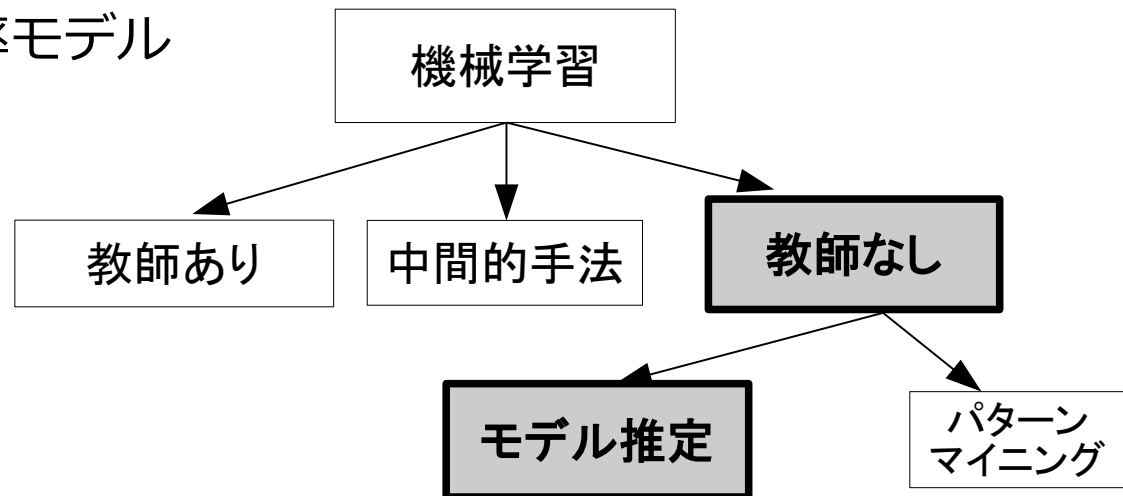


機械学習の発展的な話題

- 教師なし学習
- 半教師あり学習
- 深層学習

10. モデル推定

- 問題設定
 - 教師なし学習
 - 数値入力 → クラスモデル
 - クラスモデルの例
 - クラスの分割結果
 - クラスの確率モデル



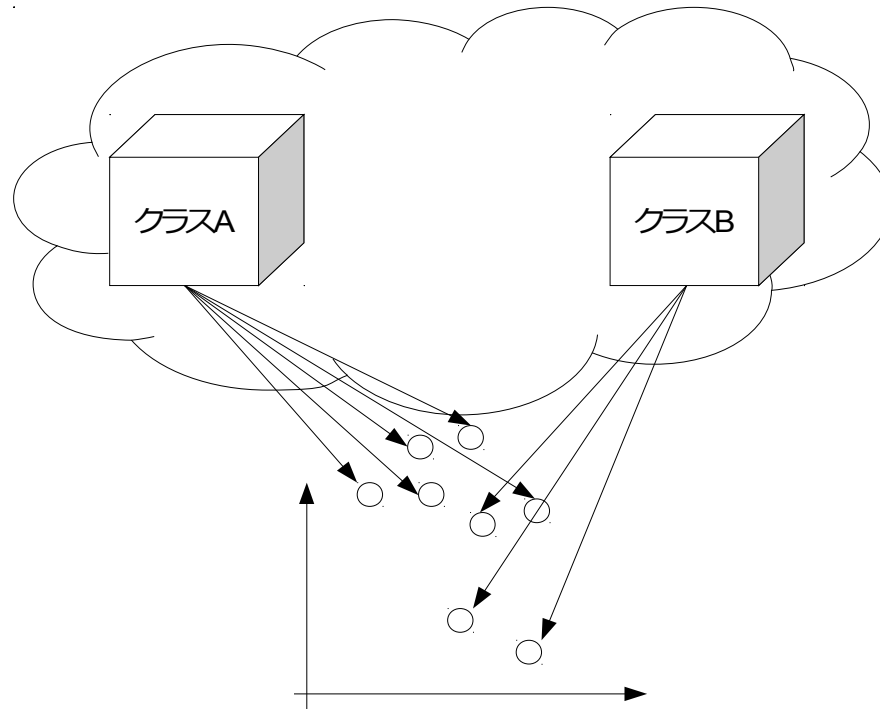
10.1 問題の定義

- 学習データ

$$\{x^{(i)}\} \quad i = 1, \dots, N$$

- 問題設定

- 特徴ベクトル x が生成された元のクラスの性質を推定する



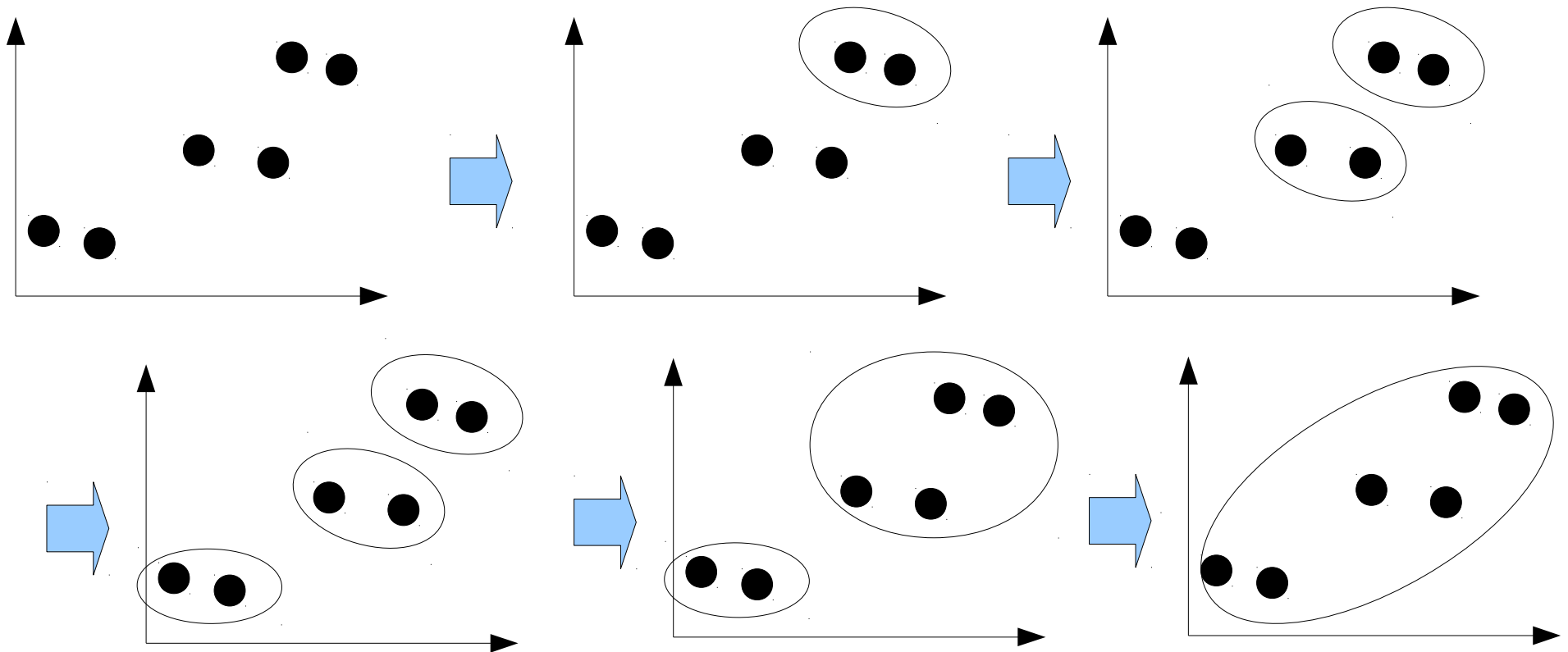
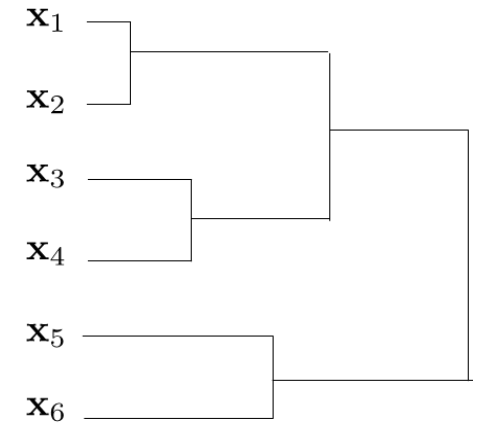
10.2 クラスタリング

- クラスタリングとは
 - 対象のデータを、
内的結合（同じ集合内のデータ間の距離は小さく）
と
外的分離（異なる集合間の距離は大きく）
が達成されるような部分集合に分割すること
- クラスタリング手法の分類
 - 階層的な手法
 - ボトムアップ的にデータをまとめてゆく
 - 分割最適化手法
 - トップダウン的にデータ集合を分割してゆく

要するに
塊を見つ
けること

10.2.1 階層的クラスタリング

- 階層的クラスタリングとは
 - 1.1 データ 1 クラスタからスタート
 2. 最も近接するクラスタをまとめる
 3. 全データが 1 クラスタになれば終了



10.2.2 分割最適化クラスタリング — k-means アルゴリズム —

- 分割最適化クラスタリングとは
 - データ分割の良さを評価する関数を定め、その評価関数の値を最適化することを目的とする
 - ただし、全ての可能な分割に対して評価値を求めることは、データ数 N が大きくなると、不可能
 - 2 分割で 2^N 通り
 - 探索によって、準最適解を求める

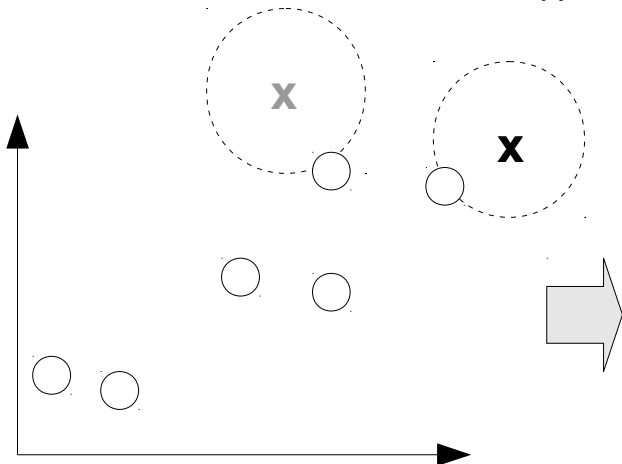
10.2.2 分割最適化クラスタリング — k-means アルゴリズム —

- k-Means アルゴリズム

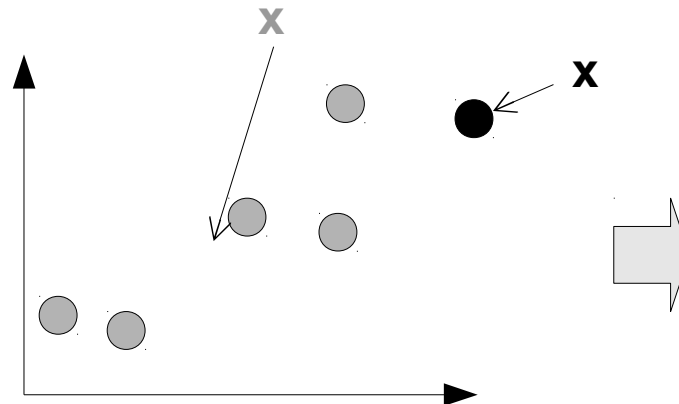
1. 分割数 k を予め与える

2. 乱数で k 個のクラスタ中心を設定し、逐次更新

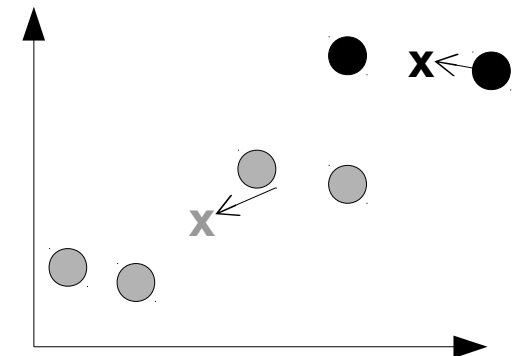
$k=2$ とし、初期値として
乱数でクラスタ中心を配置



全データを近い方のクラスタ
中心に所属させる。そして、
クラスタ中心を所属している
データの平均へ移動。



左の処理を繰り返す。



10.2.3 自動分割最適化クラスタリング — X-means アルゴリズム —

- k-means 法の問題点
 - 分割数 k を予め決めなければならない
- 解決法 \Rightarrow X-means アルゴリズム
 - 2 分割から始めて、分割数を適応的に決定する
 - 分割の妥当性の判断： BIC(Bayesian information criterion) が小さくなれば、分割を継続

