◇本日の講義スケジュール

序章　１３：００～１３：２０
・自己紹介
・時代の流れと、研究環境および研究内容や研究スタイルの変化

第一章　１３：２０～１３：４０（１０分休憩）
化学データサイエンスおよび人工知能の適用イメージ

第二章　１３：５０～１５：００（１０分休憩）
化学データサイエンスおよび人工知能の適用事例：様々なアプローチが可能

第三章　１５：１０～１５：５０（１０分休憩）
化学データサイエンスおよび人工知能の適用手順やパターン

第四章　１６：００～１６：５０
化学データサイエンスおよび人工知能の適用上での留意事項
①化学分野の留意点
②データサイエンス実施上での留意点

まとめと提案　１６：５０～１７：００
①化学データサイエンスおよび人工知能のまとめ
②今後の展開についての「オートノマス創薬」の提案
・KY法の展開（クラス分類と重回帰型）
③自由討論

◇「ケモメトリックス」とは

ケモメトリックスとは：
計量化学（けいりょうかがく、chemometrics）とは、数理科学、統計学、機械学習、パターン認識、データマイニングなどの手法により、（広義の）化学分野における諸問題を解決しようとする分野である。

ウィキペディアより：https://ja.wikipedia.org/wiki/計量化学

＊ちなみに、chemometricsなので、「化学計量学」　by Yuta

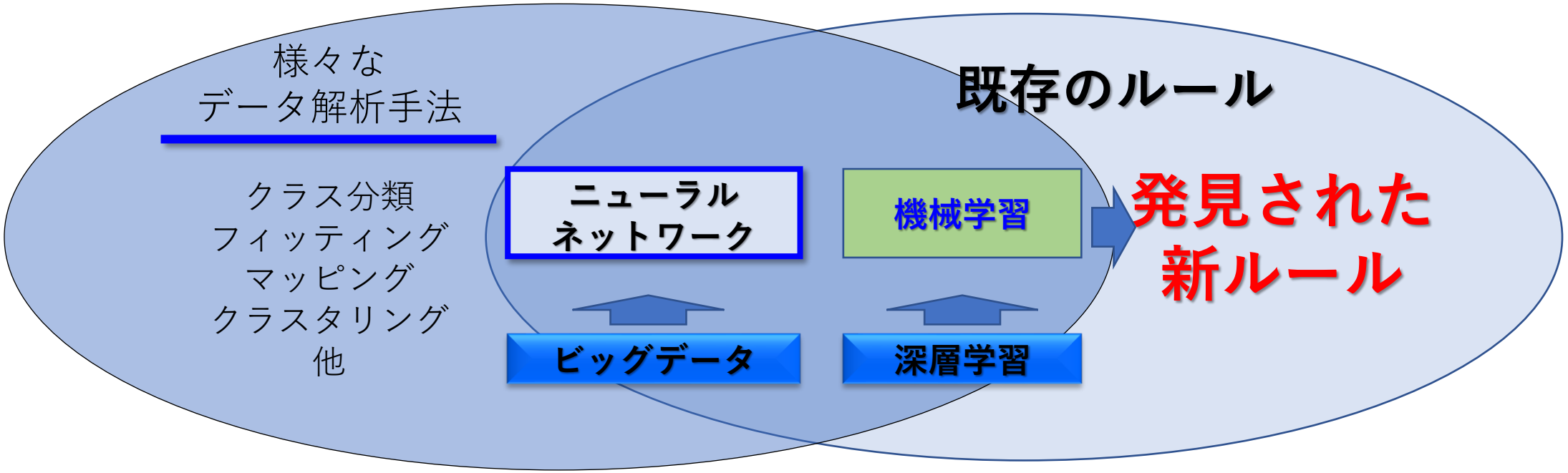
＊化学の分野でデータサイエンスを適用する時の基本的な学問分野となる。

□今後のデータサイエンス手法（DS）と人工知能の展開

現在における多変量解析/パターン認識（DS）と人工知能

多変量解析/パターン認識と人工知能は
機械学習により繋がっている

多変量解析/パターン認識（DS）　　**人工知能**



□本日の解説内容と順番

１．AI（人工知能）に関する概要

２．データサイエンスで利用される化学パラメータ

３．データサイエンス手法の概要

４．KY法の説明

□AI（人工知能）の種類と適用

AIは学習し、内容を**まとめる**ことは出来る。

AIは学習して、**理解**することは困難である。

この場合のAIは「ニューラルネットワーク型」である。

□パーセプトロンからニューラルネットワーク

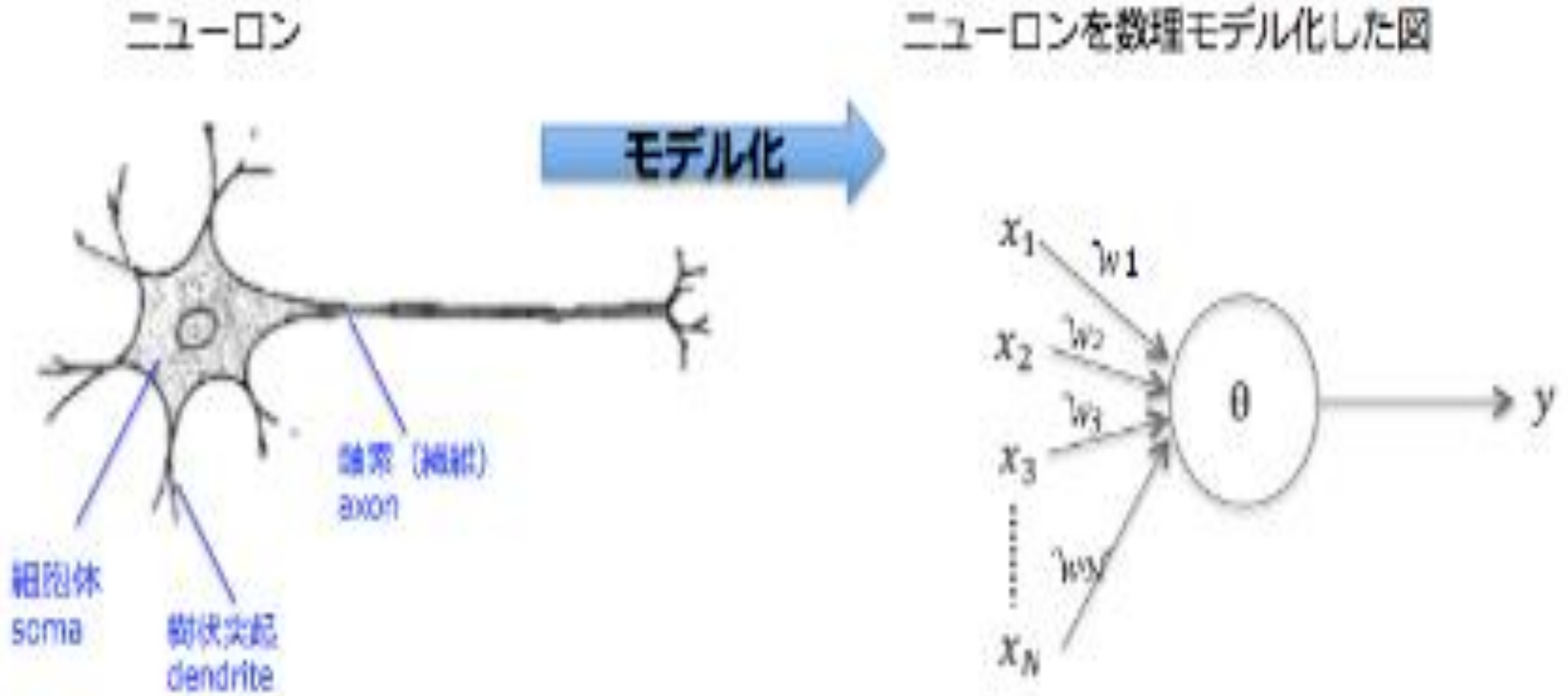
＊ニューラルネットワークは単純なネットワーク構造を利用したパーセプトロンを
　基本として発展
＊パーセプトロンは簡単な二分類問題も解決できないということで、衰退
＊パーセプトロンの限界を打破したアプローチとしてニューラルネットワークが提案

