

# 計算知能 (COMPUTATIONAL INTELLIGENCE)

第 1 3 回 決定木  
教員： 谷口彰

# 第13回 決定木

- 決定木の学習
- 情報理論の利用
- 決定木学習の現実応用

# 様々な学習

いろいろ  
出来るよ！



Explainable AI

思考過程が  
説明できます！



## ■ ニューラルネットワーク

非線形関数 $y = f(x; w)$ の数値パラメータ $w$ の値を調整

関数近似

## ■ 強化学習

状態 $s$ で行動 $a$ をとるときの報酬の推定値 $Q(s, a)$ を調整

非線形方程式

数値の学習

## ■ 決定木

属性値から概念を識別するための木構造を生成

## ■ 説明に基づく学習

将来の探索を削減するための効率良いルールを生成

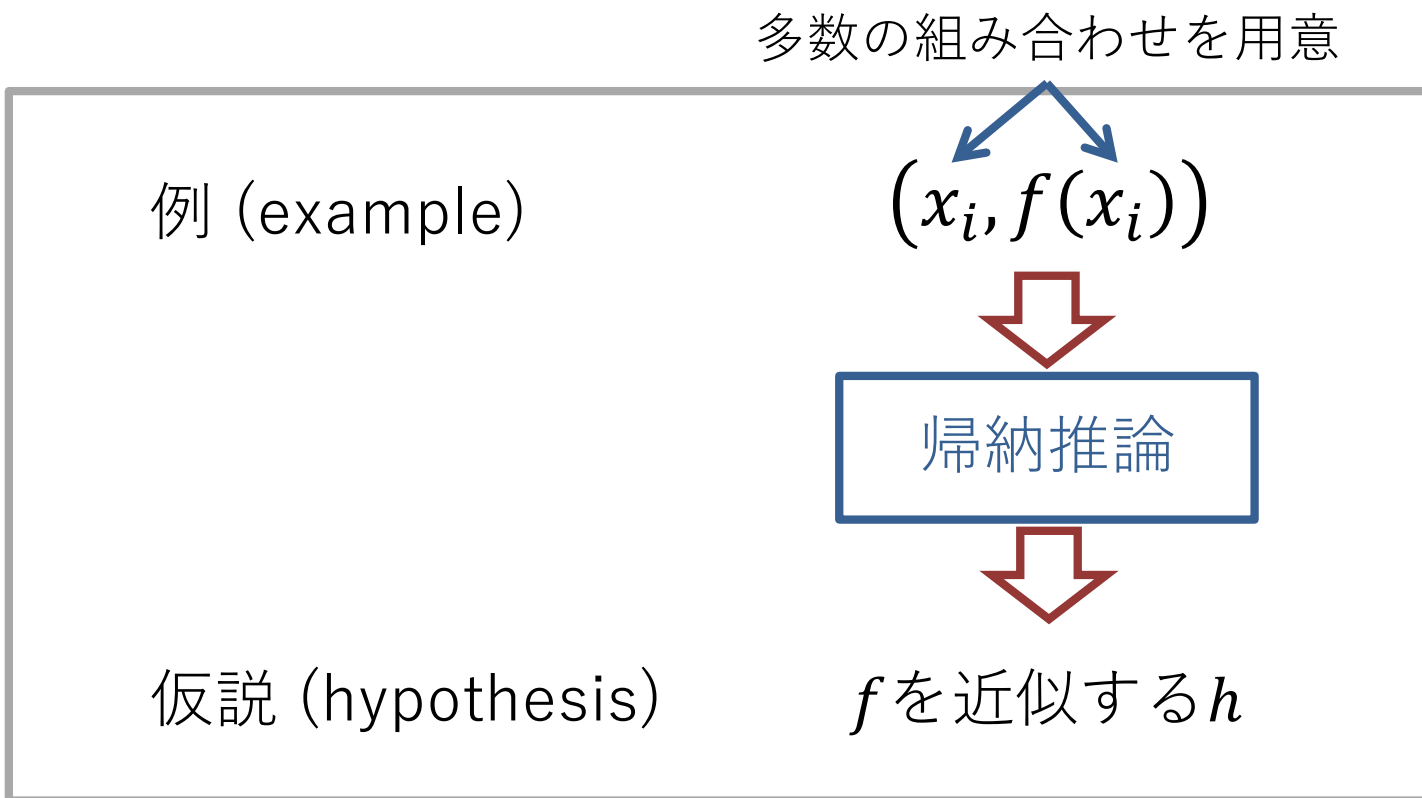
## ■ 帰納論理プログラミング

与えられた入出力関係を満たす数学的機能パターンを生成

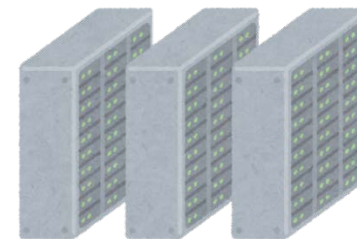
構造の学習

# 帰納推論

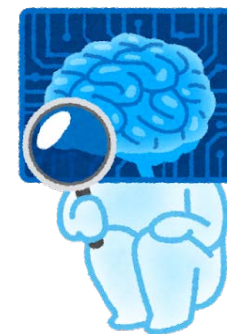
- 多くの学習アルゴリズムは帰納推論の形式をとる



事例の獲得



仮説の形成



仮説の評価



- 複数の仮説の中から、ある仮説を他の仮説より優先して選ぶ基準を設定して適切な $h$ を出力する

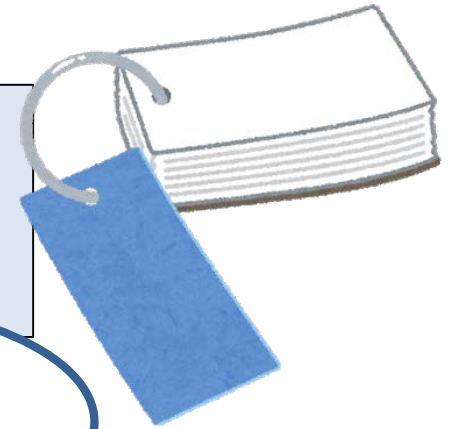
# 丸暗記学習アルゴリズム（単純な推論）

```
学習(x, y){  
    (x, y)をデータベースDBに記録する  
}
```

```
予測(x){  
    if(DBに(x)が記録されている)  
        return y;  
    else{  
        DB中のデータから多数決などでyを決める  
        return y  
    }  
}
```

これに勝てない  
アルゴリズムも  
存在

未知のxに対して出力yを予測する  
目的にはほぼ役に立たない



# 決定木(1/3)

- 構造を学習するアルゴリズムの中でもっとも一般的な学習法
- 決定木の目的は、属性とその値の組  
{属性1=値1, ..., 属性n=値n} によって表現されたデータを  
いくつかのクラスと呼ぶものに分類すること