$$BIC = -2 \log L + q \log N$$

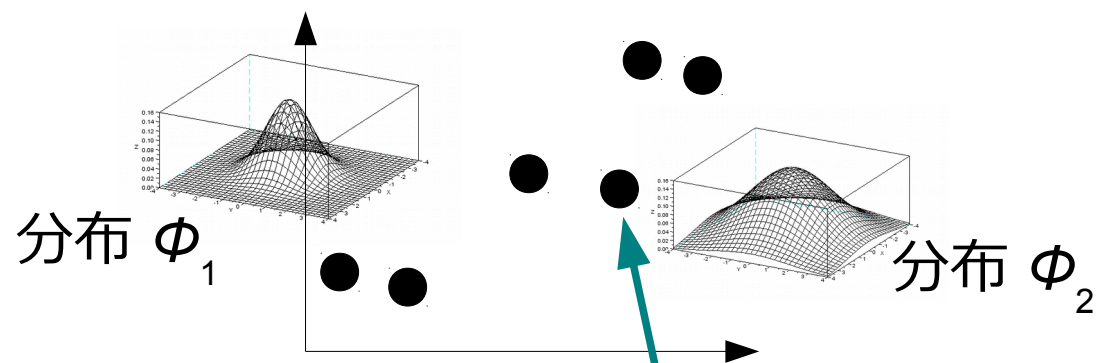
- L : モデルの尤度
- q : モデルのパラメータ数
- N : データ数

パラメータで表される
統計モデルの選択基準
(小さいほどよいモデル)

10.4 確率密度推定

- 教師なし学習で識別器を作る問題
 - クラスタリング結果からは、1クラス1プロトタイプ
の単純な識別器しかできない
 - 各クラスの事前確率や確率密度関数も推定したい

➡ EM アルゴリズム



分布 ϕ_1 の再計算の際、
重み 0.2 だけ寄与する

$$0.2\phi_1 + 0.8\phi_2$$

10.4 確率密度推定

- k-means 法の一般化
 - k 個の平均ベクトルを乱数で決める
⇒ k 個の正規分布を乱数で決める
 - 平均ベクトルとの距離を基準に、各データをいずれかのクラスタに所属させる
⇒ 各分布が各データを生成する確率を計算し、各クラスタにゆるやかに帰属させる
 - 所属させたデータをもとに平均ベクトルを再計算
⇒ 各データのクラスタへの帰属度に基づき各分布のパラメータ（平均値、共分散行列）を再計算

Weka Explorer

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

Open file... | Open UR... | Open DB... | Generate... | Undo | Edit... | Save...

Filter: Choose | None | Apply

Current relation: Relation: iris, Instances: 150, Attributes: 5, Sum of weights: 150

Selected attribute: Name: class, Missing: 0 (0%), Distinct: 3, Type: Nominal, Unique: 0 (0%)

No.	Label	Count	Weight
1	Iris-setosa	50	50.0
2	Iris-versicolor	50	50.0
3	Iris-virginica	50	50.0

Attributes: All | None | Invert | Pattern

No.	Name
1	sepallength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Class: class (Nom) | Visualize All

Remove selected attributes.

Status: OK | Log | x 0

class ラベルを消す

Remove

Weka の HierarchicalCluster

The screenshot shows the Weka Explorer interface with the 'Cluster' tab selected. The 'Clusterer' dropdown is set to 'HierarchicalClusterer -N 2 -L WARD -P -A "weka.core.EuclideanDistance -R first-last"'. The 'Cluster mode' section has 'Use training set' selected, and 'Store clusters for visualization' is checked. The 'Clusterer output' pane shows the following details:

```
Scheme:      weka.clusterers.HierarchicalClusterer -N 2 -L WARD -P -A "weka.core.EuclideanDistance -R first-last"
Relation:    iris-weka, filters, unsupervised, attribute, Remove-R5
Instances:   150
Attributes:  4
              sepallength
              sepalwidth
              petallength
              petalwidth
Test mode:   evaluate on training data
```

==== Clustering model (full training set) ====

Cluster 0
(((((((0.2:0.03254,0.2:0.03254):0.01891,((0.2:0.02778,0.2:0.02778):0.00843,

Cluster 1
(((((((1.4:0.07344,1.5:0.07344):0.08446,((1.5:0.09914,1.4:0.09914):0.0122, (

Time taken to build model (full training data) : 0.13 seconds

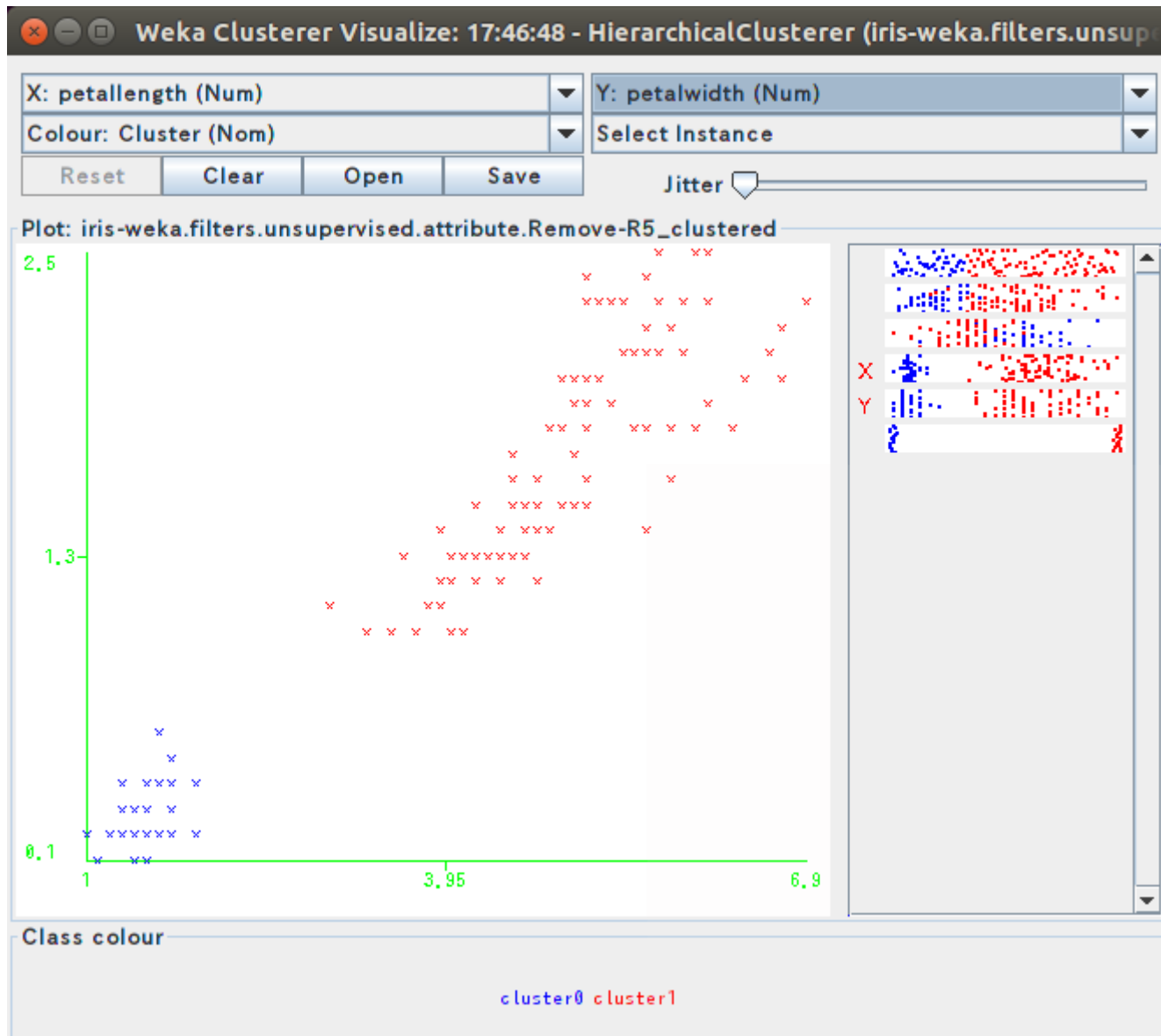
==== Model and evaluation on training set ====

Clustered Instances

0	50	(33%)
1	100	(67%)

The 'Result list' shows two entries for 'HierarchicalClusterer' at 17:46:06 and 17:46:48. The 'Status' bar at the bottom indicates 'OK'.

Weka の HierarchicalCluster



Weka の SimpleKMeans

Clusterer

Choose SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 3 -A 'weka.core

Cluster mode

Use training set

Supplied test set Set...

Percentage split % 66

Classes to clusters evaluation

(Num) petalwidth

Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

17:49:29 - HierarchicalClusterer

17:50:11 - SimpleKMeans

Clusterer output

Within cluster sum of squared errors: 6,998114004826762

Initial starting points (random):

Cluster 0: 6,1,2,9,4,7,1,4

Cluster 1: 6,2,2,9,4,3,1,3

Cluster 2: 6,9,3,1,5,1,2,3

Missing values globally replaced with mean/node

Final cluster centroids:

Attribute	Full Data	Cluster#		
		0	1	2
	(150,0)	(61,0)	(50,0)	(39,0)
sepal.length	5,8433	5,8885	5,006	6,8462
sepal.width	3,054	2,7377	3,418	3,0821
petal.length	3,7587	4,3967	1,464	5,7026
petal.width	1,1987	1,418	0,244	2,0795

Time taken to build model (full training data) : 0,02 seconds

=== Model and evaluation on training set ===

Clustered Instances

0	61 (41%)
1	50 (33%)
2	39 (26%)

クラスタ数 = 3

クラスタ中心が得られている

Status OK

Log x 0

Weka の EM

Weka Explorer

Preprocess | Classify | **Cluster** | Associate | Select attributes | Visualize

Clusterer

Choose **EM** -I 100 -N 3 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100

Cluster mode

- Use training set
- Supplied test set
- Percentage split %
- Classes to clusters evaluation
- Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

- 17:49:29 - HierarchicalClusterer
- 17:50:11 - SimpleKMeans
- 17:51:30 - EM
- 17:51:50 - EM**

Clusterer output

Attribute	Cluster		
	0	1	2
	(0.41)	(0.33)	(0.25)
=====			
sepal length			
mean	5.9275	5.006	6.8085
std. dev.	0.4817	0.3489	0.5339
sepal width			
mean	2.7503	3.418	3.0709
std. dev.	0.2956	0.3772	0.2867
petal length			
mean	4.4057	1.464	5.7233
std. dev.	0.5254	0.1718	0.4991
petal width			
mean	1.4131	0.244	2.1055
std. dev.	0.2627	0.1061	0.2456