パーセプトロン：線形分類機

ニューラルネットワーク：非線形分類機

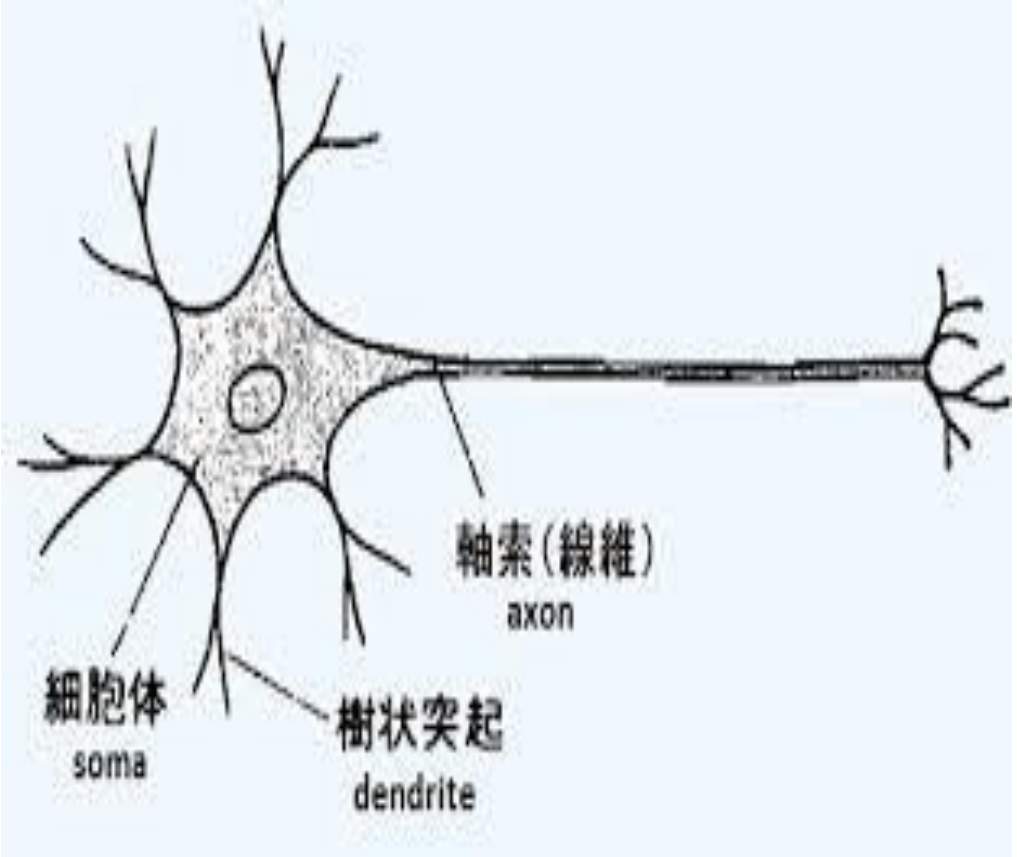
＊パーセプトロンは脳の機能を模したアプローチとして開発され、最終目標は
　　AI(人工知能）への展開であった
＊ニューラルネットワークはその改良系で、ネットワーク構造が多層構造となった
＊最近の深層学習はニューラルネットワークのネットワーク構造をさらに複雑にした

ロパーセプトロン

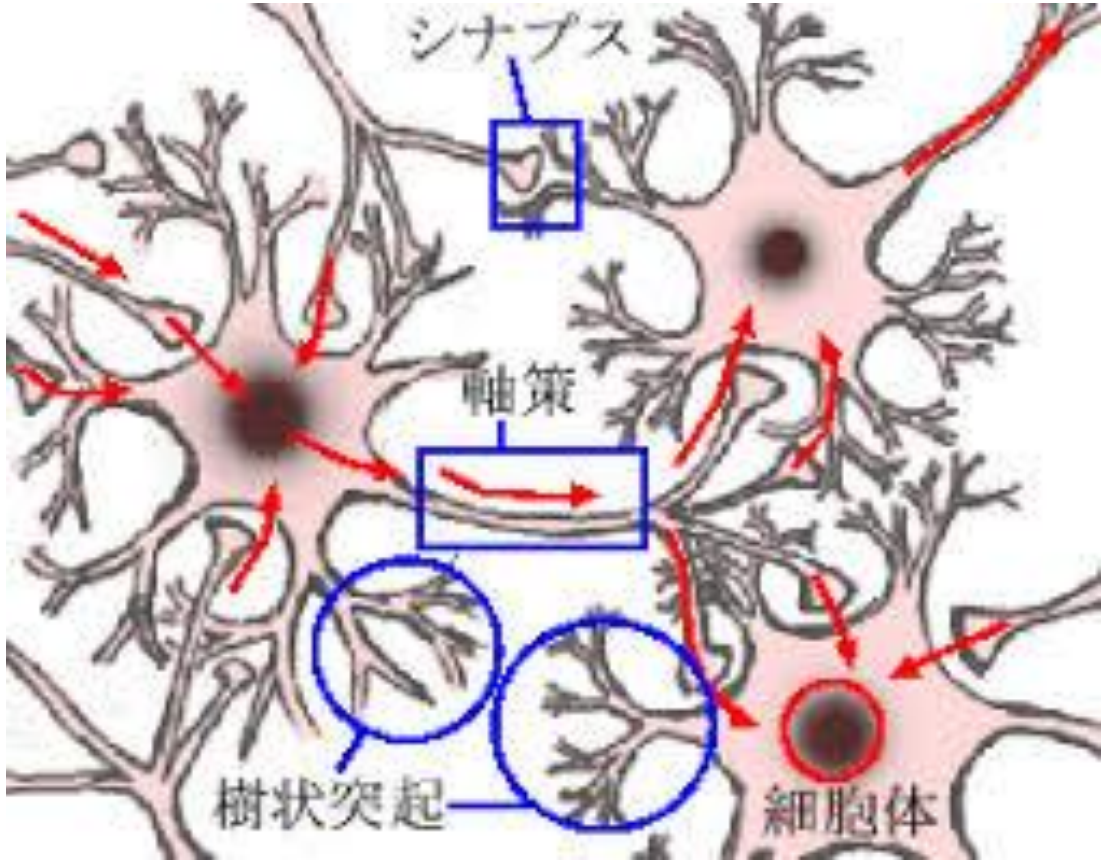


https://udemy.benesse.co.jp/ai/neural-network.html

単ニューロンモデル



ネットワークニューロンモデル



http://www.tamagawa.ac.jp/teachers/aihara/kouzou.html

http://www.sys.ci.ritsumei.ac.jp/project/theory/nn/nn.html

パーセプトロン

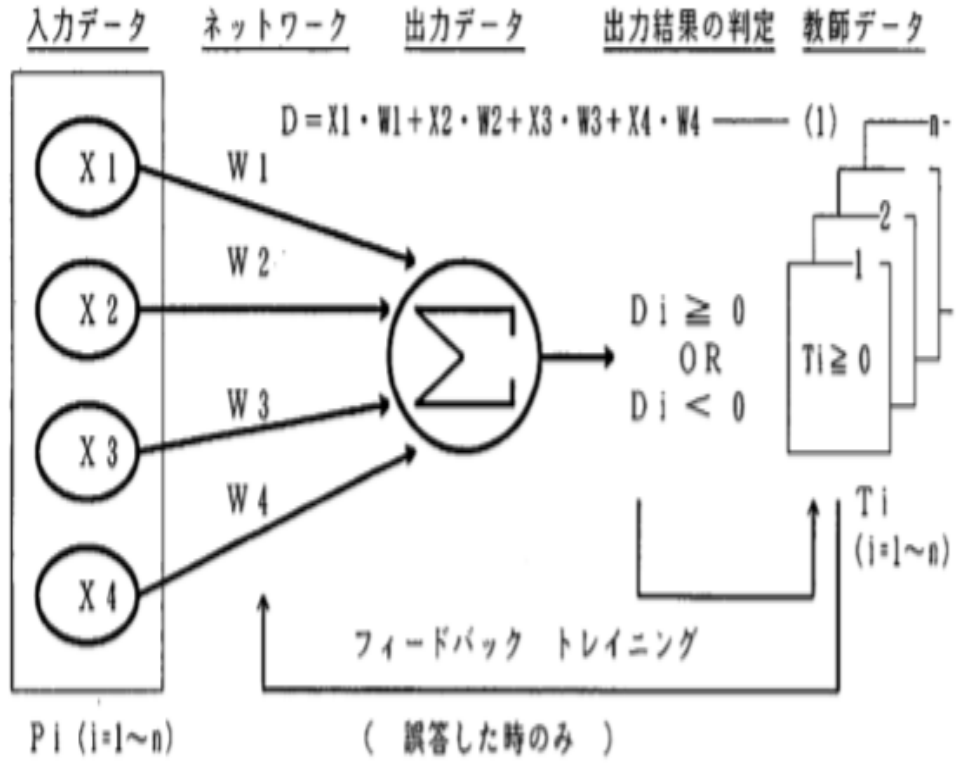
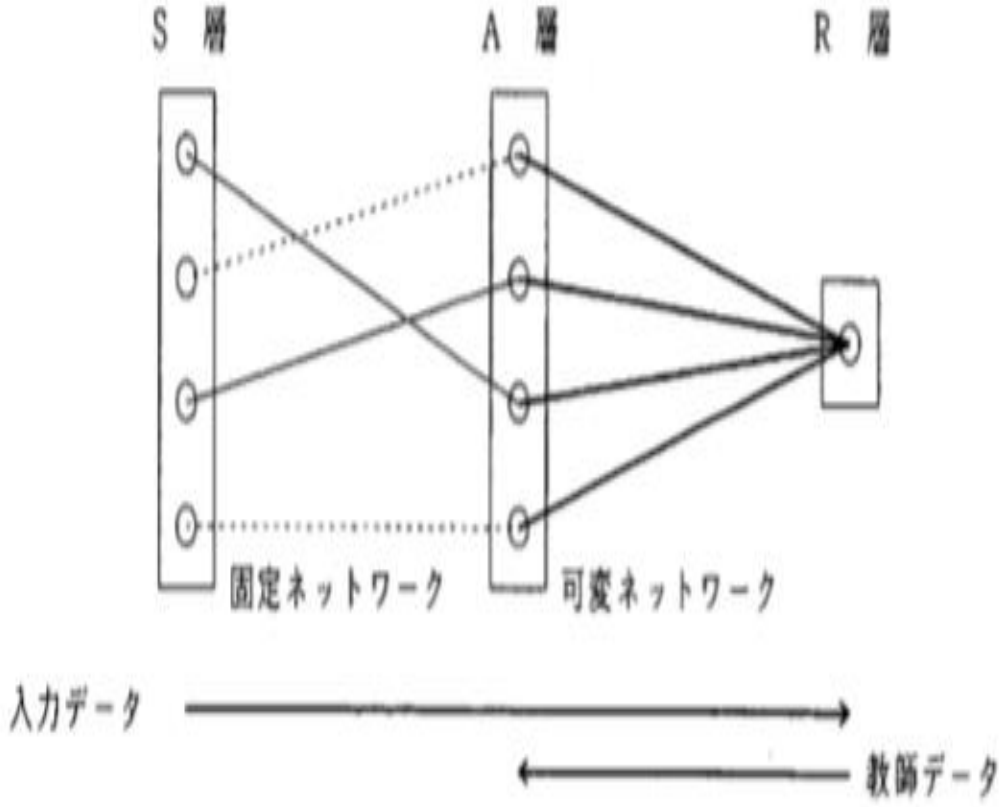