例：花

{色=ピンク、花卉の数=5} 等のデータを用いて花の種類を特定



# 決定木(2/3)

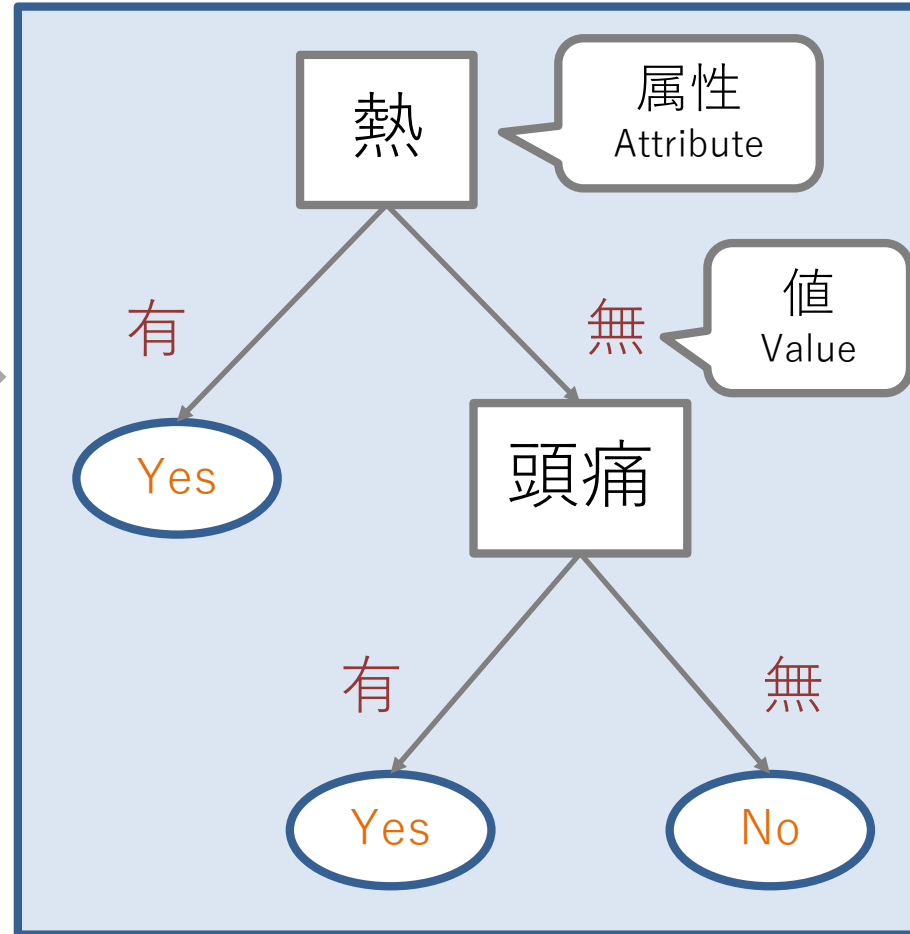


決定木：

質問：  
現在の症状について

熱 = 無  
頭痛 = 有

INPUT



OUTPUT

判断：  
病院に行くか否か

Yes  
or  
No

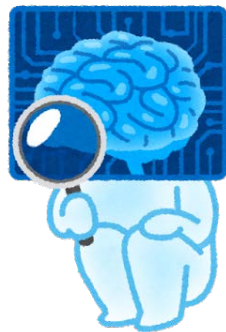
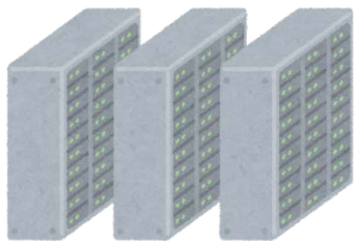


# 決定木(3/3)

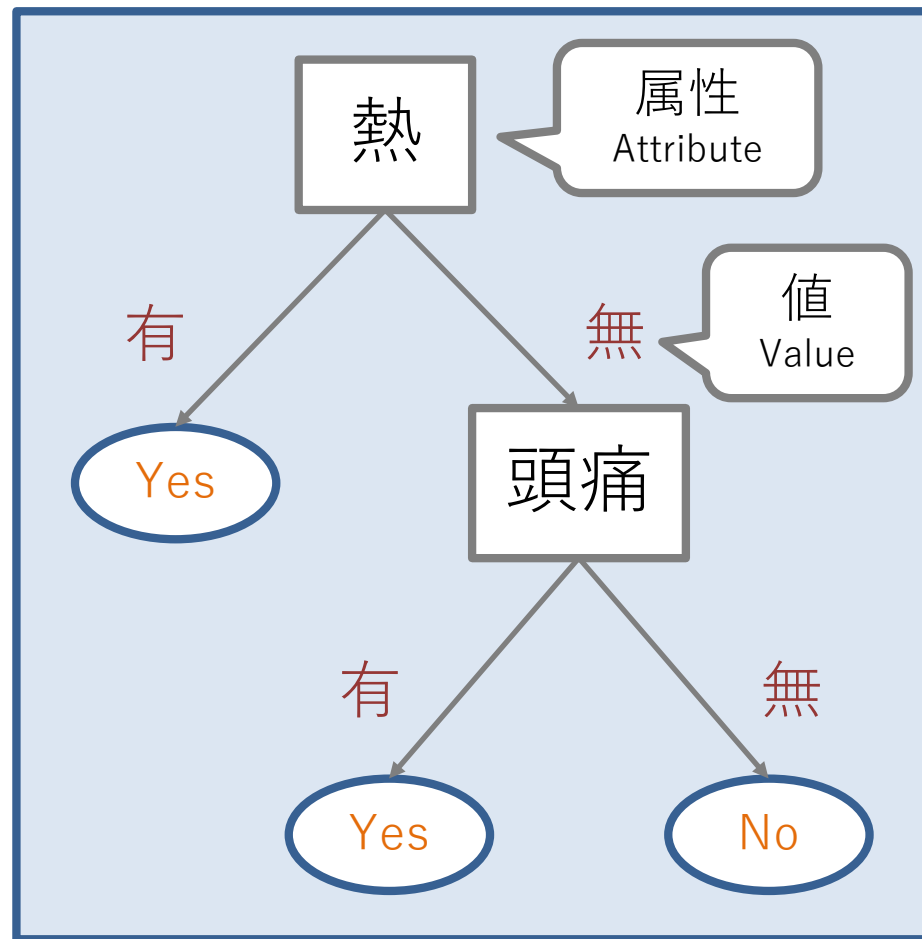
- 決定木の非終端ノードには属性のラベルがつけられ、そこから出ている枝にはその属性の取り得る値がつけられる
- 終端ノードには最終的なクラス分類がつけられる
- すべての属性が揃った {属性1=値1, ..., 属性n=値n} 形式のデータを質問データとすることで、属性の値をテストしながら最終的なクラス分類を出力可能



# 決定木の学習(1/2)



決定木：



訓練例(training set)

{熱=無, 頭痛=有} : Yes  
{熱=無, 頭痛=無} : No  
:

学習  
アルゴリズム

# 決定木の学習(2/2)

- 決定木の学習は、具体的な判断事例から決定木を生成すること

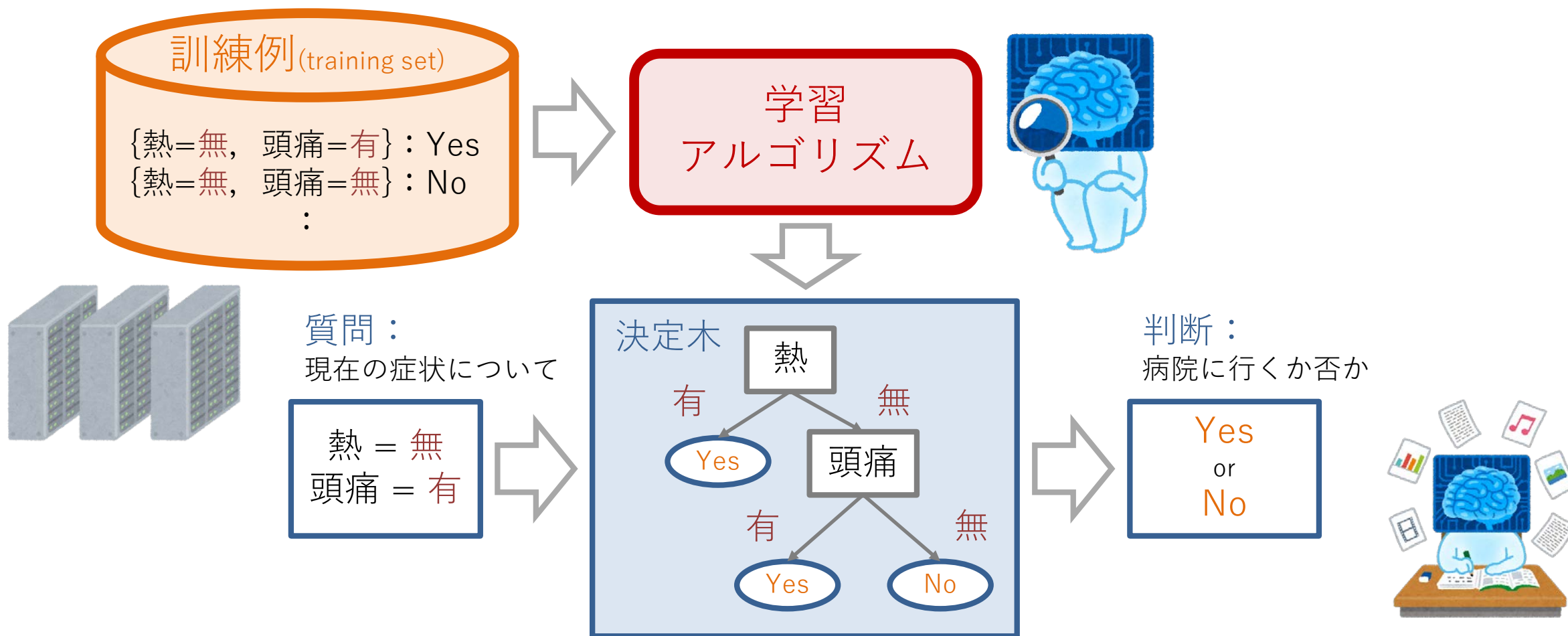
## 事例

入力 = {属性1=値1, ..., 属性n=値n} : 出力 = 判別結果のような形式のデータで、入力の属性値に応じた判別結果を指定

- このようなデータの集まりを、学習アルゴリズムに対する訓練例という

# 学習システム(1/2)

## 学習・運用システムの全体像



# 学習システム(2/2)

- システム設計者は、事前に学習アルゴリズムに訓練例を与えて、決定木を生成
- システム運用時には、質問データ  
{属性1=値1, ..., 属性n=値n} が与えられると、決定木を使用して分類結果を出力

# 訓練例

## ■ 薬物事件の犯人の属性

	性別	年齢	薬物	犯罪歴	実刑
A	男性	未成年	覚醒剤	初犯	No
B	女性	未成年	覚醒剤	初犯	No
C	女性	老人	麻薬	初犯	No
D	男性	成年	覚醒剤	初犯	Yes
E	男性	老人	シンナー	累犯	Yes
F	男性	老人	麻薬	初犯	Yes
G	女性	成年	シンナー	累犯	Yes
H	男性	未成年	シンナー	累犯	Yes
I	男性	老人	麻薬	累犯	Yes
J	女性	未成年	麻薬	累犯	Yes

負例

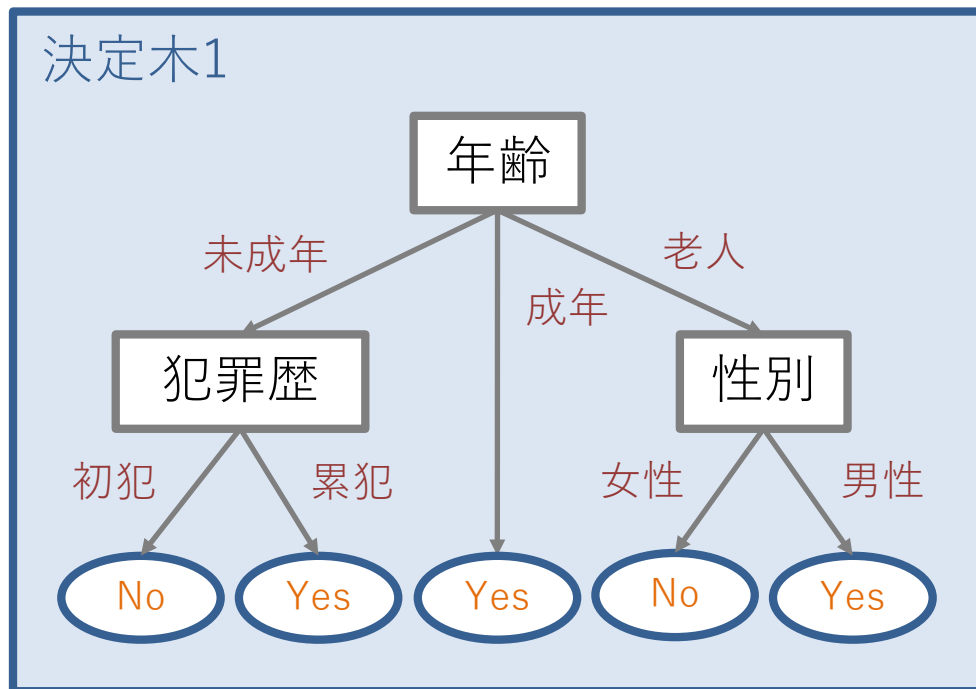
正例

刑務所

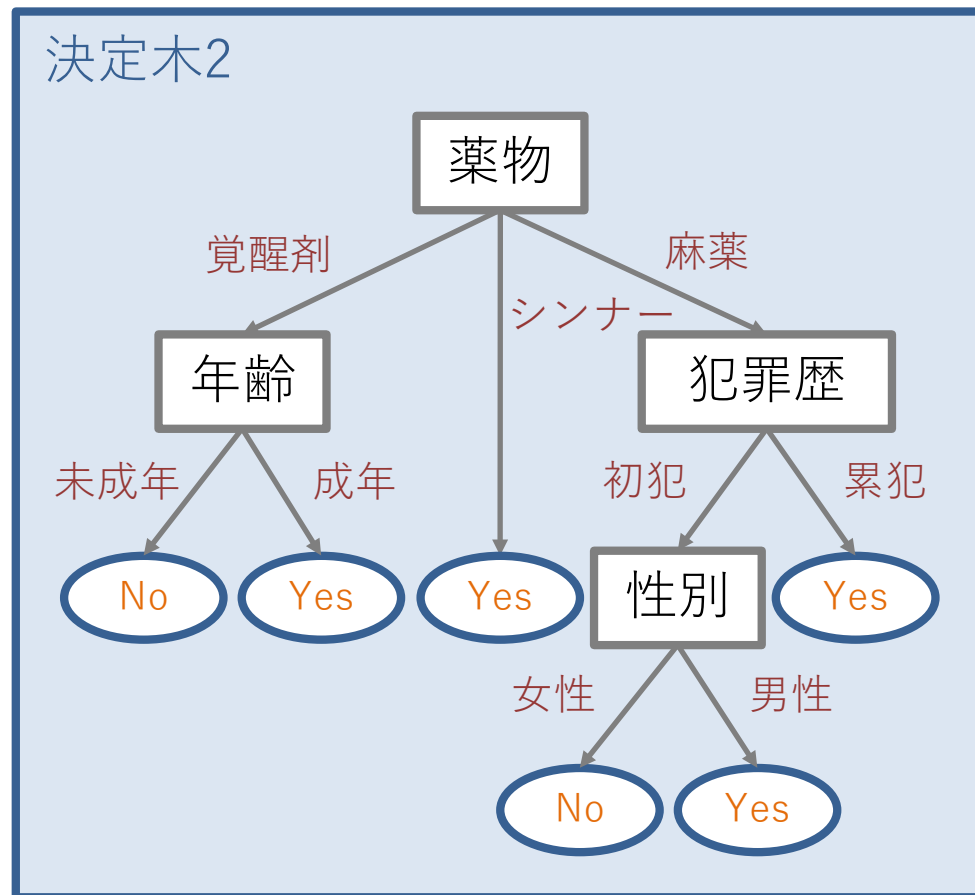


# 複数の決定木(1/2)

決定木1



決定木2



どちらも訓練例を正しく分類している

→ 未知の例を正しく分類できるか

(女性、老人、シンナー、初犯)