Time taken to build model (full training data) : 0.01 seconds

=== Model and evaluation on training set ===

Clustered Instances

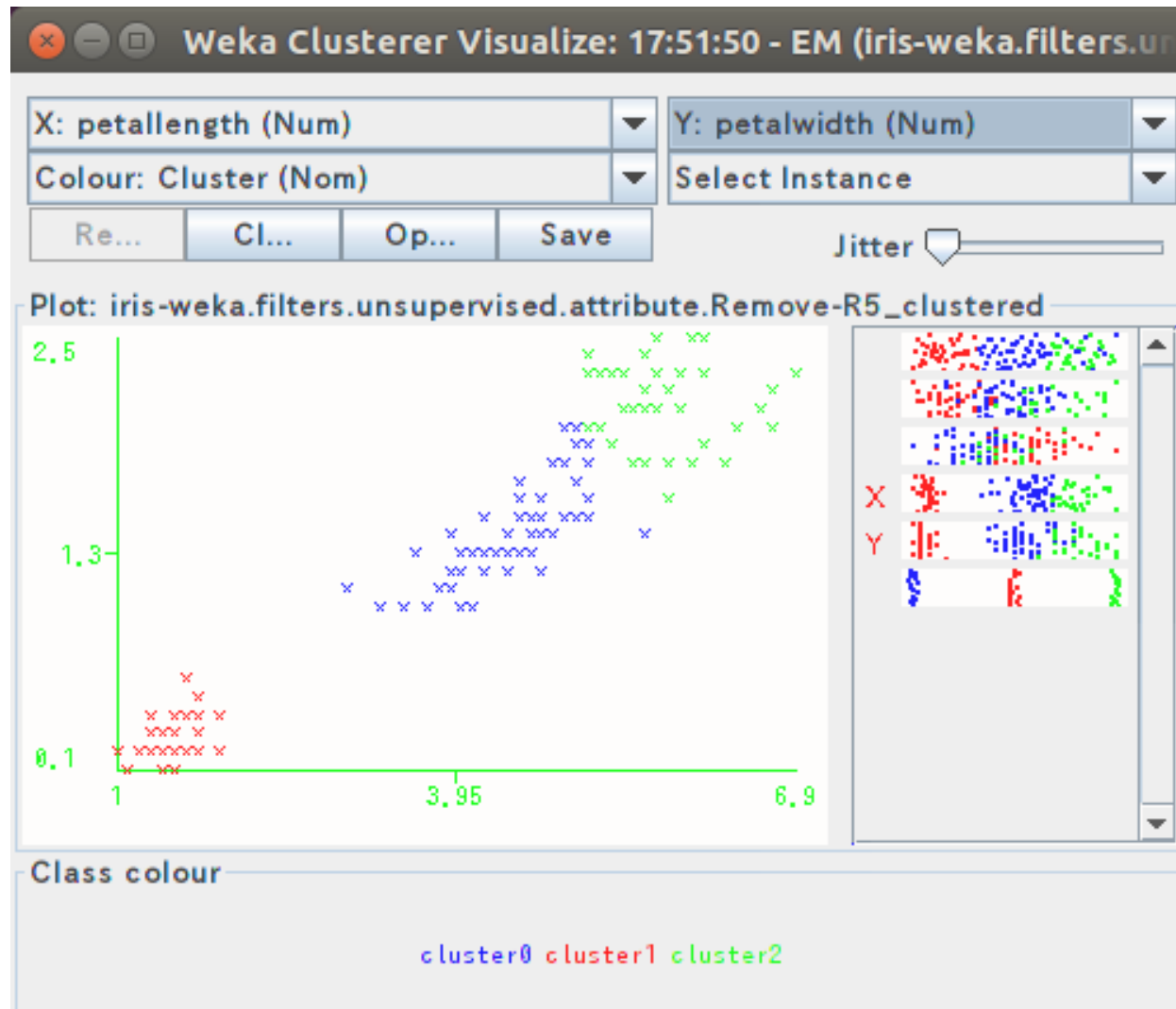
0	64 (43%)
1	50 (33%)
2	36 (24%)

確率分布が得られている

Status OK

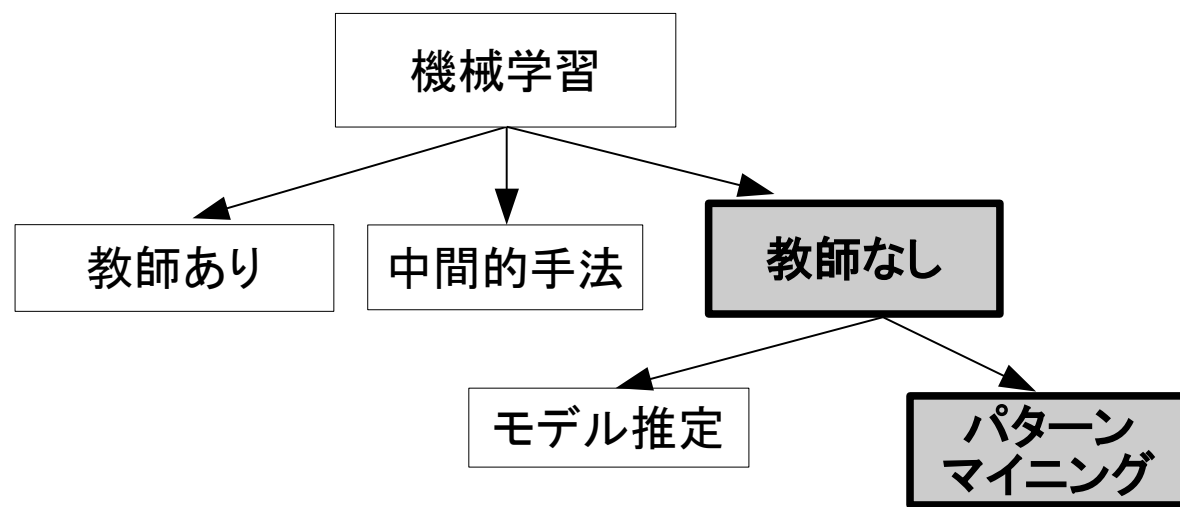
Log x 0

Weka の EM



11章 パターンマイニング

- パターンマイニングの問題設定
 - 入力：ラベル特徴の教師なしデータ
 - 出力：頻出項目、連想規則



No.	ミルク	パン	バター	雑誌
1	t	t		
2		t		
3				t
4		t	t	
5	t	t	t	
6	t	t		

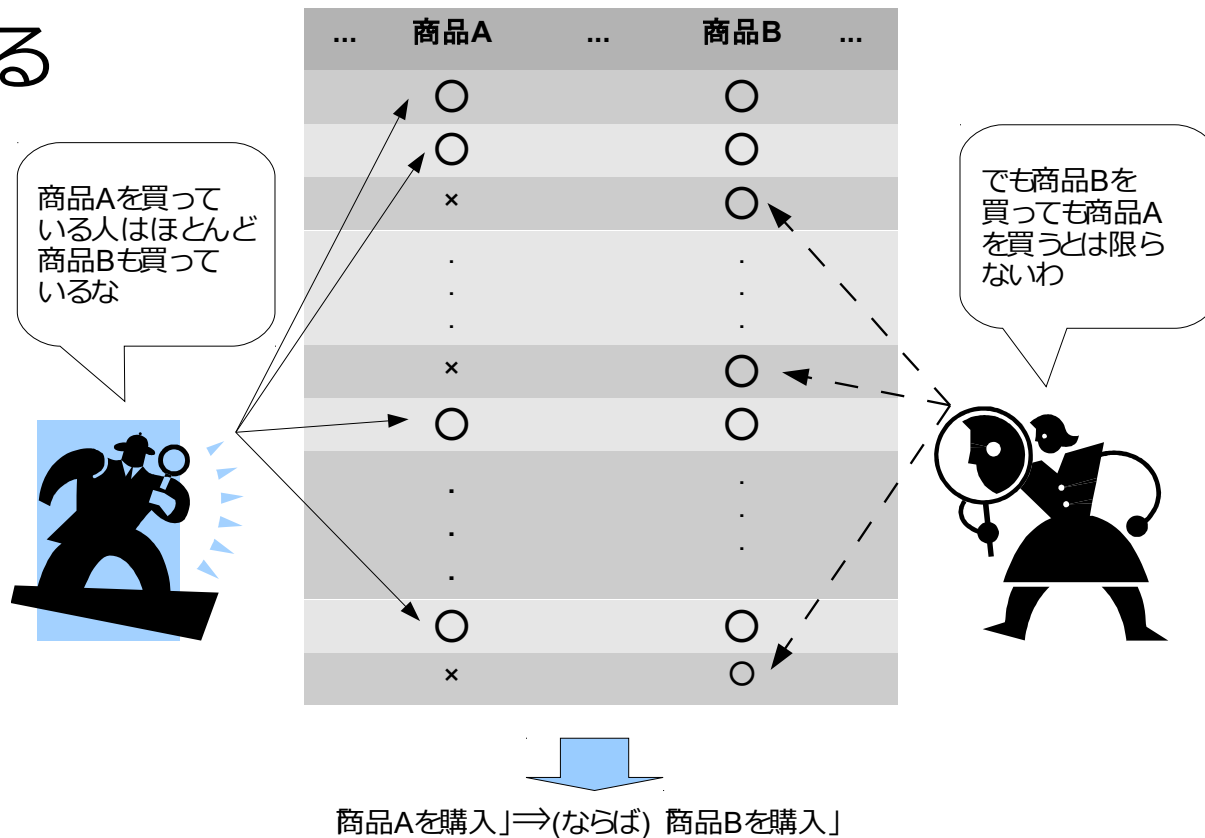
問題の定義

- 学習データ

$$\{\mathbf{x}^{(i)}\} \quad i = 1, \dots, N$$

- 問題設定

- データ集合中で、一定頻度以上で現れるパターンを抽出する



11.2 Apriori アルゴリズムによる頻出項目抽出

- 例題：バスケット分析

No.	ミルク	パン	バター	雑誌
1	t	t		
2		t		
3				t
4		t	t	
5	t	t	t	
6	t	t		

- 支持度

- 全データ T に対して、ある項目集合が出現する割合
- 支持度が閾値以上の項目集合を抽出したい

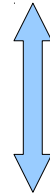
$$\text{support}(items) = \frac{T_{items}}{T}$$