図2．パーセプトロンの"学習"の流れ

単ニューロンモデル

ニューラルネットワーク



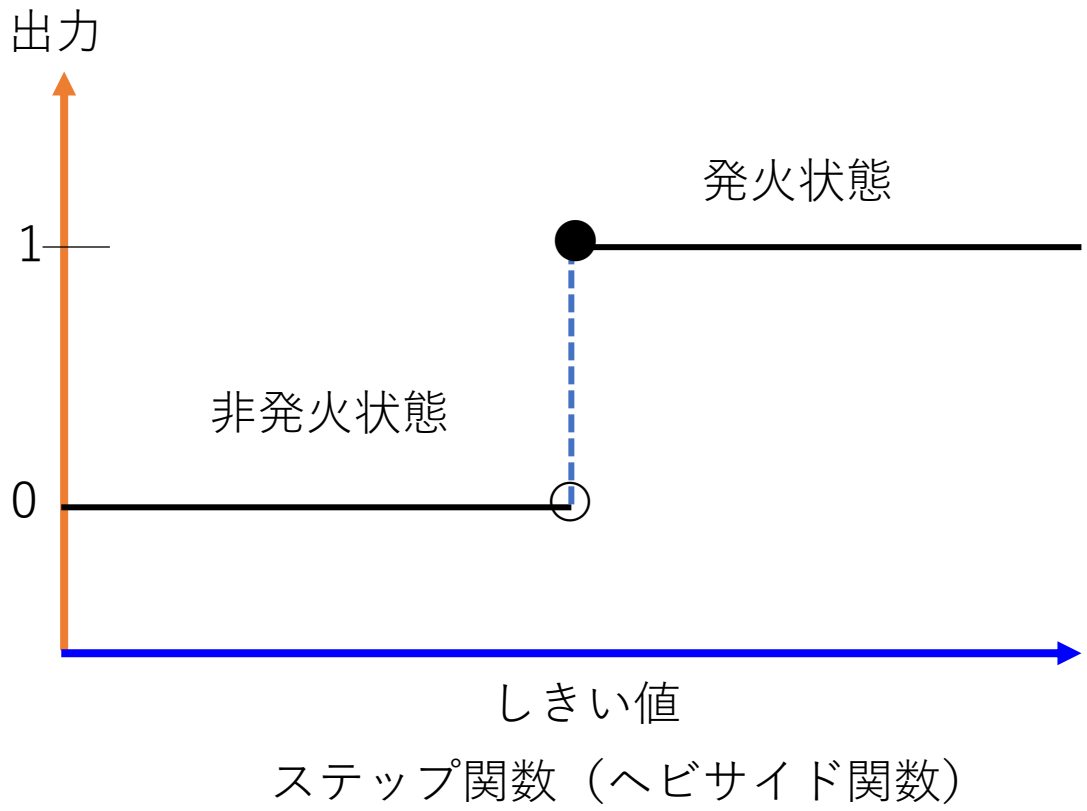
ネットワークニューロンモデル

□形式ニューロンモデル

$$Y = \sigma_i \left(\sum w_i x_i + b \right)$$

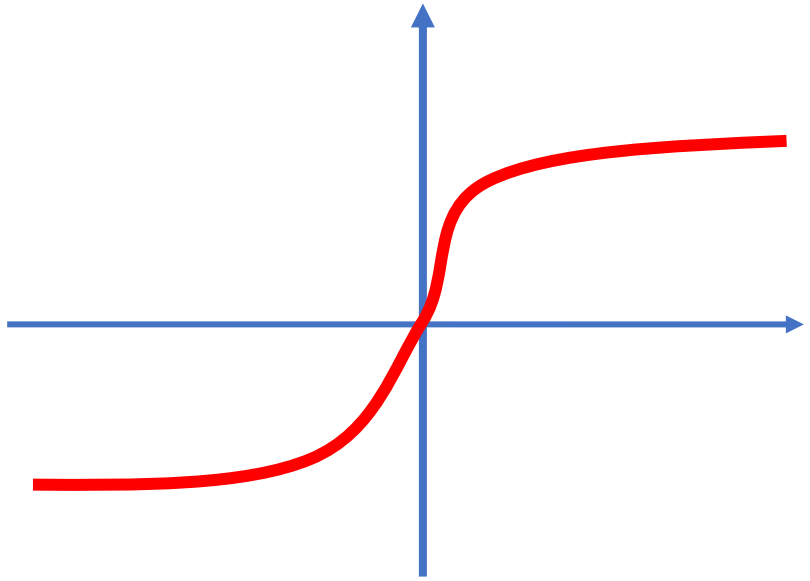
マカロック・ピッツ（McCullochとPitts）の形式ニューロンモデル

活性化関数

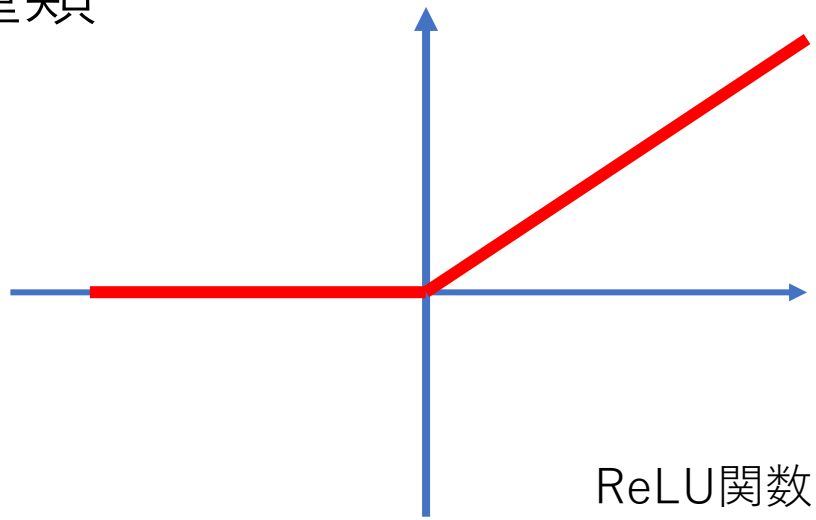


ステップ関数（ヘビサイド関数）

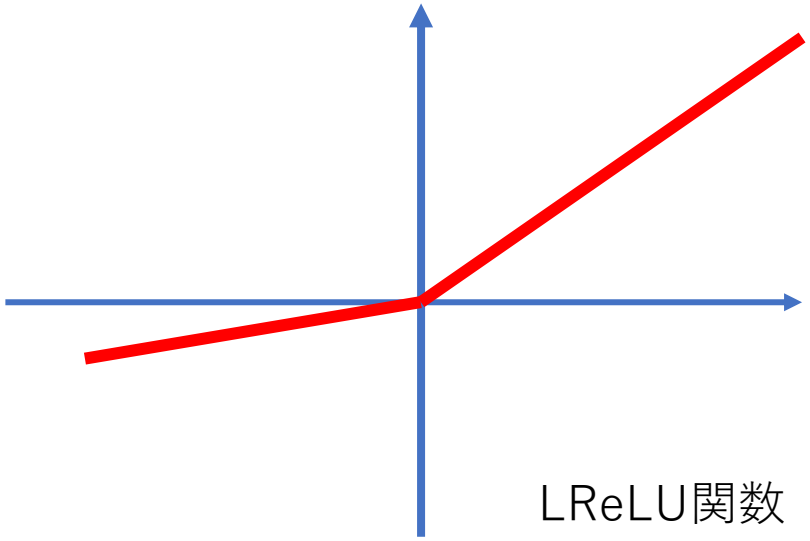
ニューラルネットワークの機械学習で利用されている ステップ関数の種類



シグモイド関数



ReLU関数



LReLU関数

□機械学習型人工知能

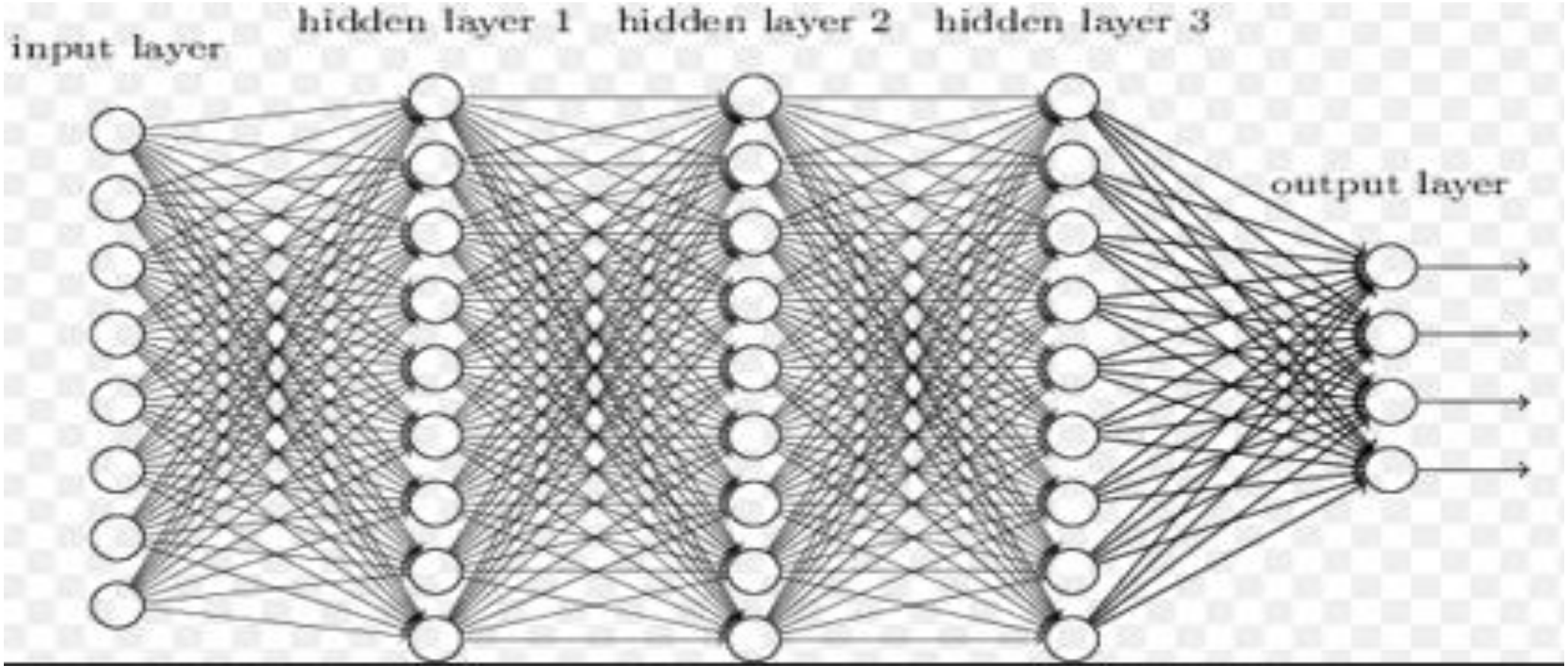