→ テストの回数はどのようになるか (効率)

(男性、成年、麻薬、初犯)

## 複数の決定木(2/2)

- 属性をどのような順番でテストするかによって、生成される決定木が異なる

オッカムのかみそり：

「ある事柄を説明するためには、必要以上に多くを仮定するべきでない」

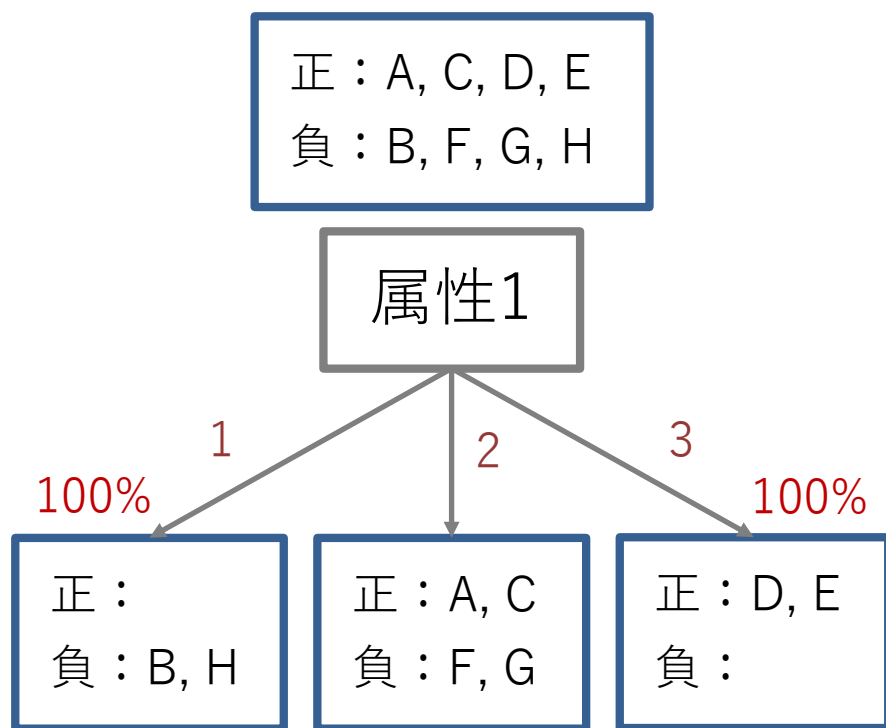
- すべての訓練例に矛盾のない決定木の中で、もっとも簡潔な（サイズの小さな）ものが、もっともありそうな仮説である。しかし、この取得には膨大な計算量が必要となる



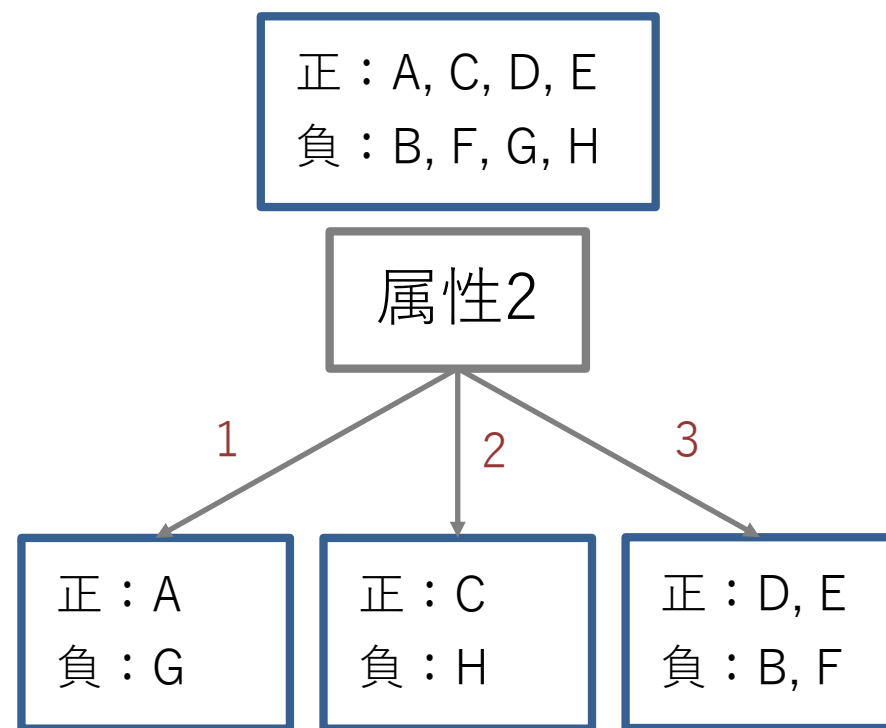
- 単純なヒューリスティックにより最小でなくともそれに近い決定木の取得法が提案されている

# 欲張り法による識別順の決定

- 識別力の高い順に属性をテストする



識別力が高い



識別力が低い



# 情報理論の利用(1/3)

- 識別力を定義する際のポイントは情報理論
- データが $n$ 個のクラスのどれか1つに分類できるとし、クラス $i$ に属する確率を $P_i$ と表す
- あるデータがクラス $i$ に属することを知るには、最低 $\log(1/P_i)$ ビットの情報量が必要（ $\log$ の底は2）
- 事象 $i$ が起こったときに受け取る選択情報量 $I(P_i)$ で表す

# 情報理論の利用(2/3)

- 各データをクラス1～ $n$ のいずれかに分類するのに必要な情報量は、その期待値を計算
- これを平均情報量（エントロピー）といい、 $I(P_1, \dots, P_n)$ で表す  
(ex. コインの表裏の分類には1ビットの平均情報量が必要)
- 決定木は、正例（Yes）と負例（No）の2クラスしか扱わない特殊ケース



# 情報理論の利用(3/3)

## ■ 平均情報量 (エントロピー)

$$I(P_1, \dots, P_n) = \sum_{i=1}^n P_i \log_2 \frac{1}{P_i} = - \sum_{i=1}^n P_i \log_2 P_i \quad (\text{ビット})$$

例：コイン投げ

表  $P_1 = 1/2$

裏  $P_2 = 1/2$

$$I\left(\frac{1}{2}, \frac{1}{2}\right) = \frac{1}{2} \log_2 2 + \frac{1}{2} \log_2 2 = 1$$

