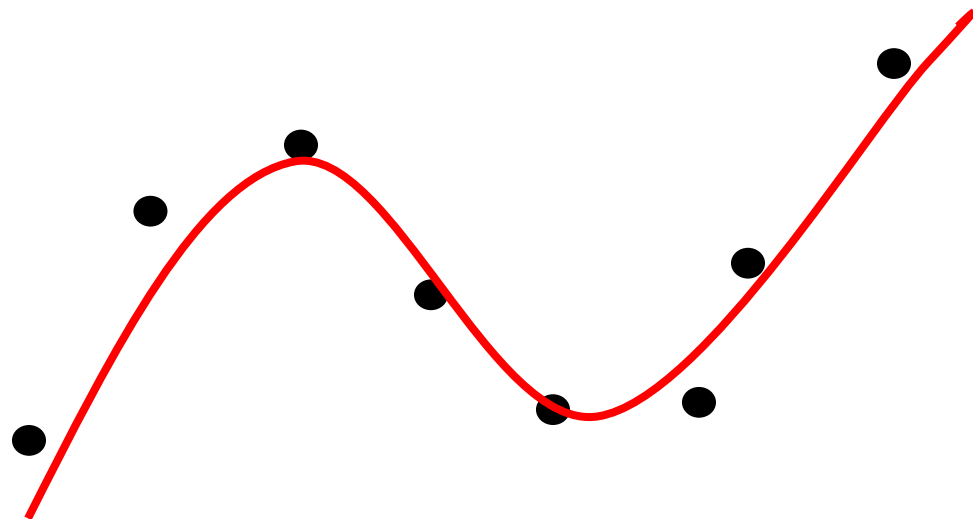
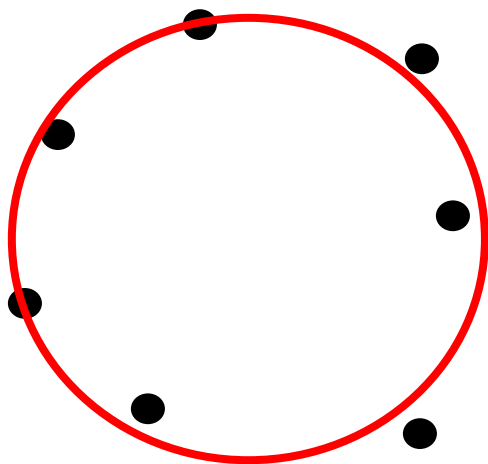


2億5千万円

# 回帰分析



# 様々な回帰モデル



[https://kwichmann.github.io/ml\\_sandbox/linear\\_regression\\_diagnostics/](https://kwichmann.github.io/ml_sandbox/linear_regression_diagnostics/)