・実行環境：機械学習手法の発展（深層学習：ディープラーニング）

◇ニューラルネットワークから深層学習へ

中間層を増やすことで分類性が向上することは明白であったが、効率的なバックプロパゲーションアルゴリズムが無かったので展開が遅れていた。しかし、効率的に学習結果を前に戻すアルゴリズムが開発されたことで深層学習が開発された。

現在では、中間層が多層になった深層学習が次世代ニューラルネットワーク（人工知能）として脚光を浴びている。

順伝搬型ニューラルネットワーク



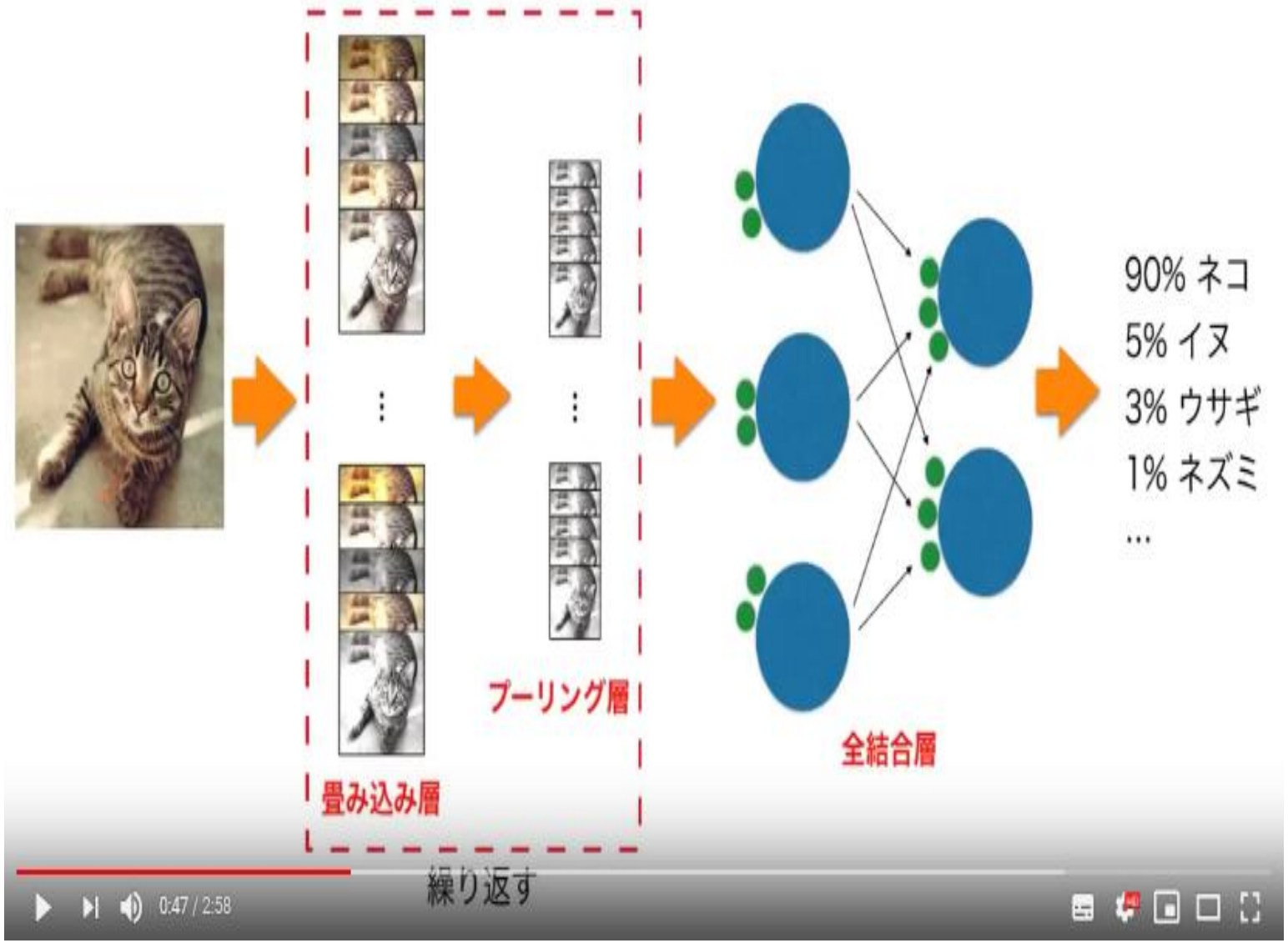
https://nnadl-ja.github.io/nnadl_site_ja/chap6.html

□機械学習型人工知能

・畳み込みニューラルネットワーク

畳み込みニューラルネットワーク（Convolutional Neural Network）

CNNは、全結合層だけでなく
畳み込み層(Convolution Layer)と
プーリング層(Pooling Layer)から
構成されるニューラルネットワーク。



