

第10章 同定木の構築

(Learning by Building Identification Tree)

10.1 データから同定木の作成

- (1) 決定木の作成方法
- (2) 不均一の最小化
- (3) 情報理論の利用

10.2 同定木からルールへ

- (1) 不必要な条件の削除
- (2) 不必要なルールの削除

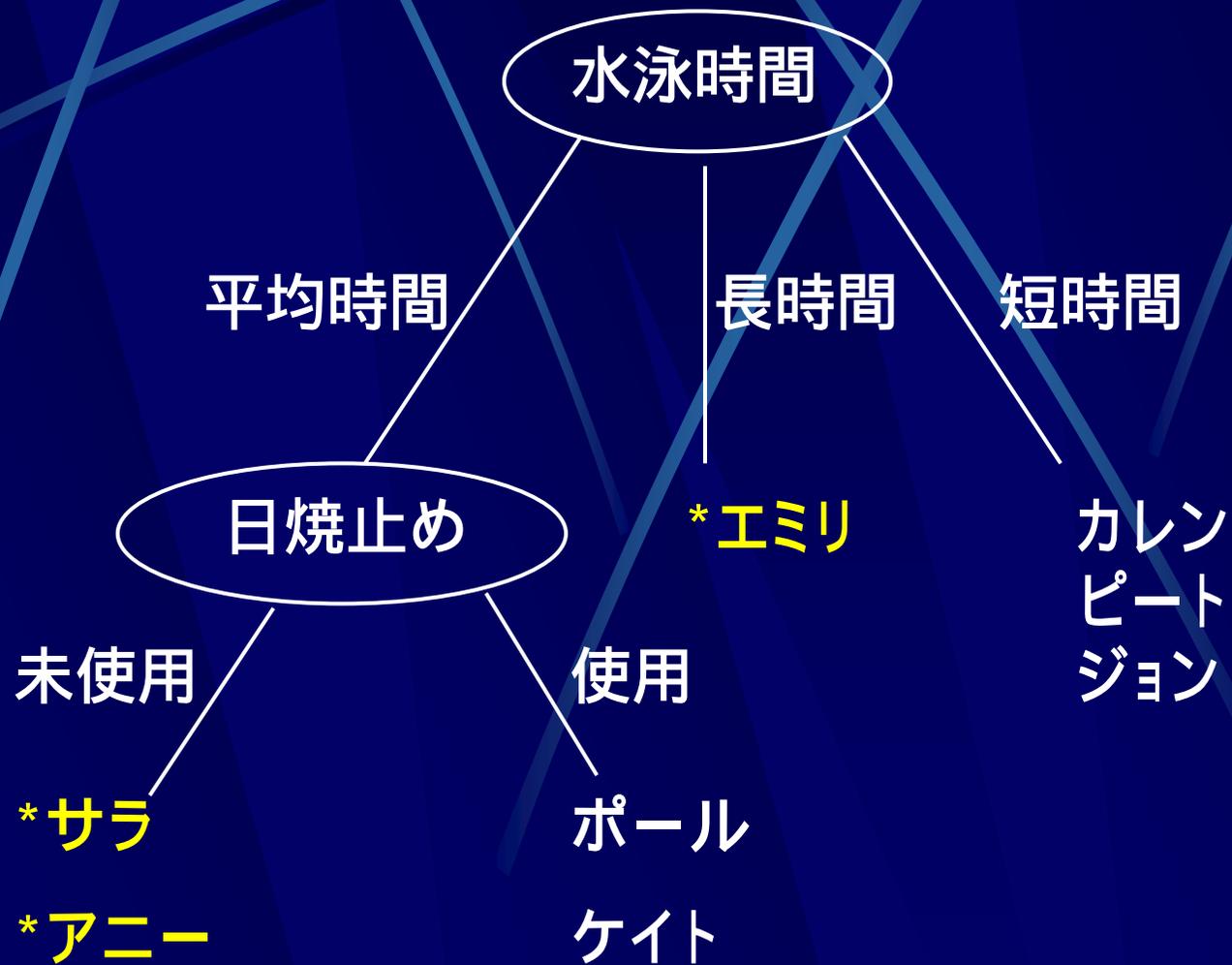
観測データ

名前	水泳時間	身長	体重	日焼止め	結果
サラ	平均時間	平均	軽い	未使用	日焼あり
ポール	平均時間	高い	平均	使用	日焼ナシ
カレン	短時間	低い	平均	使用	日焼ナシ
アニー	平均時間	低い	重い	未使用	日焼あり
エミリ	長時間	平均	重い	未使用	日焼あり
ピート	短時間	高い	重い	未使用	日焼ナシ
ジョン	短時間	平均	重い	未使用	日焼ナシ
ケイト	平均時間	低い	軽い	使用	日焼ナシ

10.1 データから同定木の作成

同定木：結論が、既知のクラスの
リストによって確立される
決定木

同定木の例 1



同定木の例 2



最小同定木を得るのは
計算上実際には不可能



どんなテストが旨く集合を
分けるか調べる

最小の同定木の作成方法

不均質さを最小にする .

不均質さ =

$$\sum_b \left(\frac{N_b}{N_t} \right) \times \left(\sum_c - \frac{N_{bc}}{N_b} \log_2 \frac{N_{bc}}{N_b} \right)$$

N_b : 枝 b のサンプル数

N_t : すべての枝のサンプル数の合計

N_{bc} : 枝 b における c のサンプル数

同定木の作成例は、
教科書p.107 ~ p.110を参照

10.2 同定木からルールの作成

条件：木の根から葉に向かって
頂点に対応するテスト

結論：葉頂点のクラス（種類）

使用する述語

述語	意味
水泳時間(X, Y)	X の水泳時間は Y である。
身長(X, Y)	X の身長は Y である。
体重(X, Y)	X の体重は Y である。
日焼止め(X, Y)	X は日焼止めを Y である。
結果(X, Y)	X の結果は Y である。

同定木からルールの作成例は、
教科書p.110～p.114を参照