# 機械学習の区分

## 機械学習

### 識別

AかBか

決定木



ナイーブベイズ



ニューラル  
ネットワーク



SVM



ロジスティック回帰



### 回帰

どのくらいの量か

重回帰分析



### 分類

どう分けるか

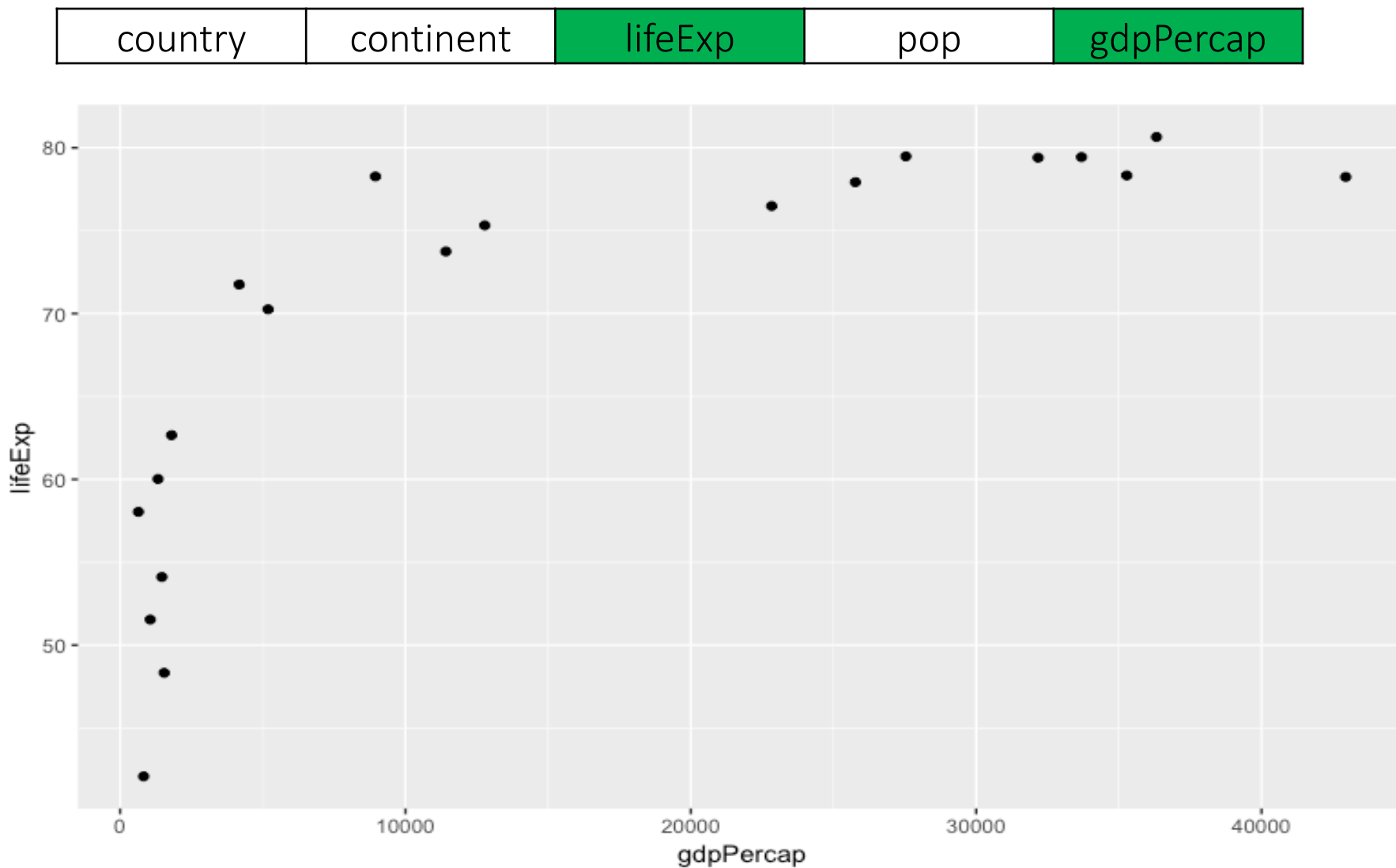
k-means法



主成分分析

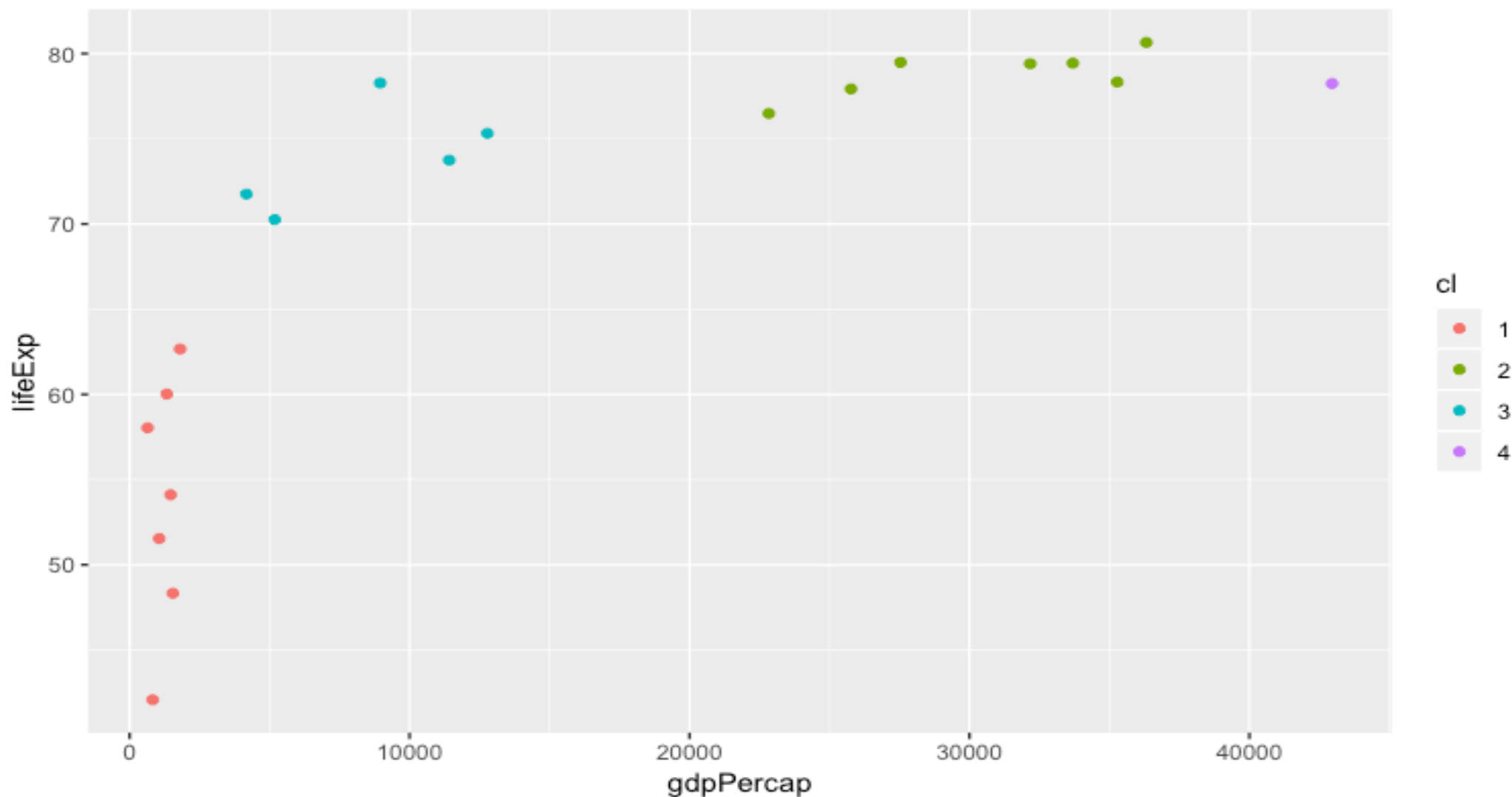


# データの可視化





# 4つのクラスターに分類



# 4つのクラスターに分類