https://www.youtube.com/watch?v=Kr84vlulqTA

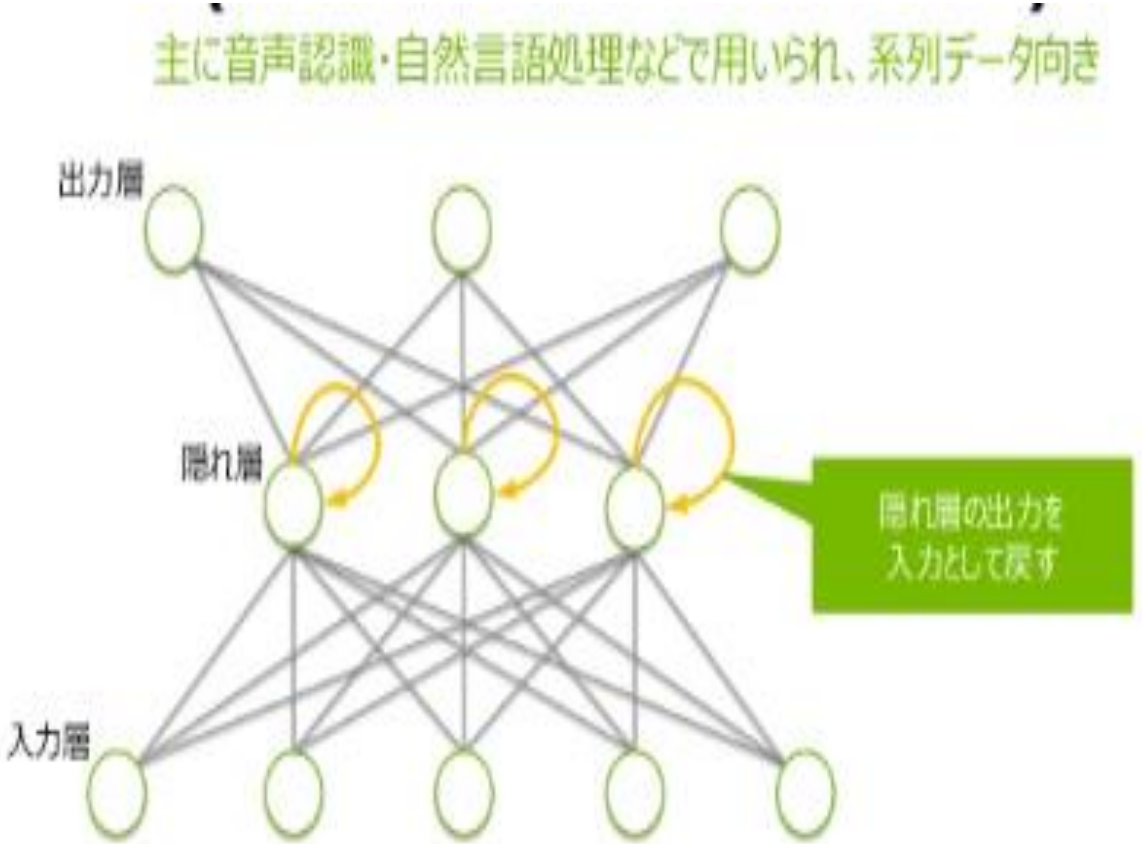
□機械学習型人工知能

・開発歴：リカレント（再帰型）ニューラルネットワーク（RNN）

◇リカレント（再帰方）ニューラルネットワーク（Recurrent Neural Networks）

・中間層の出力データを再び入力データとして利用することを特徴とする

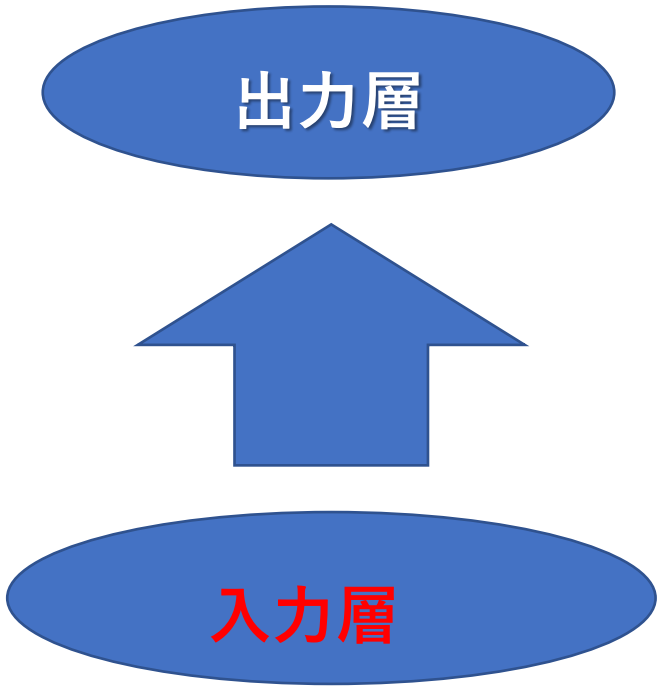
・時系列データの解析や言語処理等に利用される



https://www.slideshare.net/NVIDIAJapan/ss-75109746

□ネットワークの深層化による効果

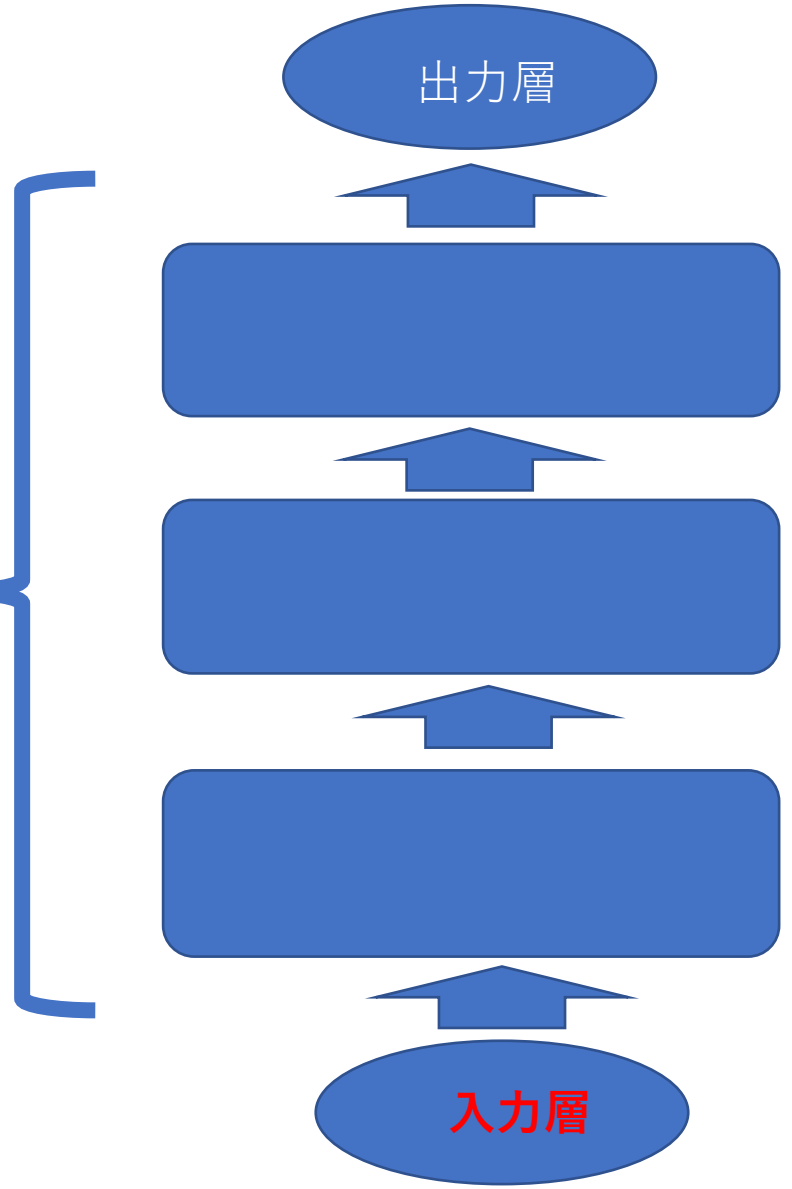
浅層ネットワークと深層ネットワーク

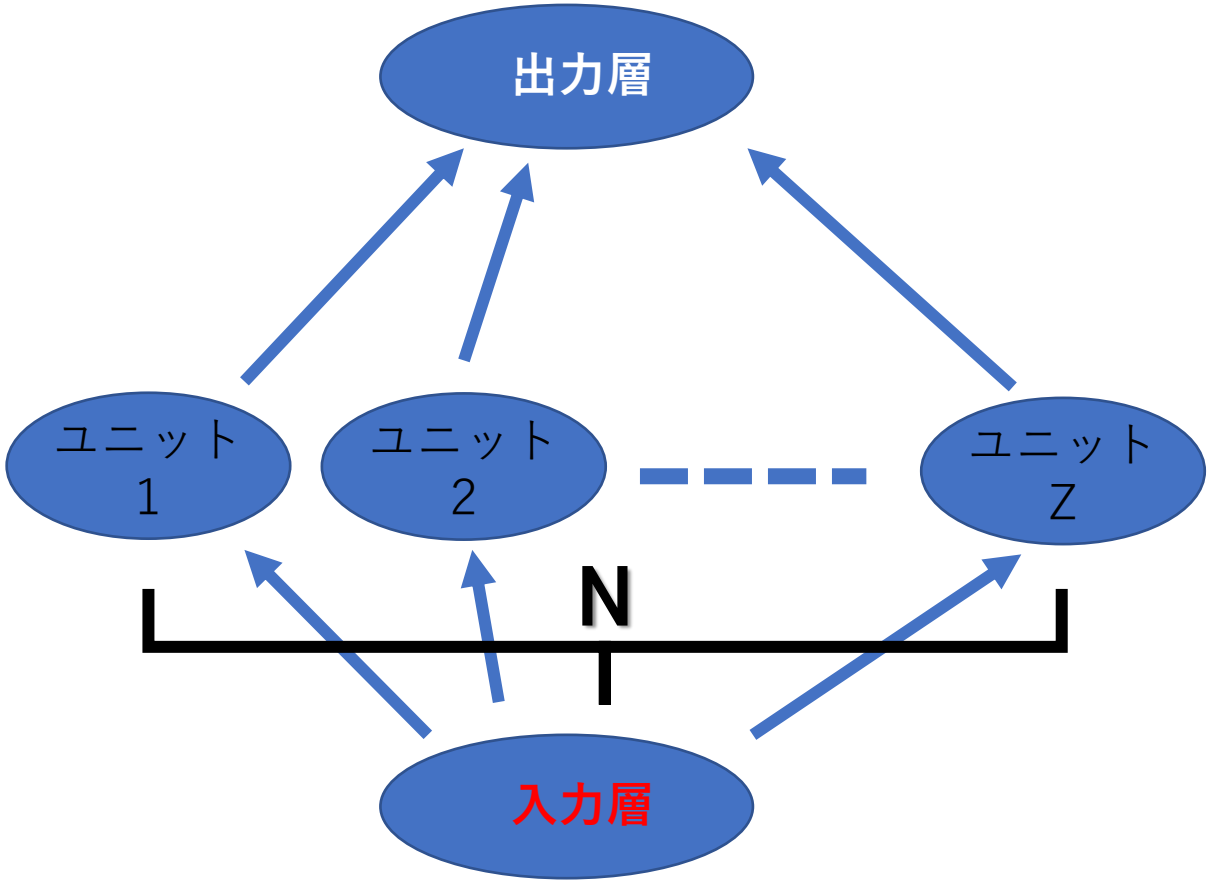


パーセプトロン

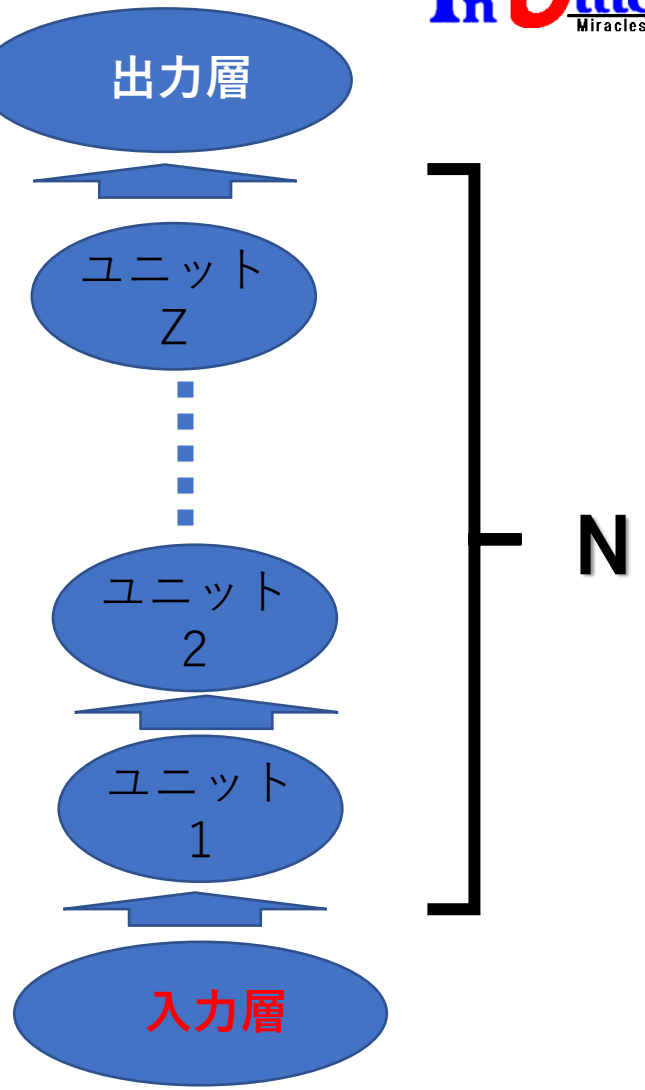
隠れ層

深層学習



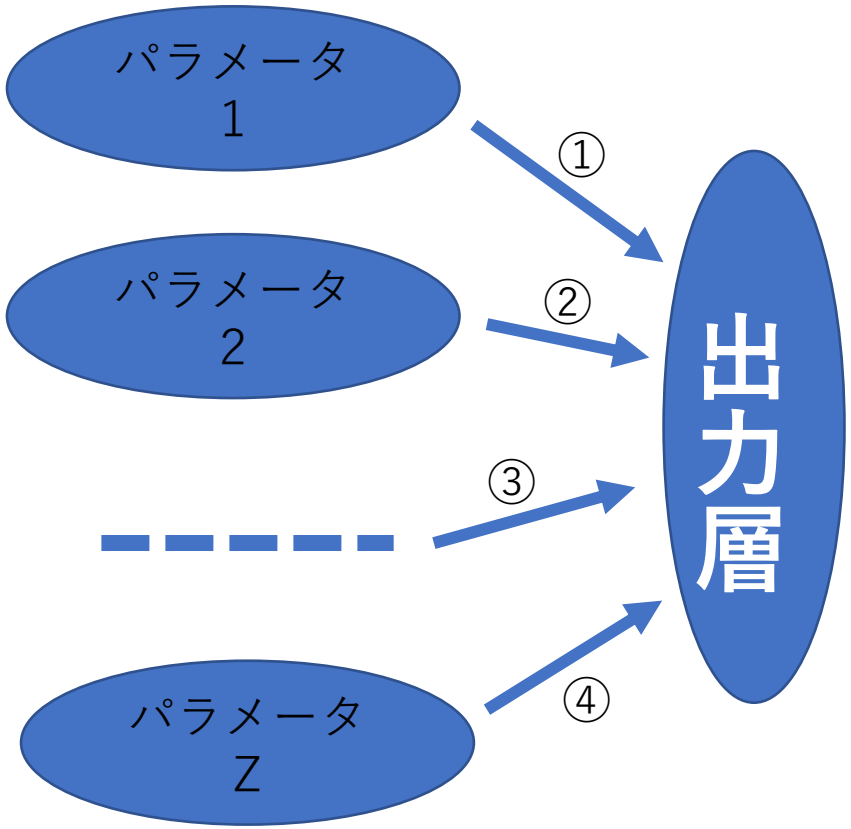