項目集合の種類数は、
2 の商品数乗

11.2 Apriori アルゴリズムによる頻出項目抽出

- a priori な原理

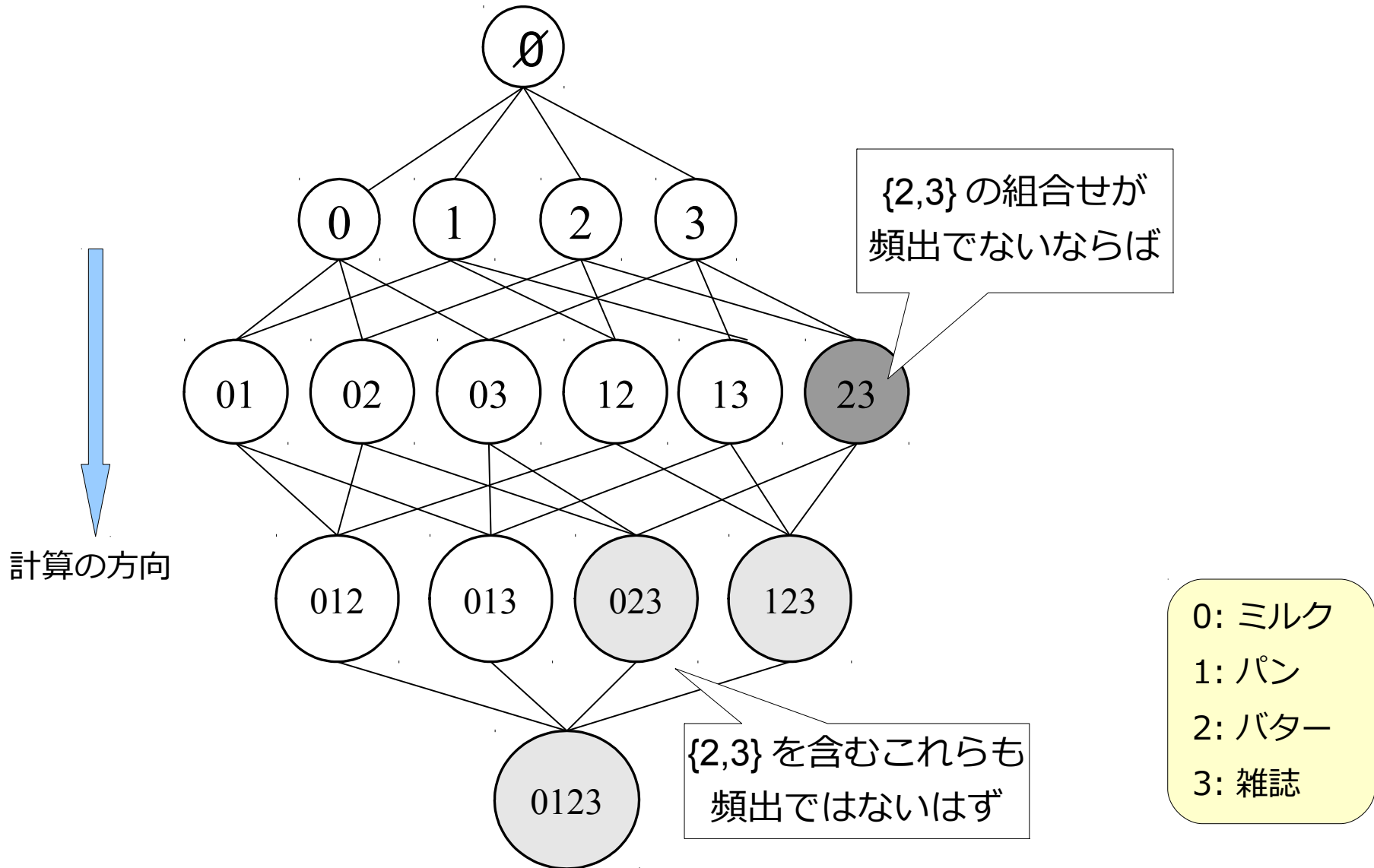
ある項目集合が頻出ならば、その部分集合も頻出である



対偶

ある項目集合が頻出でないならば、
その項目集合を含む集合も頻出でない

11.2 Apriori アルゴリズムによる頻出項目抽出



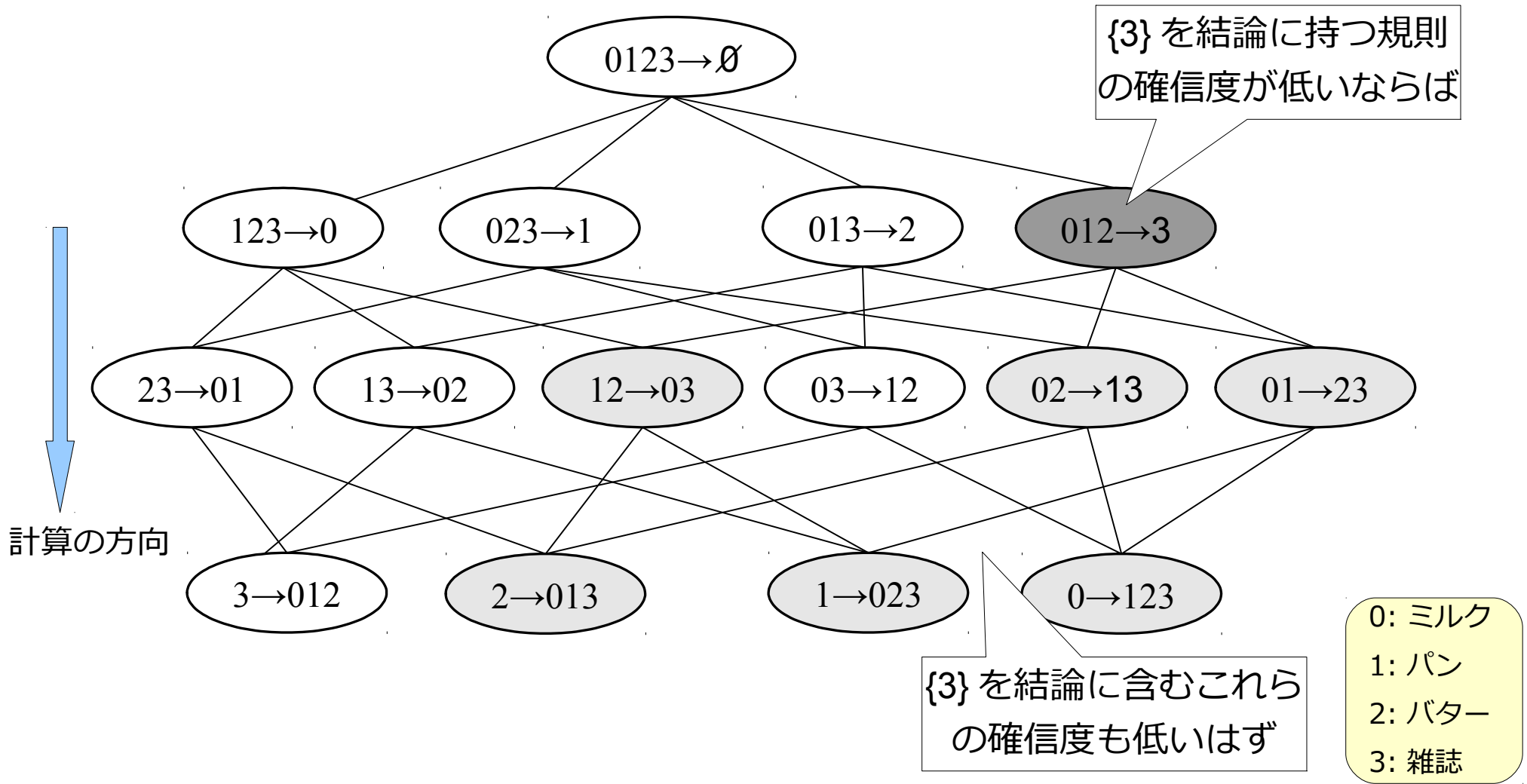
11.3 連想規則抽出

- 連想規則
 - 頻出項目集合を、条件部と結論部に分ける
 - 確信度の高い規則を抽出

$$\text{confidence}(A \rightarrow B) = \frac{\text{support}(A \cup B)}{\text{support}(A)}$$

- a priori 原理による探索

11.3 連想規則抽出



supermarket データ

Viewer

Relation: supermarket

12: baby needs Nominal	13: bread and cake Nominal	14: baking needs Nominal	15: coupons Nominal	16: juice-sat-cord-ms Nominal	17: tea Nominal	18: biscuits Nominal	19: canned fish-meat Nominal	20: canned fruit Nominal	:
	t	t			t	t	t		t
				t					
	t			t		t		t	t
t	t	t		t		t		t	t
	t	t		t		t			
				t	t	t		t	
				t		t			
	t	t		t	t	t		t	t
	t	t		t		t		t	t
	t	t							
	t				t	t			t
t				t					
	t	t							
		t			t	t		t	
		t			t	t			t
	t			t				t	t
	t	t				t			t
	t	t				t			t
	t	t				t			t
	t	t				t	t		t
t	t	t		t		t	t	t	t
t	t	t				t		t	t
	t								
	t			t			t		
	t	t		t		t			t
	t			t		t			
	t	t		t		t	t		t
t		t		t		t		t	t
	t								

Undo OK Cancel

Weka の Apriori

The screenshot shows the Weka Explorer interface with the 'Associate' tab selected. The 'Apriori' algorithm is running with the following parameters: `Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1`. The 'Associator output' pane displays the following information:

Minimum support: 0.15 (694 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:

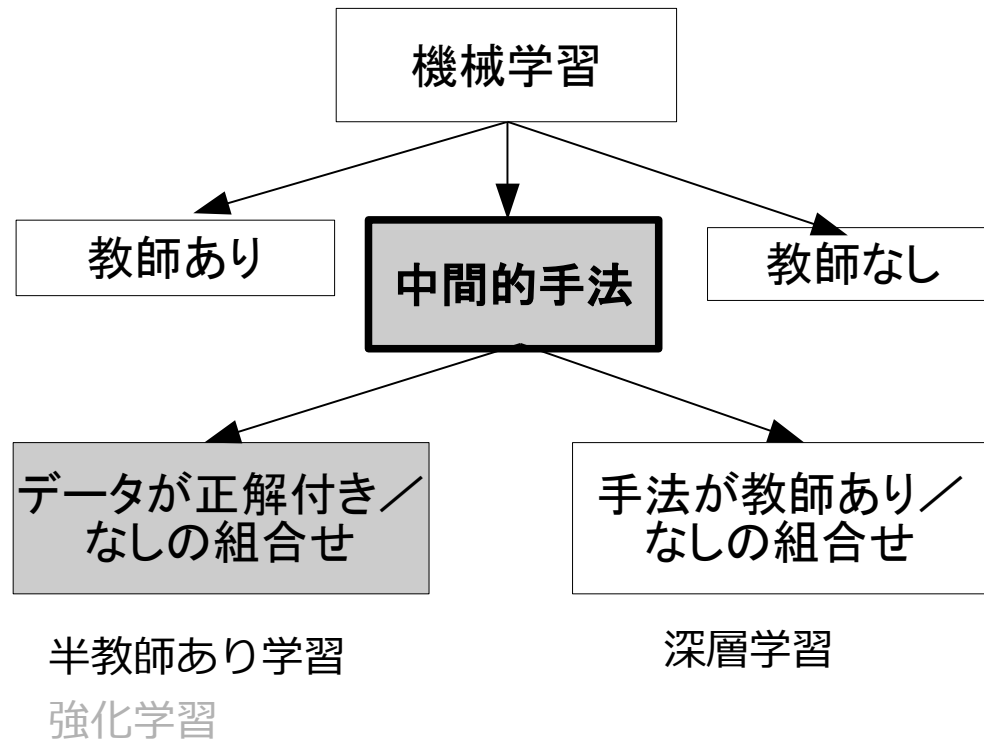
- Size of set of large itemsets L(1): 44
- Size of set of large itemsets L(2): 380
- Size of set of large itemsets L(3): 910
- Size of set of large itemsets L(4): 633
- Size of set of large itemsets L(5): 105
- Size of set of large itemsets L(6): 1

Best rules found:

1. biscuits=t frozen foods=t fruit=t total=high 788 ==> bread and cake=t 723 <conf:(0.92)> lift:(1.27) lev:(0.04)
2. baking needs=t biscuits=t fruit=t total=high 760 ==> bread and cake=t 696 <conf:(0.92)> lift:(1.27) lev:(0.04)
3. baking needs=t frozen foods=t fruit=t total=high 770 ==> bread and cake=t 705 <conf:(0.92)> lift:(1.27) lev:(0.04)
4. biscuits=t fruit=t vegetables=t total=high 815 ==> bread and cake=t 746 <conf:(0.92)> lift:(1.27) lev:(0.04)
5. party snack foods=t fruit=t total=high 854 ==> bread and cake=t 779 <conf:(0.91)> lift:(1.27) lev:(0.04)
6. biscuits=t frozen foods=t vegetables=t total=high 797 ==> bread and cake=t 725 <conf:(0.91)> lift:(1.27) lev:(0.04)
7. baking needs=t biscuits=t vegetables=t total=high 772 ==> bread and cake=t 701 <conf:(0.91)> lift:(1.27) lev:(0.04)
8. biscuits=t fruit=t total=high 954 ==> bread and cake=t 866 <conf:(0.91)> lift:(1.26) lev:(0.04)
9. frozen foods=t fruit=t vegetables=t total=high 834 ==> bread and cake=t 757 <conf:(0.91)> lift:(1.26) lev:(0.04)
10. frozen foods=t fruit=t total=high 969 ==> bread and cake=t 877 <conf:(0.91)> lift:(1.26) lev:(0.04)

The 'Result list (right-click to open)' pane shows a single entry: `17:53:31 - Apriori`. The 'Status' bar at the bottom indicates 'OK' and a 'Log' button is visible.