## ■ 正例が $p$ 個、負例が $n$ 個のときの平均情報量

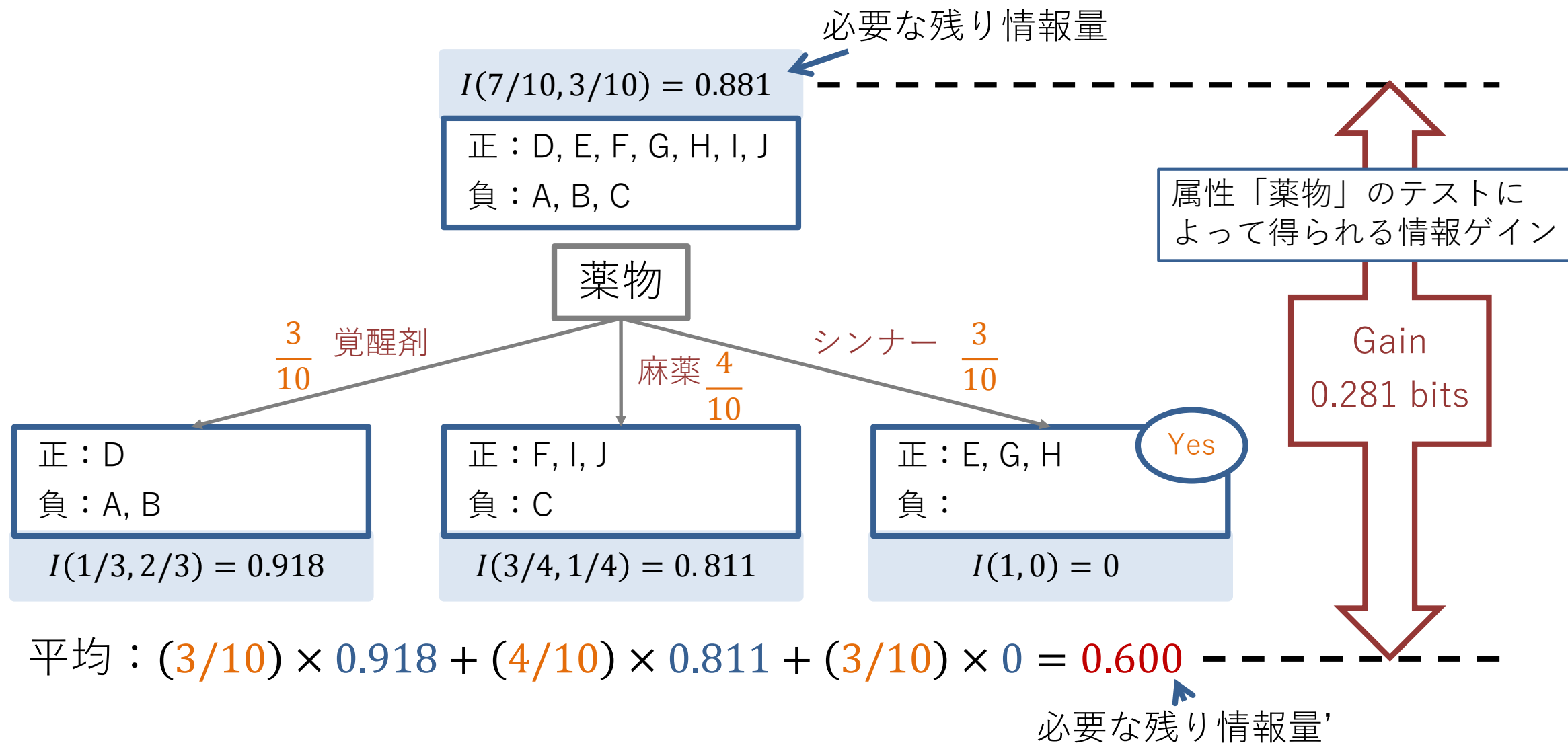
$$I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) = \frac{p}{p+n} \log_2 \frac{p+n}{p} + \frac{n}{p+n} \log_2 \frac{p+n}{n}$$

# 練習問題13-1

- データ数10のデータ集合を3つ、2つ、5つに分類できる特徴がある。この時、平均情報量（エントロピー）はいくらか

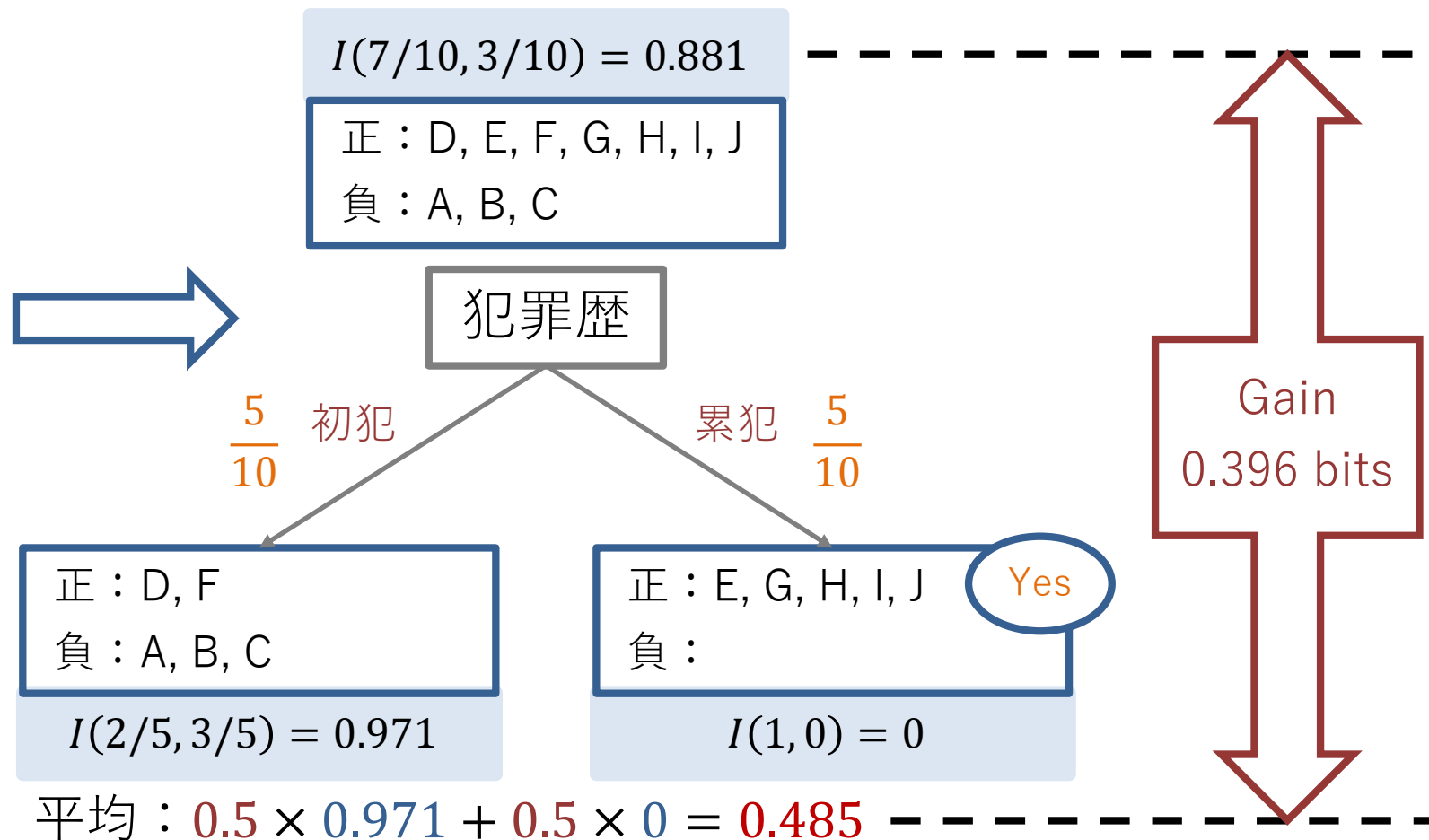
$$(\log_2 \frac{10}{3} = 1.7, \log_2 \frac{10}{2} = 2.3, \log_2 \frac{10}{5} = 1)$$