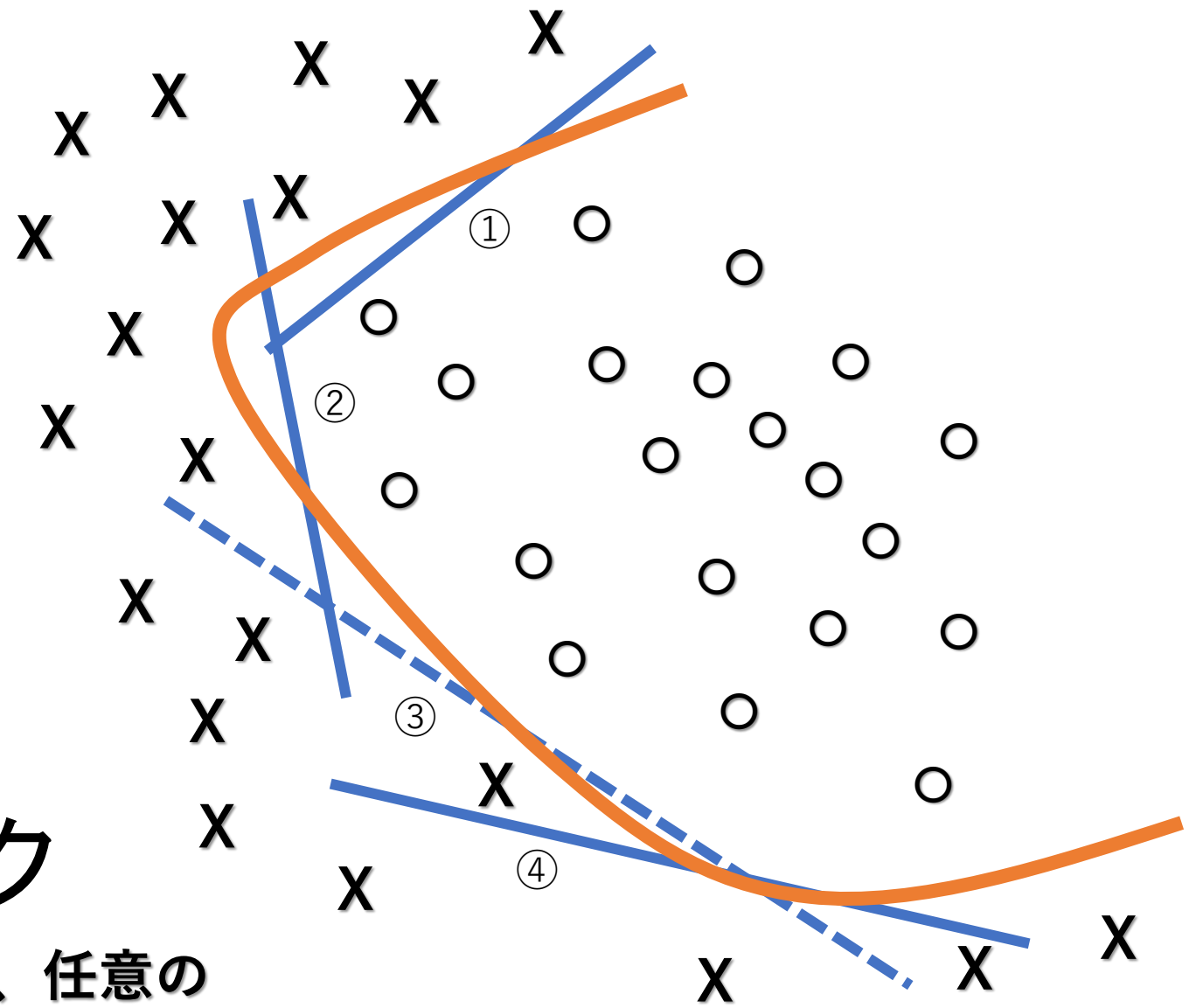
浅層ネットワーク
初期型ニューラルネットワーク



深層ネットワーク

□浅層ネットワークによる非線形化

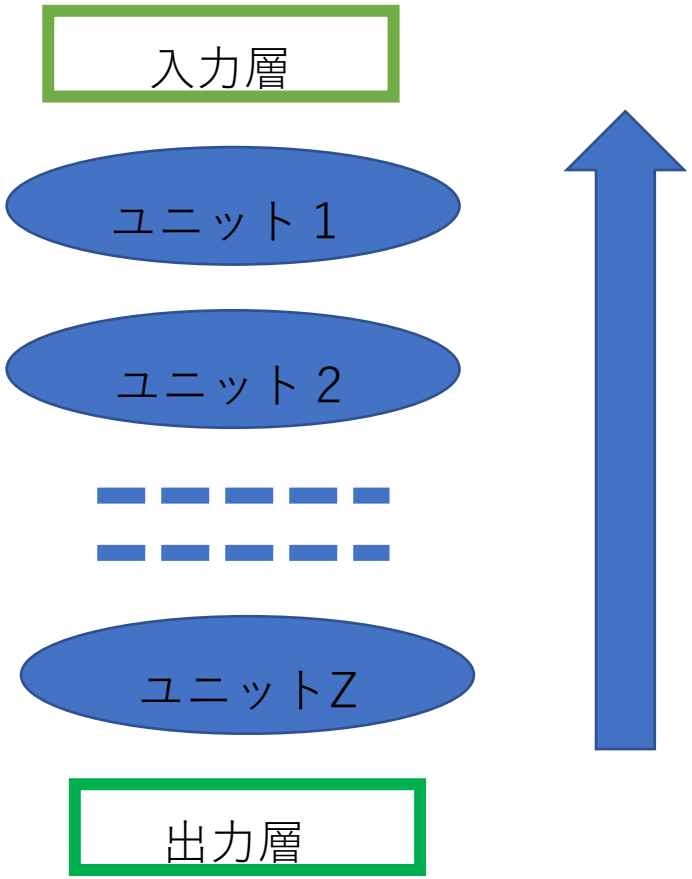


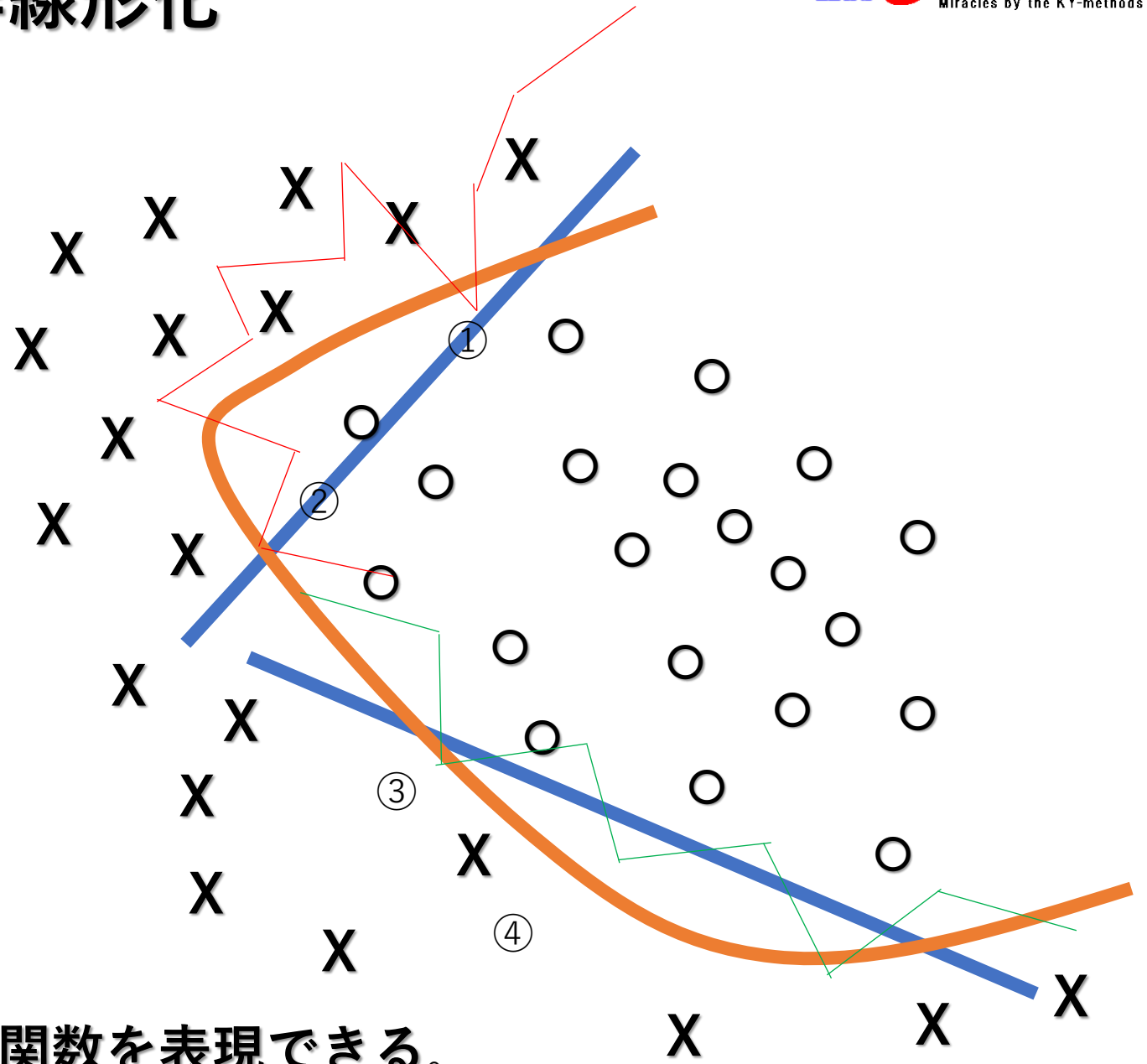


浅層ネットワーク

中間層のユニット数が大きければ、任意の関数を表現できる。非線形性が高い。

□深層ネットワークによる非線形化

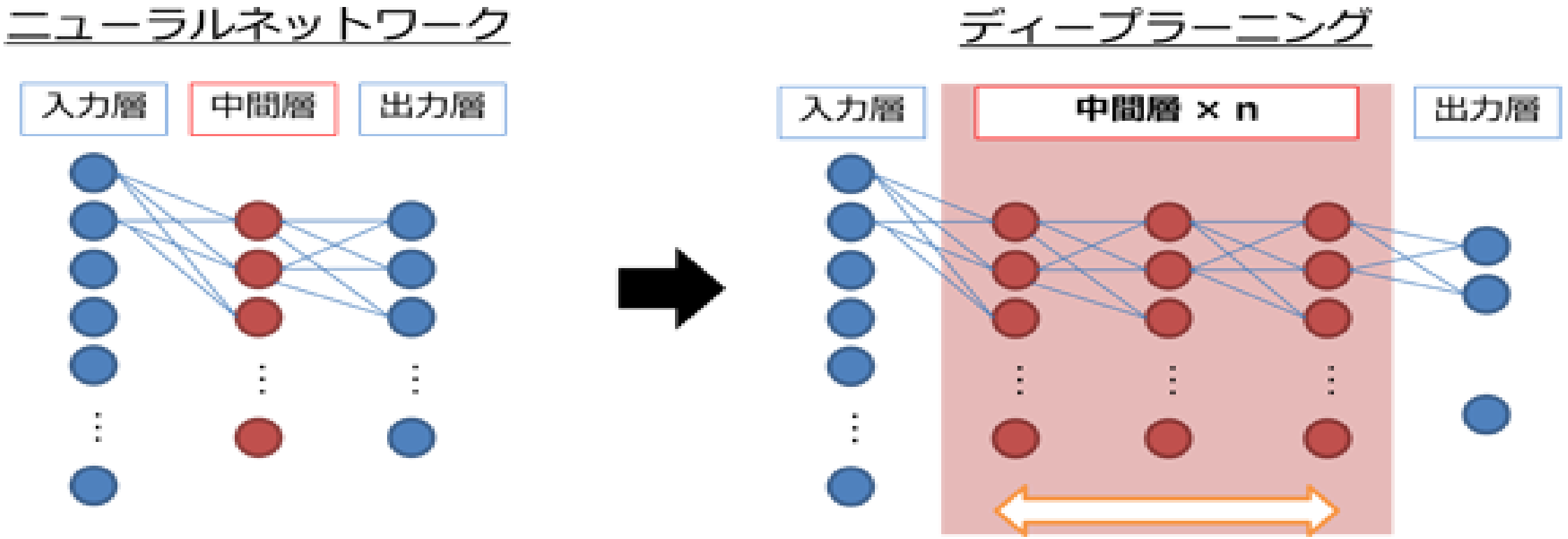




深層ネットワーク

浅層ネットワークよりも遥かに複雑な関数を表現できる。

□機械学習における問題



多層化による**勾配消失**と**過剰適合**の問題があったが、近年、アルゴリズムの改良とデータ量の増大、そして膨大なデータを処理できる計算装置（コンピュータ）の爆発的な性能向上によって問題が解消されてきた。

https://markezine.jp/article/detail/24185より抜粋

In Silico Data
Miracles by the KY-methods