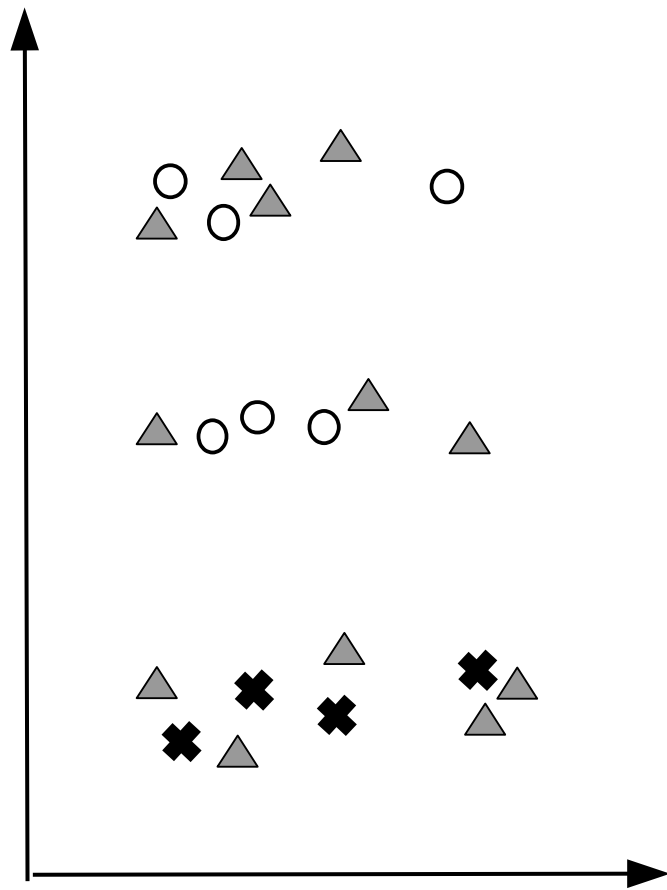
13 章 半教師あり学習



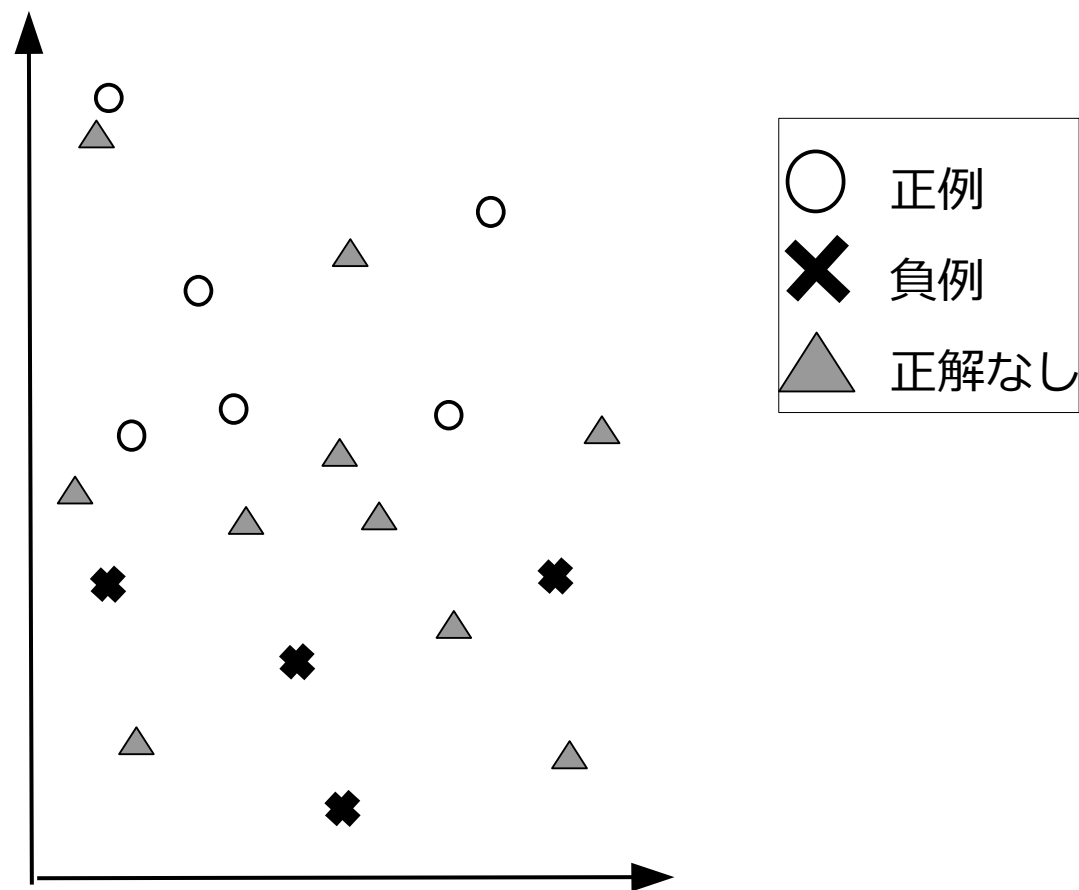
13.1 半教師あり学習とは

13.1.1 数値特徴の場合

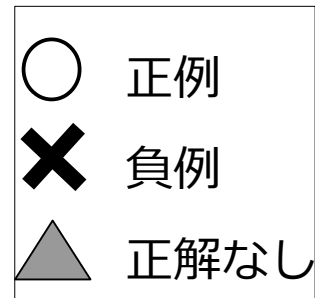
- 半教師あり学習に適した数値特徴データの性質



半教師あり学習に適するデータ



半教師あり学習に適さないデータ



13.1.1 数値特徴の場合

- 半教師あり学習が可能なデータ
 - 半教師あり平滑性仮定
 - 二つの入力が高密度領域で近ければ、出力も関連している
 - クラスタ仮定
 - もし入力が同じクラスタに属するなら、それらは同じクラスになりやすい
 - 低密度分離
 - 識別境界は低密度領域にある
 - 多様体仮定
 - 高次元のデータは、低次元の多様体上に写像できる

13.1.2 ラベル特徴の場合

- オーバーラップ

Positive ○

... よかった。 ..
...
高性能 ..
...
... 満足

?

...
...
高性能 ..
... 満足 .
....

?

....
...
高性能 ..
...
... よかった。

Negative ✕

... 壊れた。 ..
...
不満 ..
...
... 買わない

?

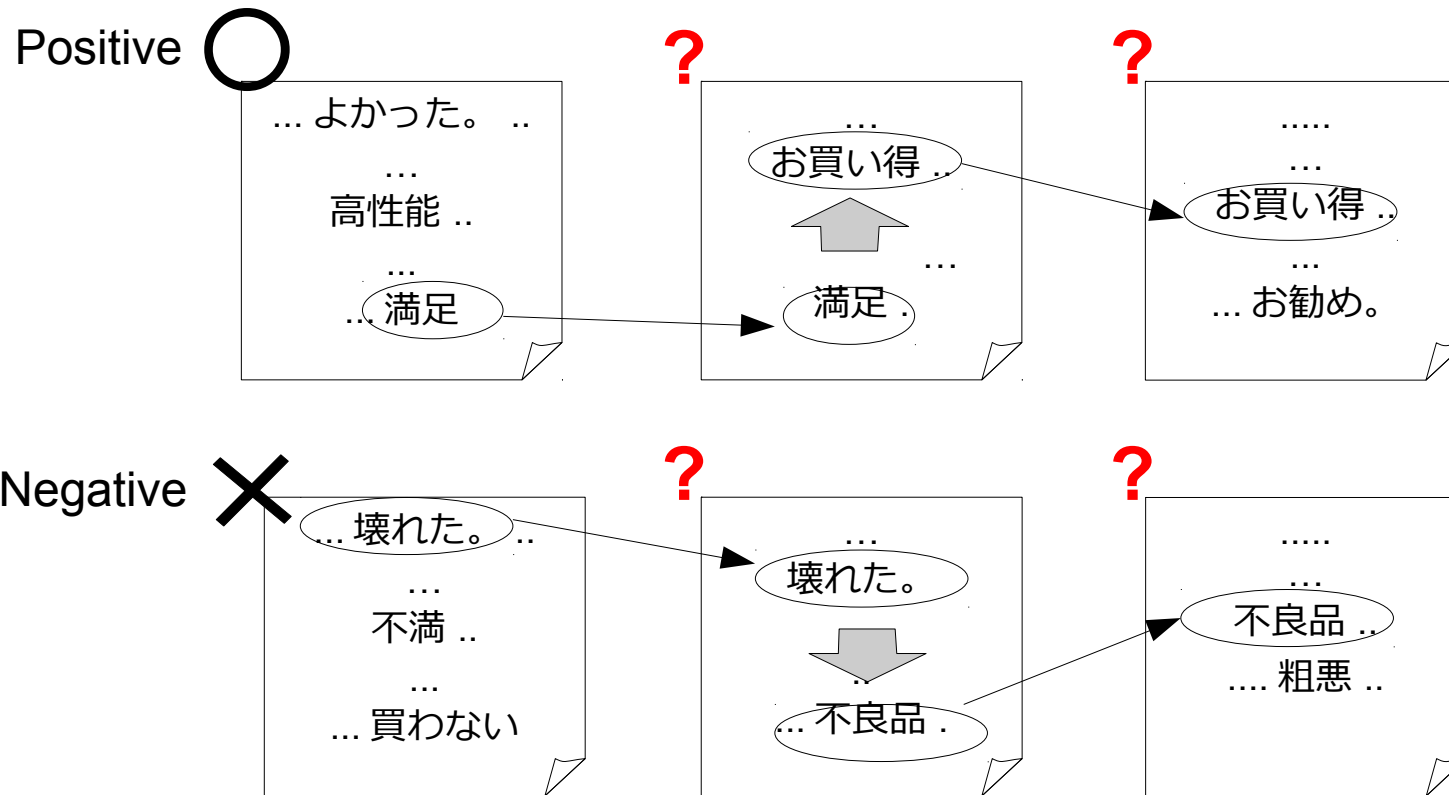
...
...
壊れた。 ..
... 買わない .
....

?

....
...
不満 ..
...
... 買わない

13.1.2 ラベル特徴の場合

- 伝播



13.1.3 半教師あり学習のアルゴリズム

- 半教師あり学習の基本的な考え方
 - 正解付きデータで識別器を作成
 - 正解なしデータで識別器のパラメータを調整
- 識別器に対する要求
 - 確信度の出力：正解なしデータに対する出力を信用するかどうかの判定に必要