# 識別力=情報ゲイン



# ゲインが最大の属性を木の根に

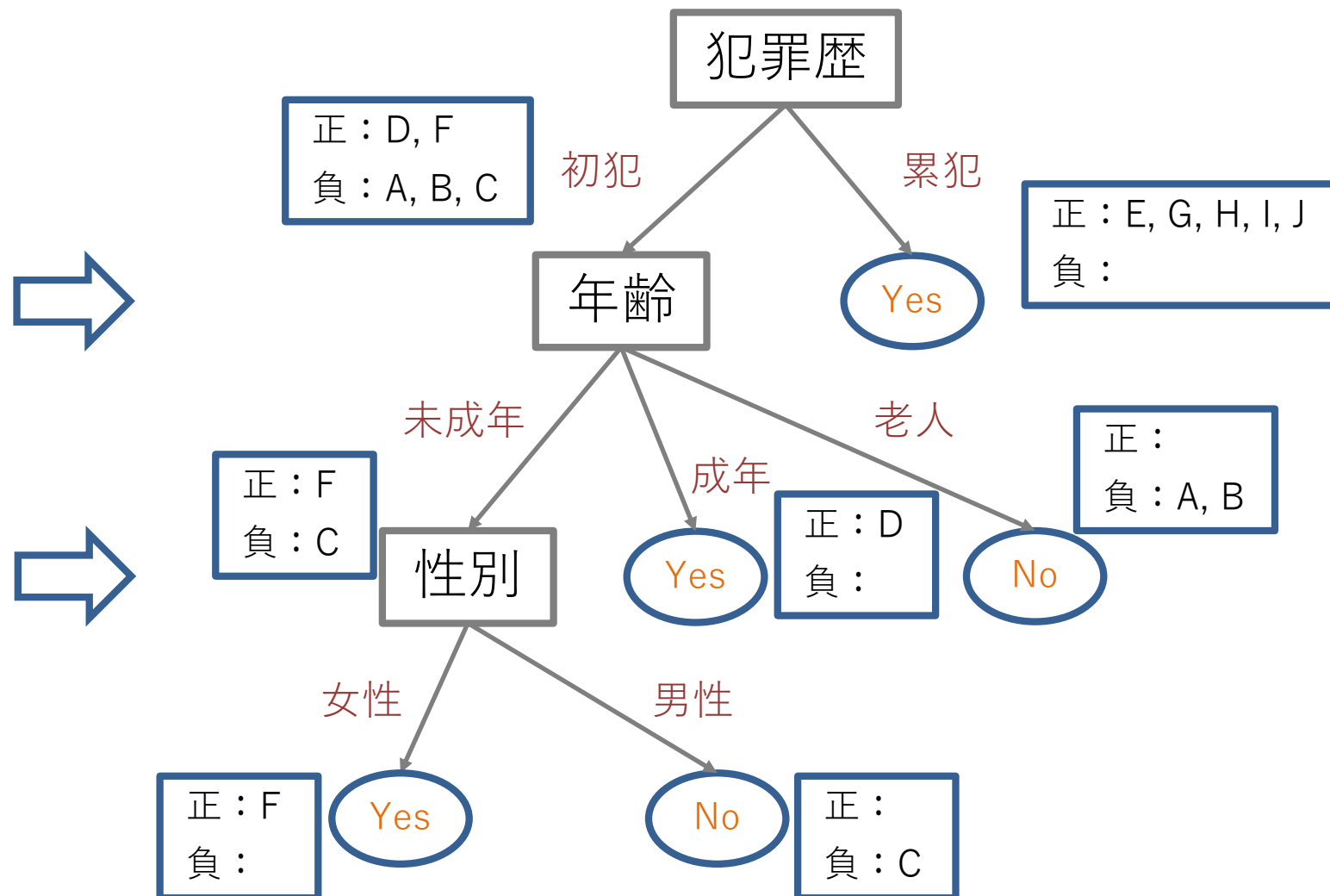
属性	ゲイン
性別	0.091
年齢	0.157
薬物	0.281
犯罪歴	0.396



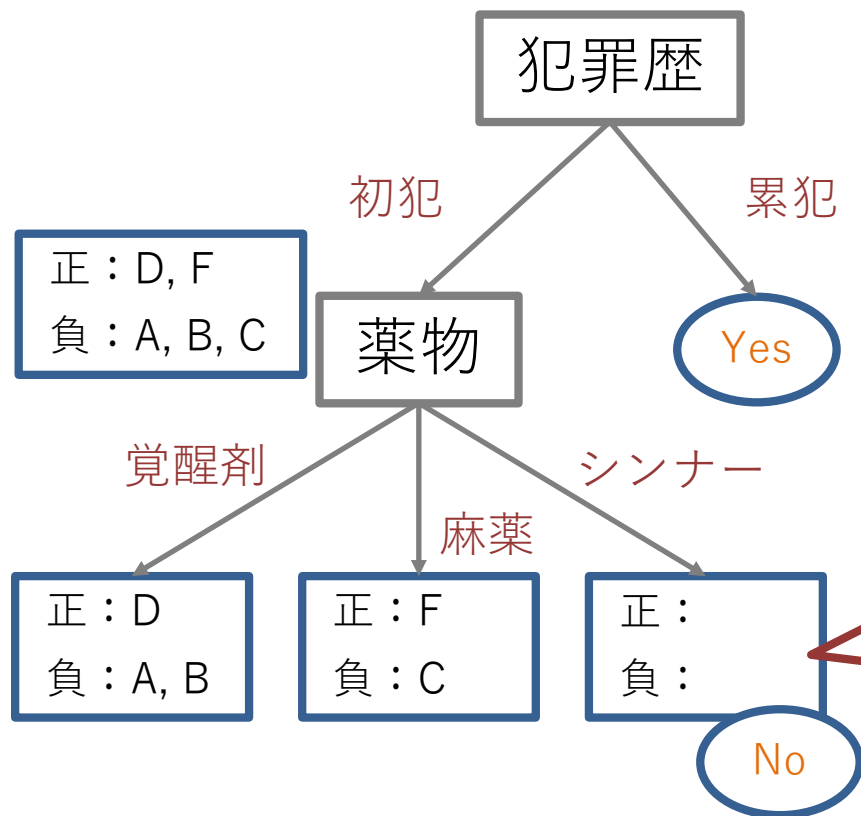
# 分割された訓練例の集合について再帰

属性	ゲイン
性別	0.042
年齢	0.571
薬物	0.020

属性	ゲイン
性別	1.000
薬物	0.000



# 決定不可能ならば多数決(1/2)



訓練例が空集合の場合：

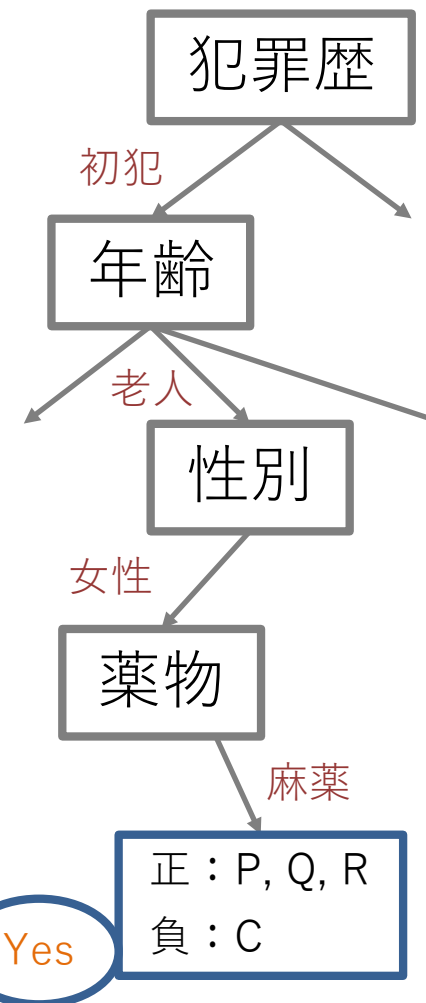
親ノードの多数決  
により、2対3で  
No  
とする



# 決定不可能ならば多数決(2/2)

誤り（ノイズ）のあるデータ：

	性別	年齢	薬物	犯罪歴	実刑
C	女性	老人	麻薬	初犯	No
P	女性	老人	麻薬	初犯	Yes
Q	女性	老人	麻薬	初犯	Yes
R	女性	老人	麻薬	初犯	Yes