□本日の解説内容と順番

１．AI（人工知能）に関する概要

２．データサイエンスで利用される化学パラメータ

３．データサイエンス手法の概要

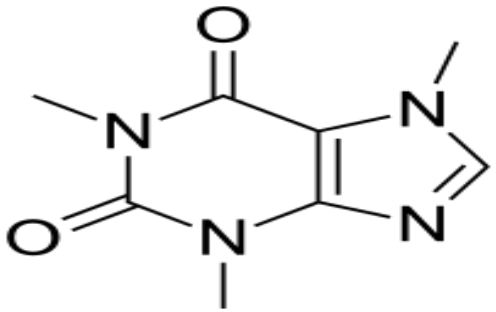
４．KY法の説明

□化学データ（パラメーター）の種類

化合物に起因するデータを「化学データ（パラメーター）」とする
構造式の有無や内容により、以下の３種類に分類される

構造式不要

物性データ
機器スペクトルデータ
医療関連データ



２次元構造式関連データ

分子式関連パラメーター
トポロジカルパラメーター
部分構造パラメーター
フラグメントパラメーター

３次元構造式関連データ

トポグラフィカルパラメーター
３次元構造関連データ
力場関連パラメーター
量子化学関連パラメーター

□化学データ（パラメーター）の種類

◇化合物に由来する化学データ（パラメーター）

化合物関連データ解析を行うにはパラメーター、即ち数値データが必要となる

化合物構造式由来のパラメーター

■ トポロジカルデータ
分子構造インデックス：原子数（原子種）、結合数（結合種）、リング数、その他
様々なインデックス値：ＨＯＳＯＹＡインデックス、分子結合インデックスＭＣ値
パス値インデックス、