13.2 自己学習

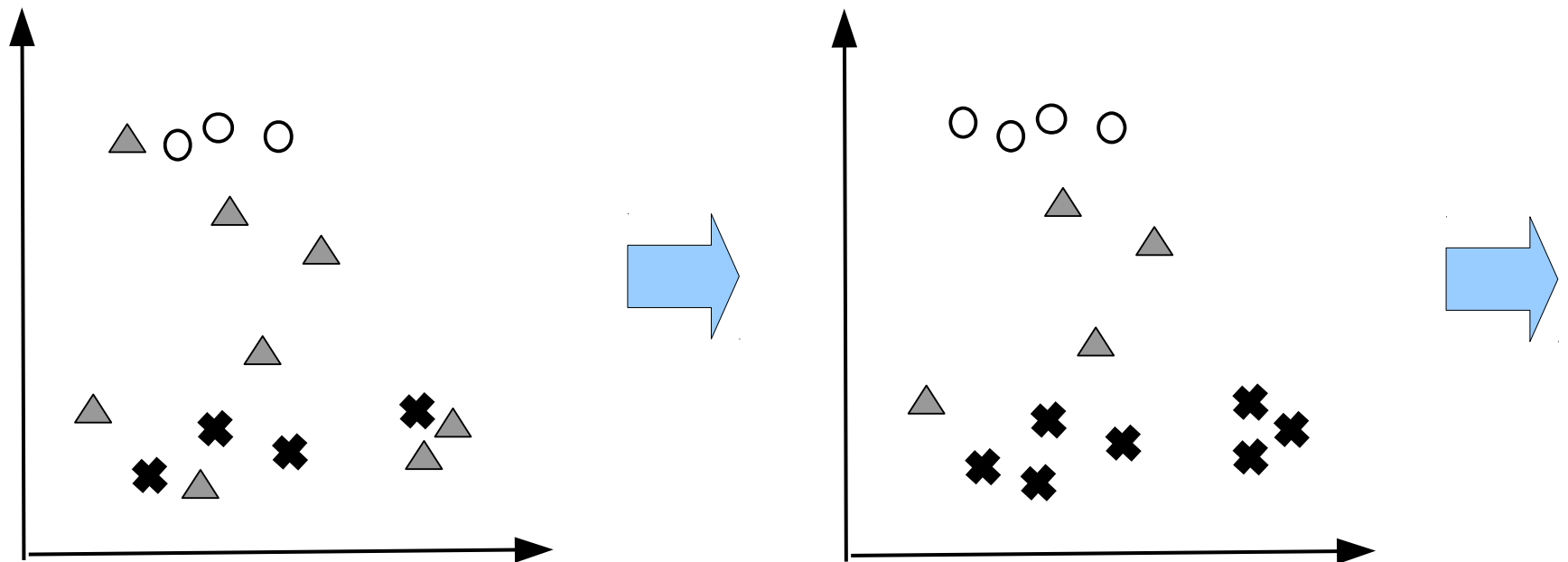
- 自己学習のアルゴリズム

1. 正解付きデータで初期識別器を作成

2. 正解なしデータの識別結果のうち、確信度の高いものを、正解付きデータとみなす

3. 新しい正解付きデータで、識別器を学習

4. 2, 3 を繰り返す

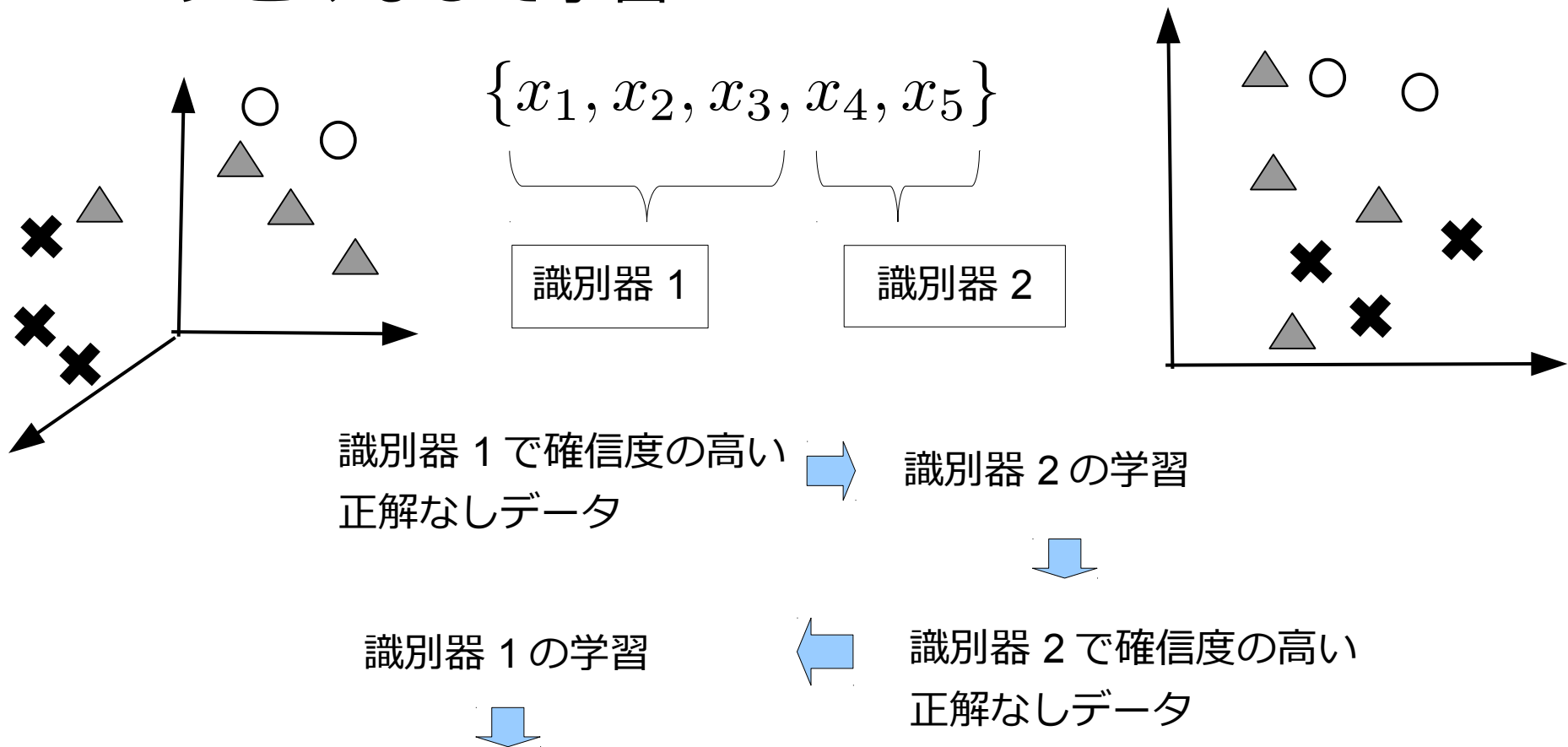


13.2 自己学習

- 自己学習の性質
 - クラスタ仮定や低密度分離が満たされるデータに対しては、高い性能が期待できる
 - 低密度分離が満たされていない場合、初期識別器の誤りが拡大してゆく可能性がある

13.3 共訓練

- 共訓練とは
 - 判断基準が異なる識別器を交互に用いる
 - 片方の確信度が高いデータを、相手が正解付きデータとみなして学習



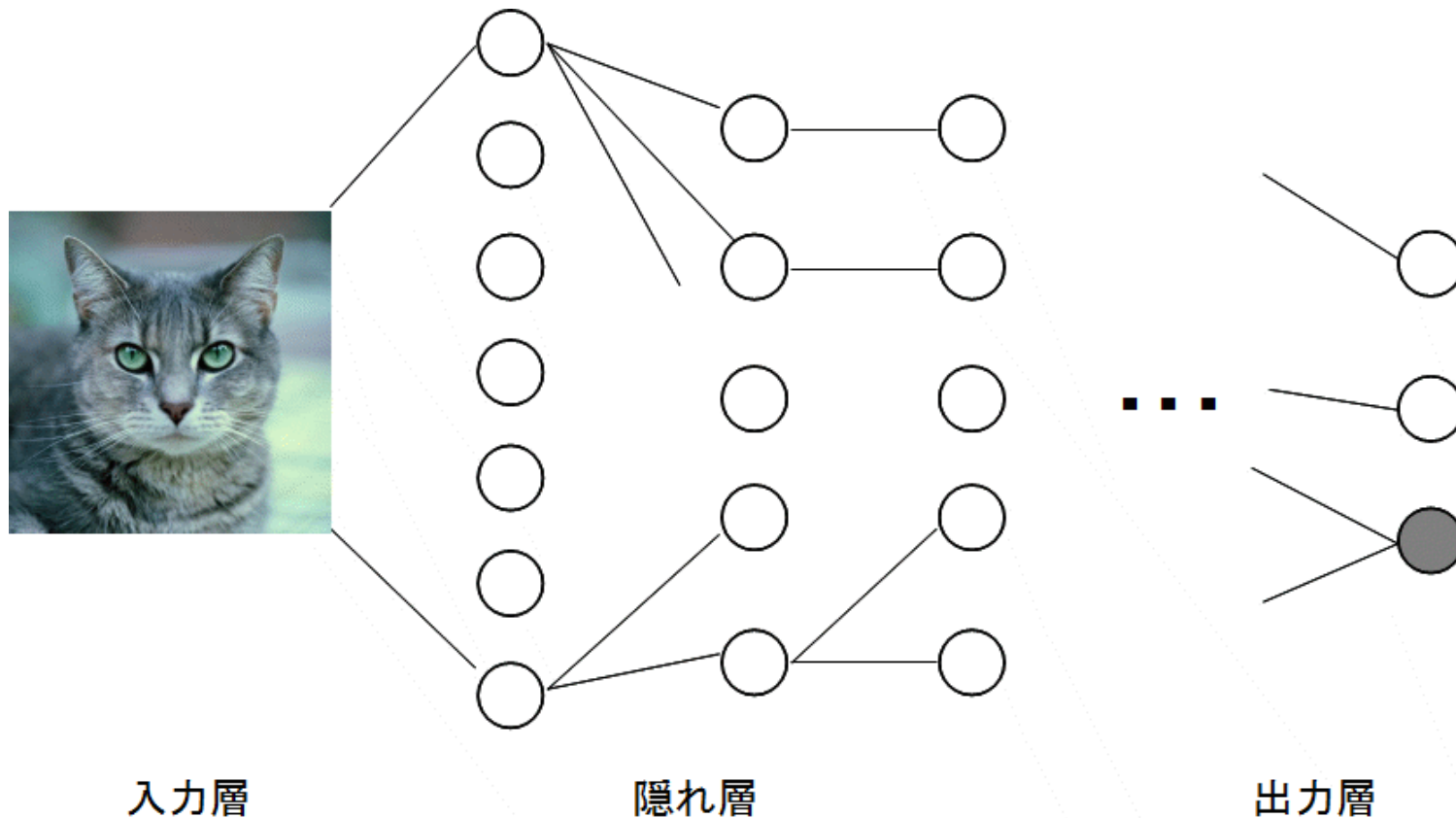
13.3 共訓練

- 共訓練の特徴
 - 学習初期の誤りに対して頑健
- 共訓練の問題点
 - それぞれが識別空間として機能する特徴集合を、どのようにして作成するか
 - 全ての特徴を用いる識別器よりも高性能な識別器が作成できるか