属性が残っておらず、  
正負の例が混在する場合

多数決

により、3対1で

Yes

とする

Yes

# ID3アルゴリズム

- これまでのルールをまとめて

$S$  : 訓練例の集合  
 $A$  : 属性の集合  
default : Yes/Noの既定値

```
決定木 ID3( $S$ ,  $A$ , default){  
  if ( $S$ が空集合) return default  
  else if ( $S$ がすべて正例) return Yes  
  else if ( $S$ がすべて負例) return No  
  else if ( $A$ が空集合) return 多数決 ( $S$ )  
  else {  
    bestA =  $A$ のうち、ゲインが最大の属性;  
    tree = new 決定木(bestA);  
    bestAdomain = bestA の取り得るすべての値;  
    for each  $v$  in bestAdomain do {  
       $S'$  =  $S$ のうちbestA= $v$ となる全データ;  
      subtree = ID3( $S'$ ,  $A$ -bestA, 多数決( $S$ ));  
      treeの下にsubtreeを $v$ の枝で連結する;  
    }  
    return tree;  
  }  
}
```

# ノイズと過剰一致(1/2)

- 訓練例にはノイズと呼ばれる誤りの一種が含まれる
- ノイズまでも再現するような細かな分類は不適切な決定木を生み出すことが多い (過学習)

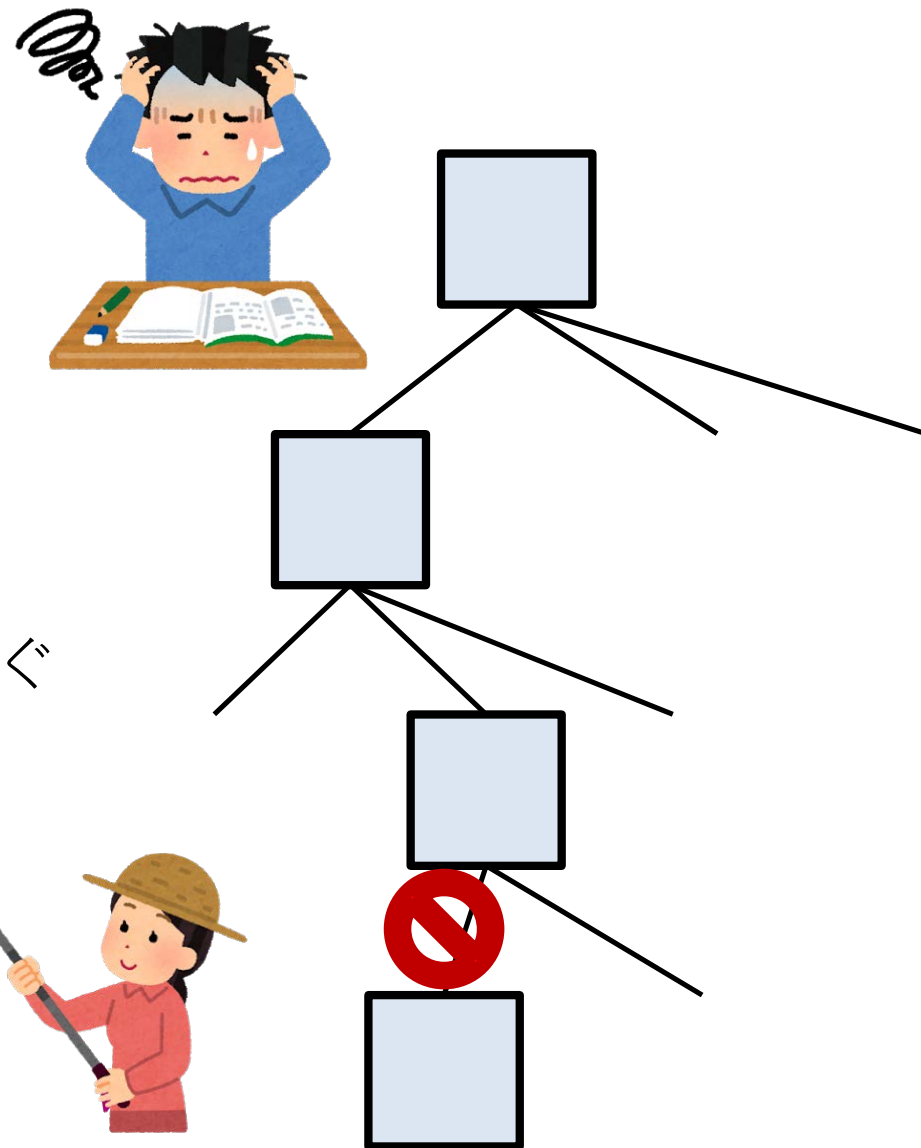


- 平均情報量のゲインが十分小さいときにはそれ以上の分類をやめ、終端ノードとする (枝刈り)