15 章 深層学習

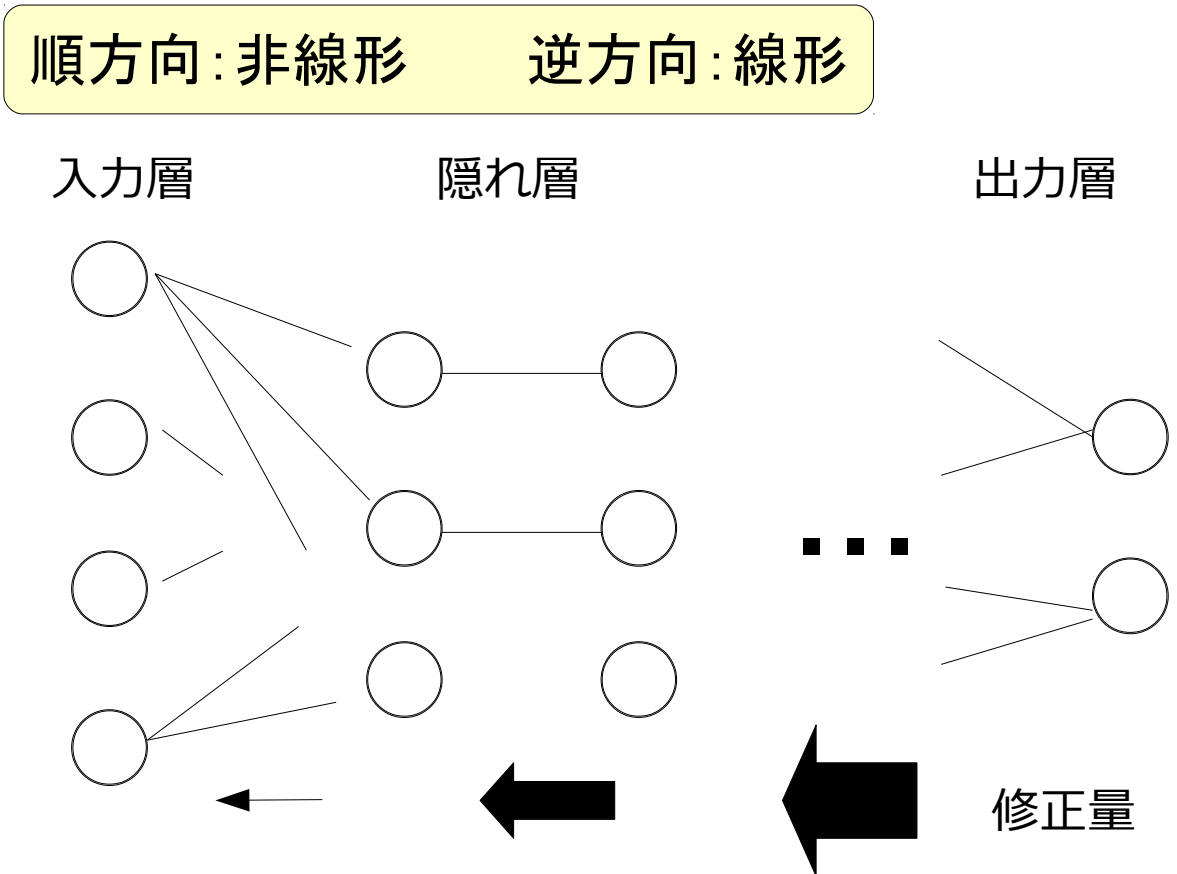
15.1 深層学習とは

- 深層学習の定義のひとつ
 - 表現学習：抽出する特徴も学習する



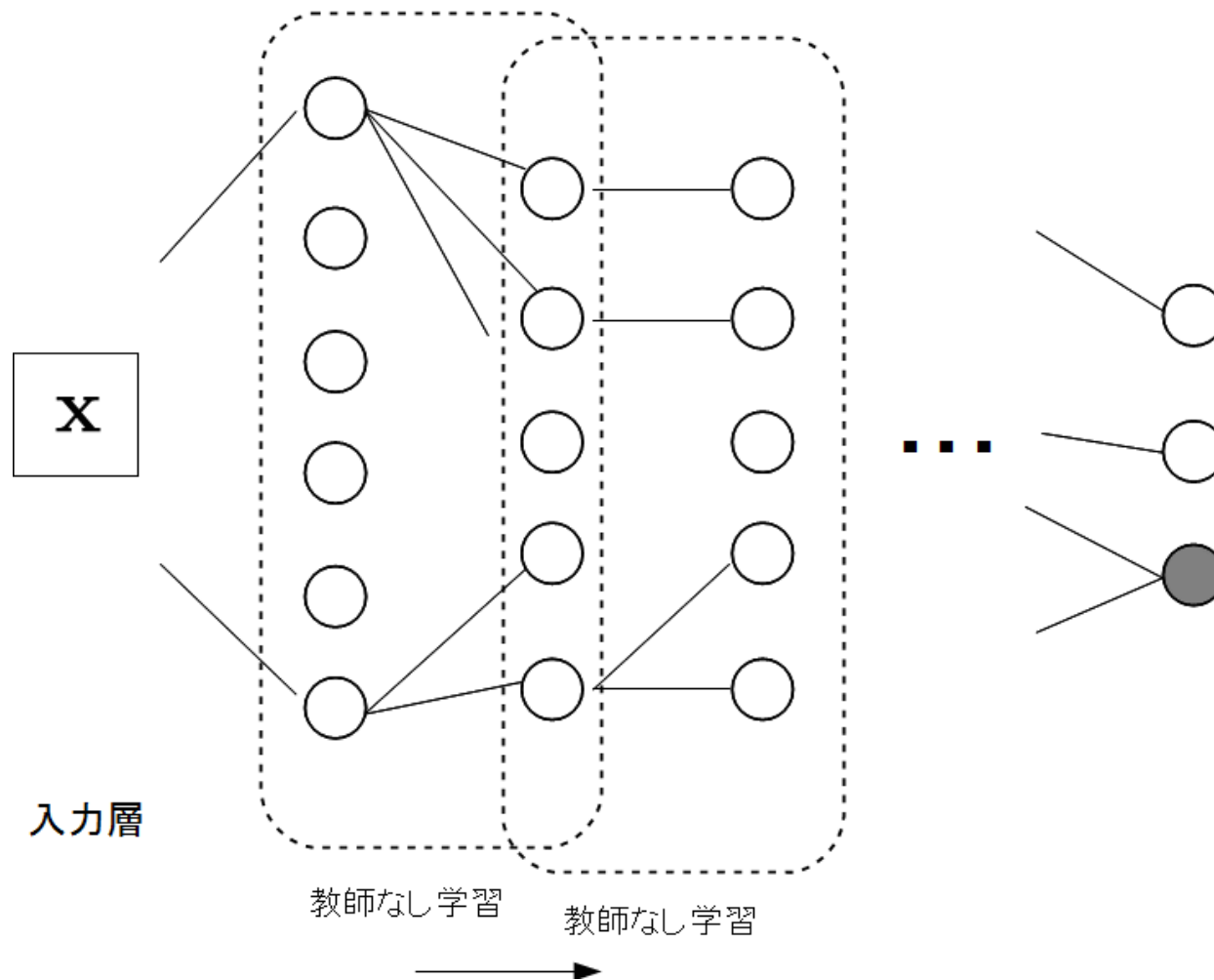
15.2 多階層ニューラルネットワークの学習

- 問題点
 - 修正量が消失／発散する



15.2 多階層ニューラルネットワークの学習

- 事前学習法のアイデア
 - 深層学習における初期パラメータ学習



15.2 多階層ニューラルネットワークの学習

Algorithm 15.1 深層学習のアルゴリズム

入力: 正解付学習データ X

出力: L 層 DNN

$F \leftarrow X$

for $l = 1$ to $L - 1$ **do**

単層特徴抽出器の学習 $\hat{\Lambda} = \arg \min_{\Lambda} h(\Lambda; F)$

単層特徴抽出器から DNN のパラメータ抽出: $\mathbf{W}_l \leftarrow \hat{\mathbf{W}}(\hat{\Lambda}), \mathbf{b}_l \leftarrow \hat{\mathbf{b}}(\hat{\Lambda})$

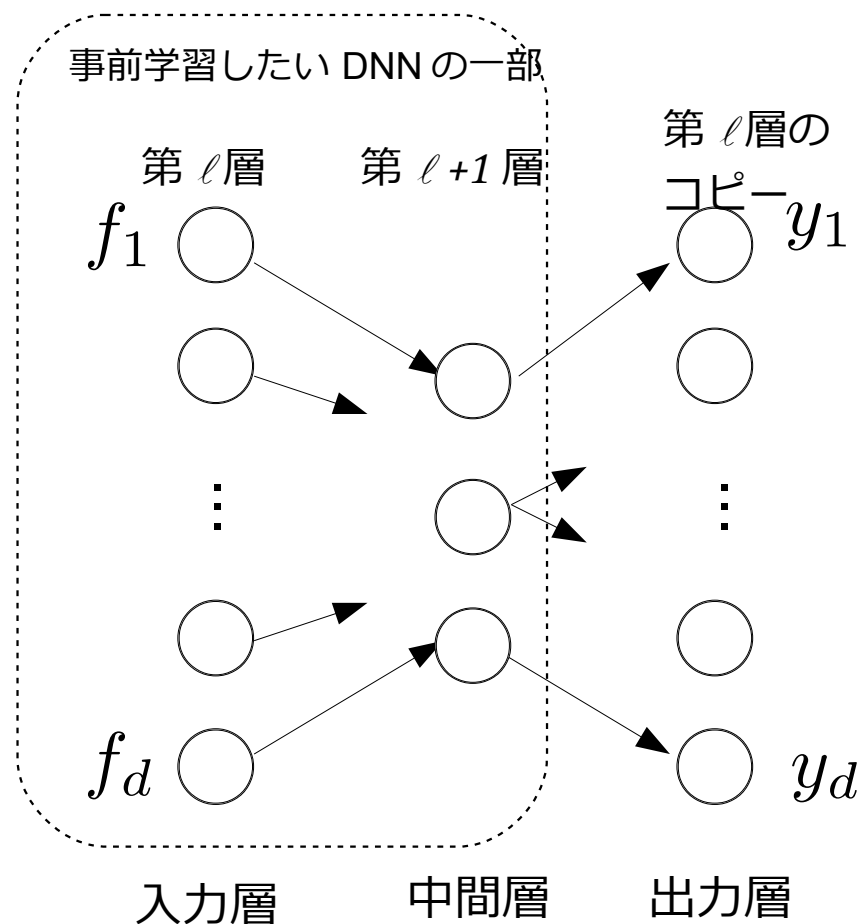
単層特徴抽出器の適用: $F \leftarrow \Phi(F)$

end for

softmax 法などで最上位層を含めた誤差逆伝播法による学習

15.3 Autoencoder

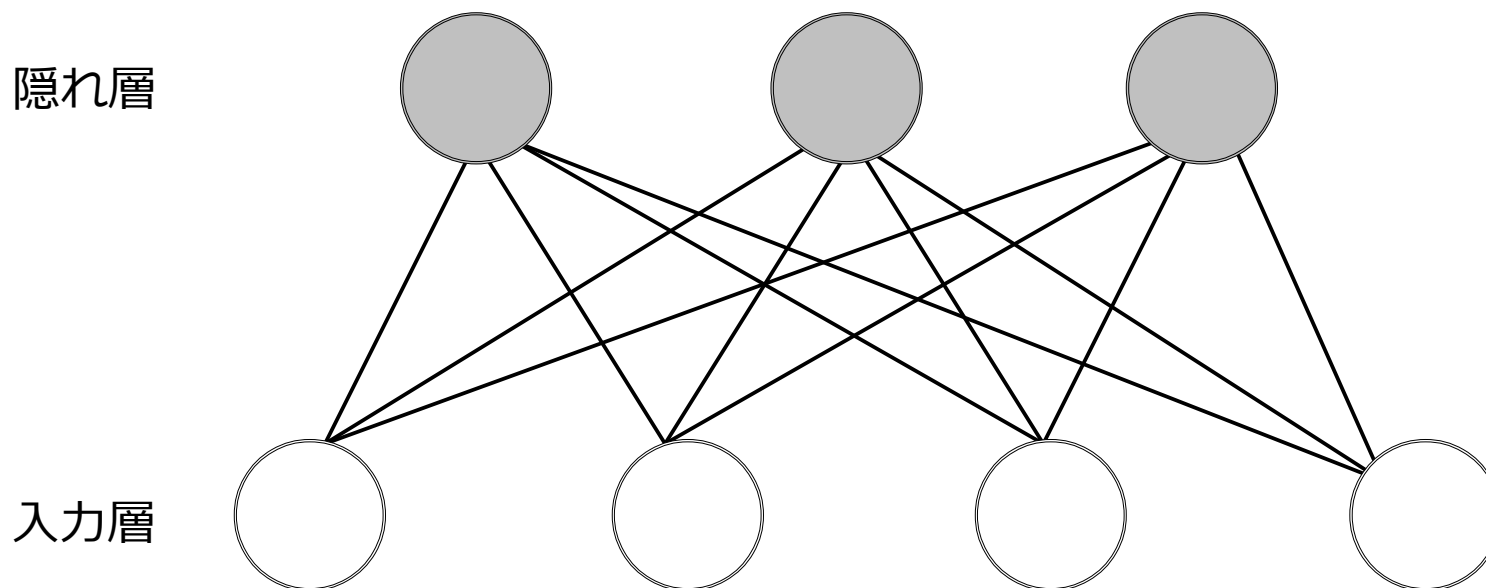
- 事前学習に用いるネットワーク
 - Autoencoder: 自己写像を行う



15.4 RBM

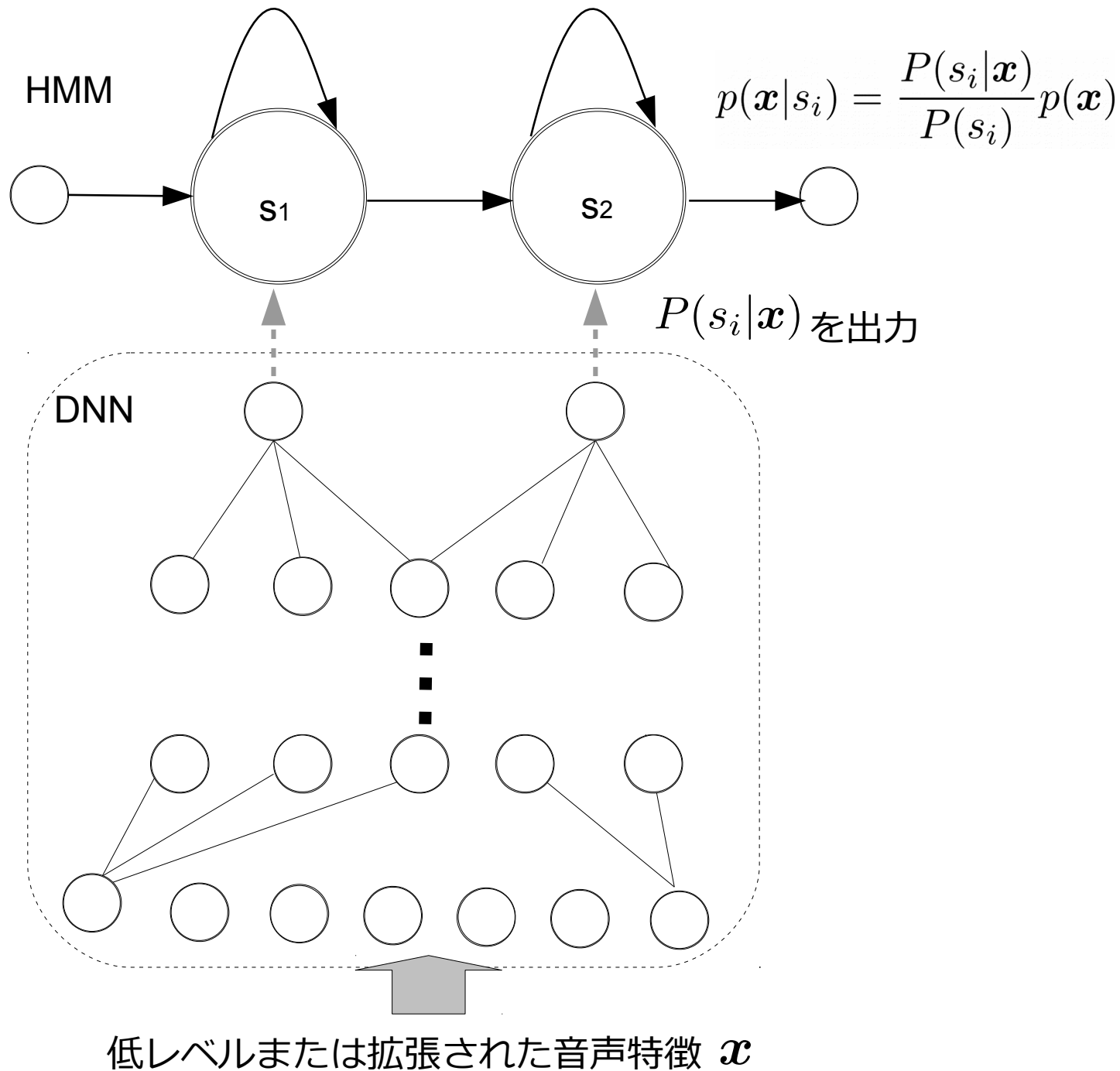
- RBM(Restricted Boltzmann Machine)
 - 生起確率の高い入力 x に対して、エネルギーが高くなるように重み w と閾値 θ を学習

$$\Phi(x|\theta, w) = - \sum_{i \in \Omega} \theta_i x_i - \sum_{(i,j) \in E} w_{ij} x_i x_j$$



15.5 深層学習の応用

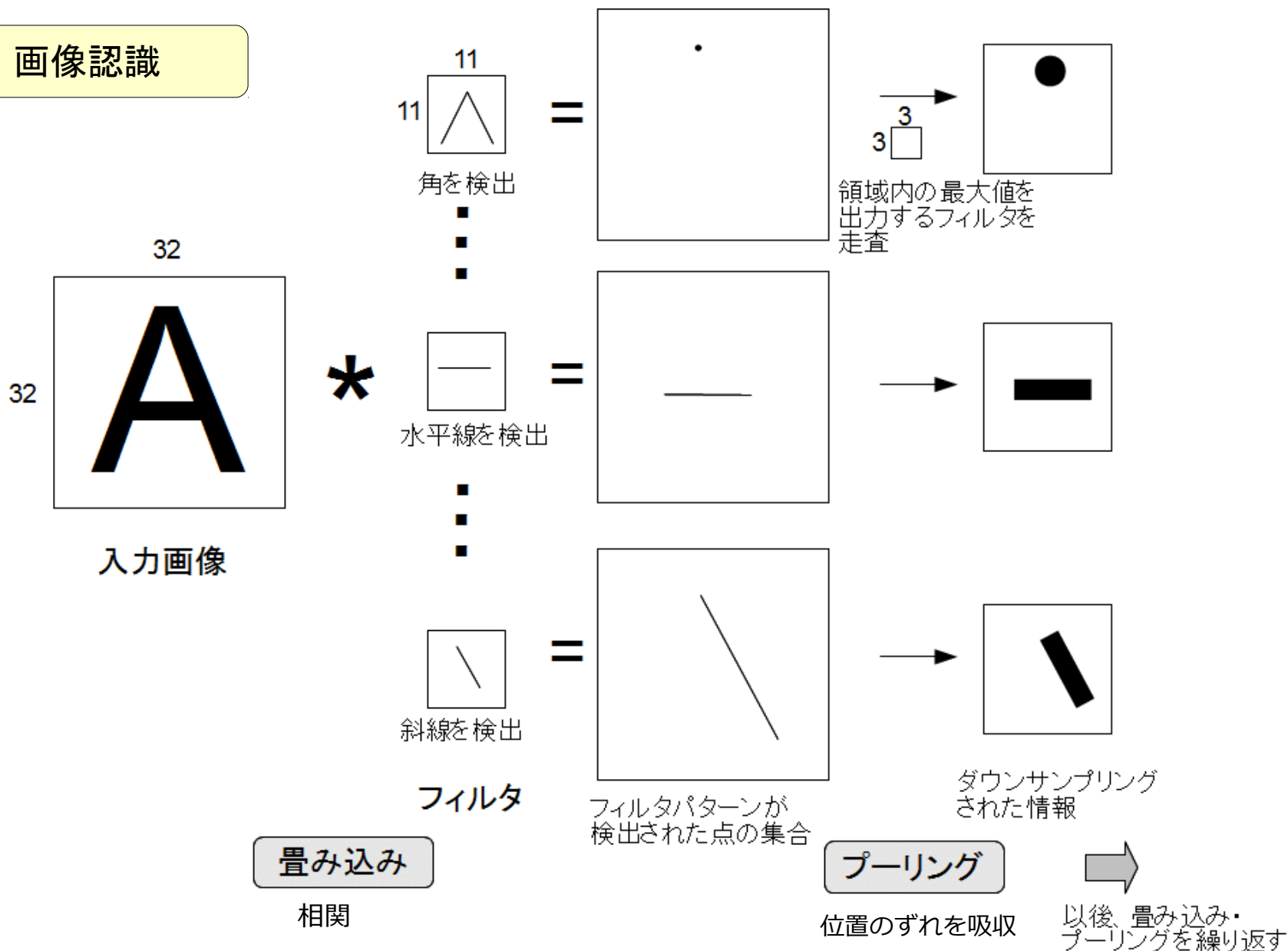
音声認識



15.5 深層学習の応用

畳み込みニューラルネット

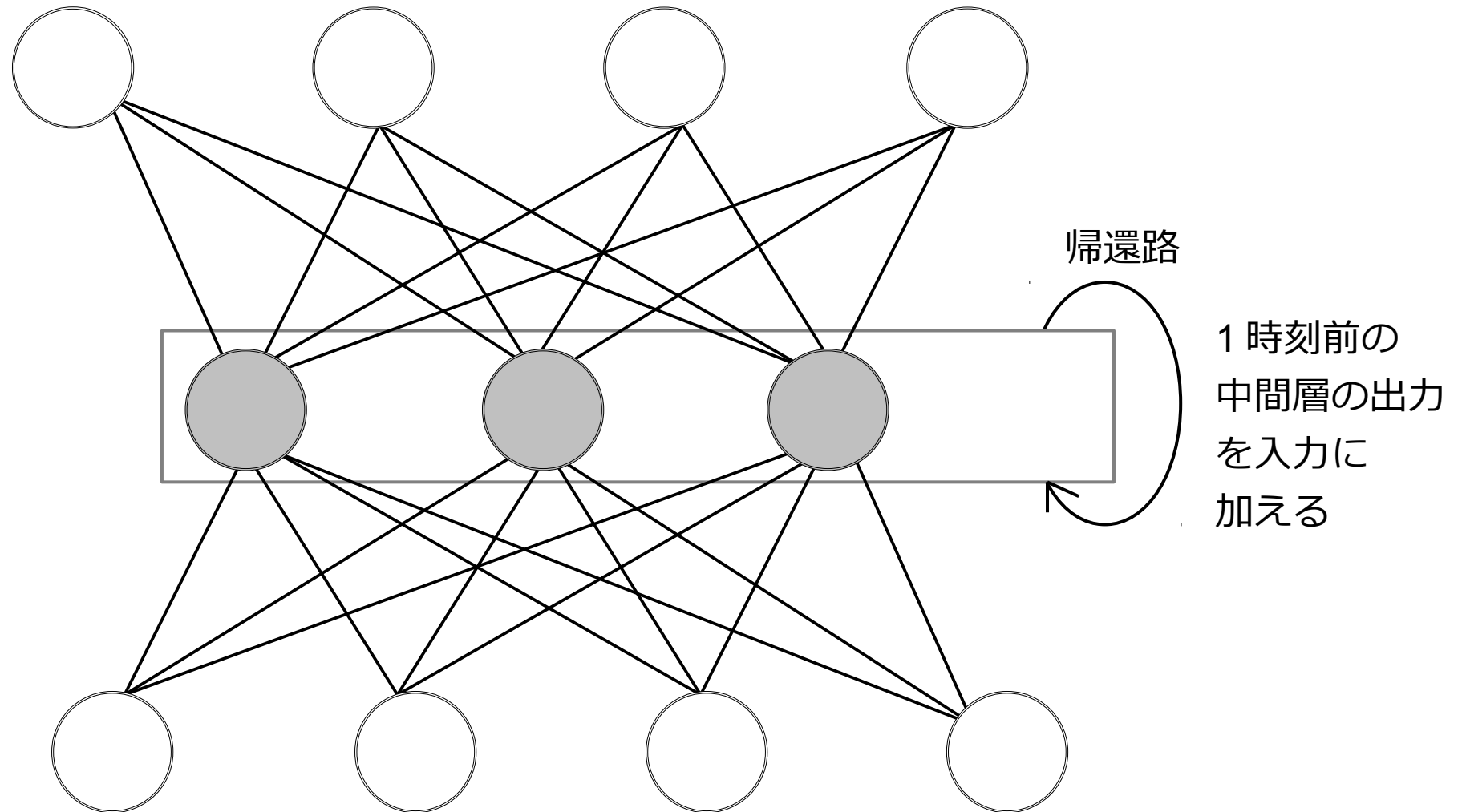
画像認識



15.5 深層学習の応用

自然言語処理

再帰型ニューラルネット



15.5 深層学習の応用

自然言語処理

機械翻訳