# ノイズと過剰一致(2/2)

## ■ 過学習 (overfitting)

細かく分類し過ぎて、意味のない仮説を生成してしまう



## ■ 枝刈り (pruning)

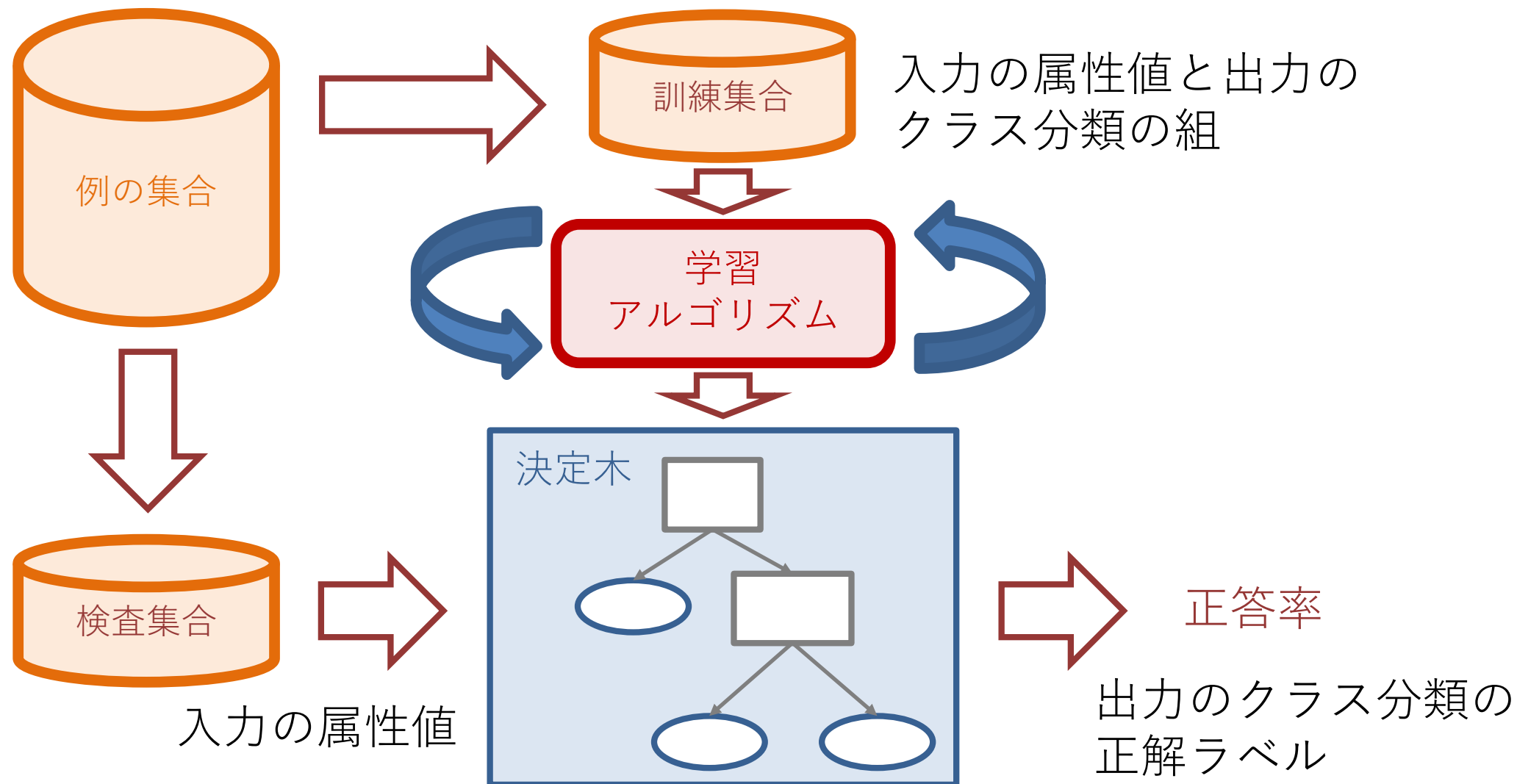
統計的仮説検定で、有意でない分類を防ぐ

# 決定木の性能評価(1/2)

1. データ収集：  
例となるデータ（入力の属性値と出力のクラス分類の組）を収集
2. データセット整理：  
例となるデータを訓練集合（トレーニング）と検査集合（テスト）に分割
3. 学習：  
訓練集合を学習アルゴリズムに入力して決定木を生成
4. 性能評価：  
検査集合を決定木に入力し、正答率を取得

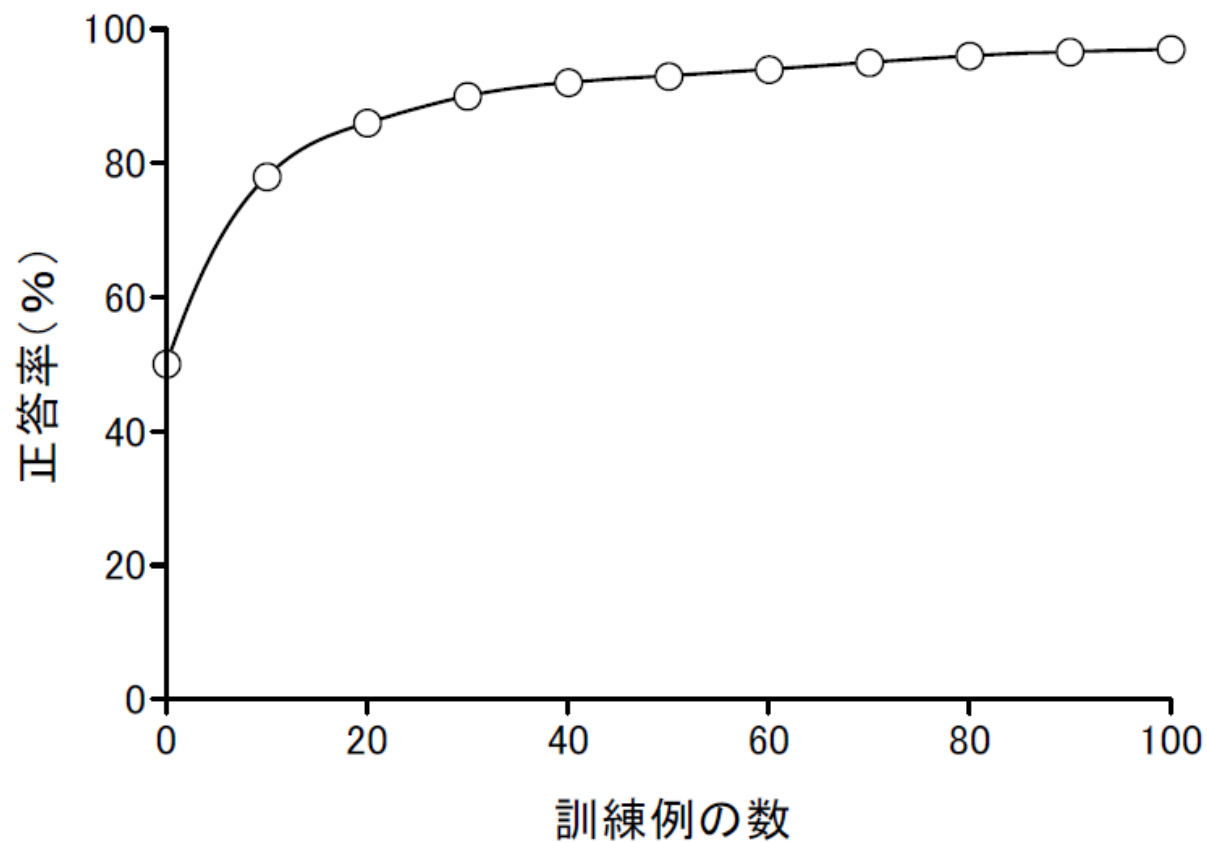
※決定木に限らず、学習アルゴリズムで一般的に使用されている

# 決定木の性能評価(2/2)



# 学習曲線

- 訓練例の数に応じた学習で正答率が推移する



# 決定木学習の現実応用(1/2)

## ■ 油田基地の機器設計（英国石油、1986）

- 油田基地のガス-石油分離システムを設計するエキスパートシステム
- ガス・石油・水の混合比、流速、圧力、密度などを設計
- 人手での設計：10人年（3650人日）→100人日
- 性能は専門家をしのぐ
- 100万ドルの節約を実現





# 決定木学習の現実応用(2/2)

## ■ 飛行学習 (1992)

- フライトシミュレータでセスナ機の操縦を学習
- 3人の熟練パイロットの30回の繰り返しから学習
- ノイズ除去効果により熟練パイロットの操縦ミスを除去し、先生である熟練パイロットよりも上手に飛べるようになった。



# まとめ

- 構造を学習する決定木について学んだ。
- 訓練例からの決定木の学習について学んだ。
- 情報理論に基づく属性の識別順の決定について学んだ。
- 決定木の学習におけるノイズと過剰一致について学んだ。

# 復習問題

1. 決定木の目的は、属性と値の組を何に分類する事か？
2. 決定木の学習の対象は何か？
3. 決定木における属性の識別力は何によって計算されるか？
4. 細かく分類し過ぎて、意味のない仮説を生成してしまう学習状態を何  
　　というか？

# 次回の講義

## ■ 強化学習

- マルコフ決定過程
- Q-learning