■ トポグラフィカルデータ
化合物の３次元的形状に関するパラメータ
化合物全体構造　：ボックスパラメータ、対称パラメータ、
立体格子パラメータ、その他
化合物部分構造　：ステリモルパラメータ、

■ 物理化学データ
分子に関する様々な物性データ　：　分子屈折率、分子量、ＬＯＧＰ、融点、沸点
分子容積、分子表面積、その他
分子軌道法より得られる様々なパラメータ：電子密度、ＨＯＭＯ、ＬＵＭＯ、他
分子力学計算から得られるパラメータ：　種々歪みエネルギー
種々スペクトルより得られるデータ：　種々スペクトルデータ

■ その他のデータ
部分構造パラメータ：　部分構造の有無、部分構造数、
部分構造単位の様々なパラメータ値計算、
演算パラメータ１　：　記述子間の演算により得られるパラメータ（＋－ｘ÷ Log）
演算パラメータ２　：　他の解析手法より算出されたパラメータ
ダミーパラメータ　：　有るパターン存在の有無（１／０）に関するパラメータ

機器スペクトル由来のパラメーター

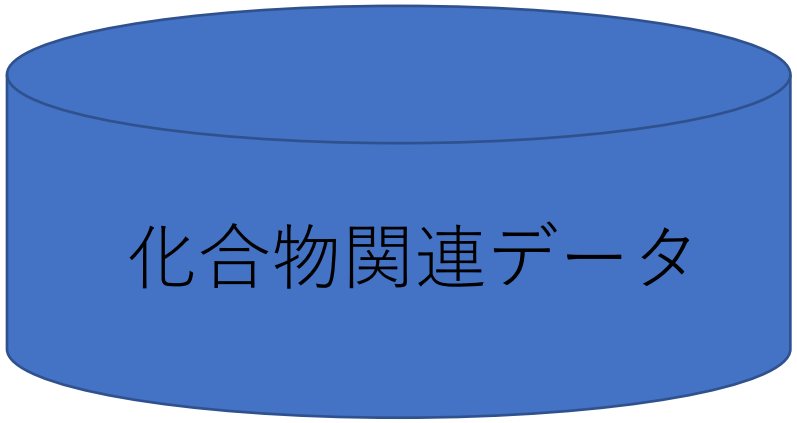
Mass
IR
H-NMR
C-NMR
UV
GC
HPLC
Raman
Ｘ線分析
その他

化合物に起因する解析目的パラメーター

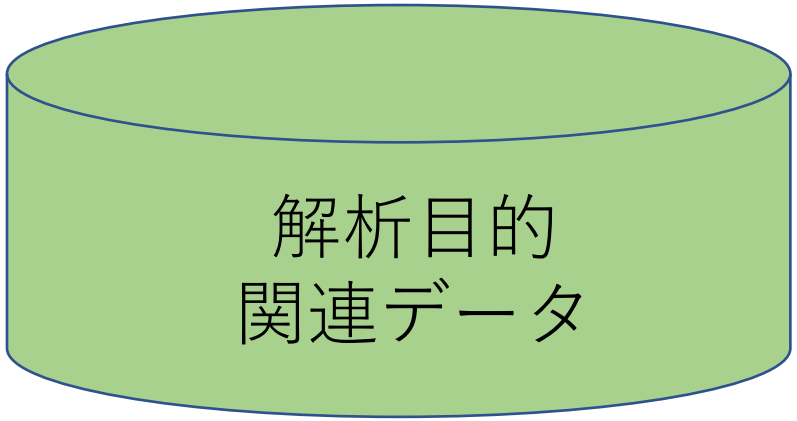
薬理活性
毒性
ADME
物性
環境毒性

□化学データ解析実施上での必要データ

◇汎用的なデータ解析の流れ図：入力データ関連

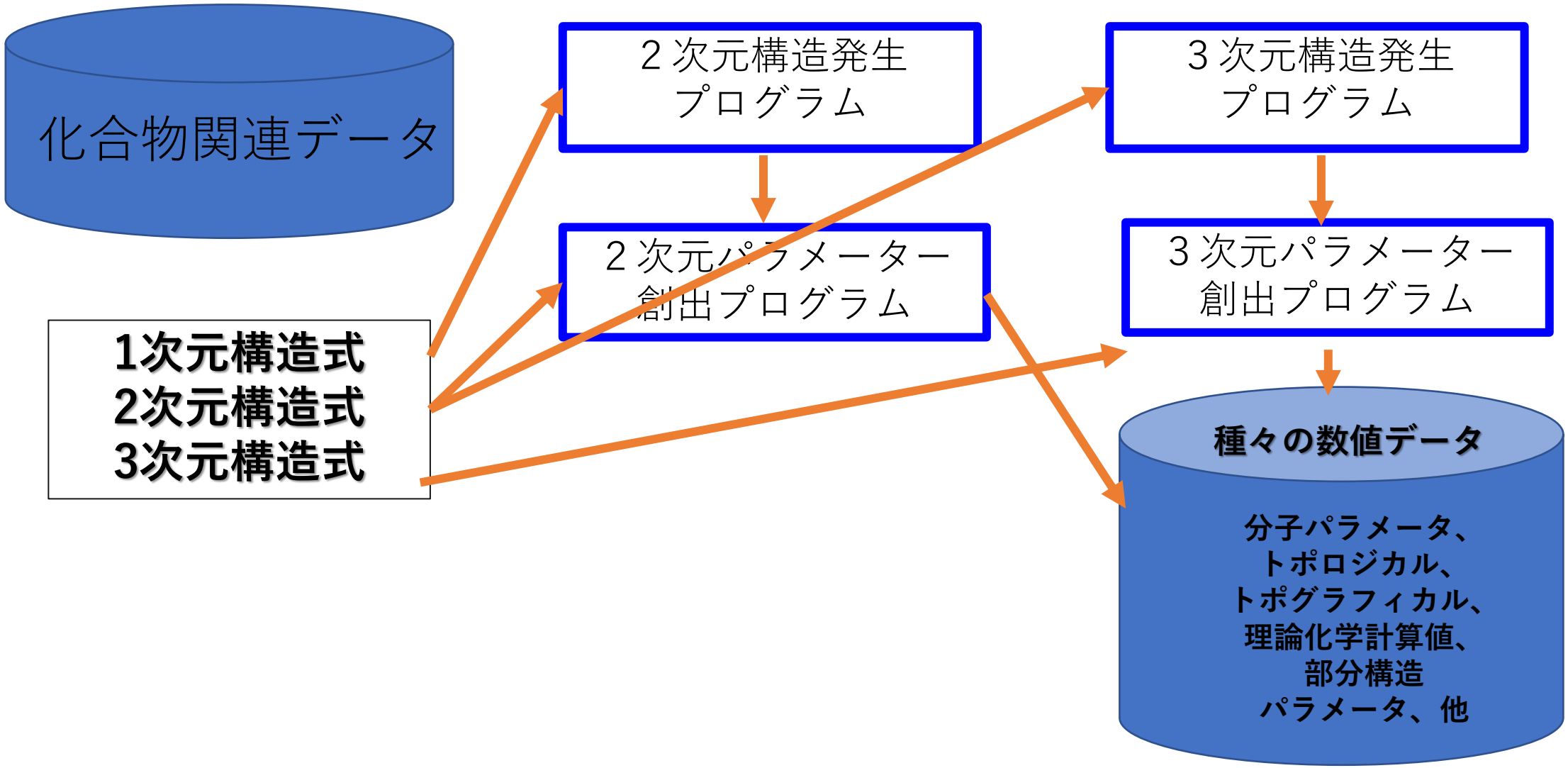


＊化合物構造式（1／2次元）
＊物性データ
＊機器スペクトルデータ
＊バイオ関連データ
＊医療関連データ（画像等）



＊薬理活性
＊毒性関連データ
＊機器スペクトルデータ
＊バイオ関連データ
＊患者関連データ

口化合物構造式を用いたパラメーター発生



◇汎用的なデータ解析の流れ図：データ前処理関連
解析目的と無関係な情報を有するパラメーター除去（特徴抽出）

パラメーター選択（特徴抽出）の実施

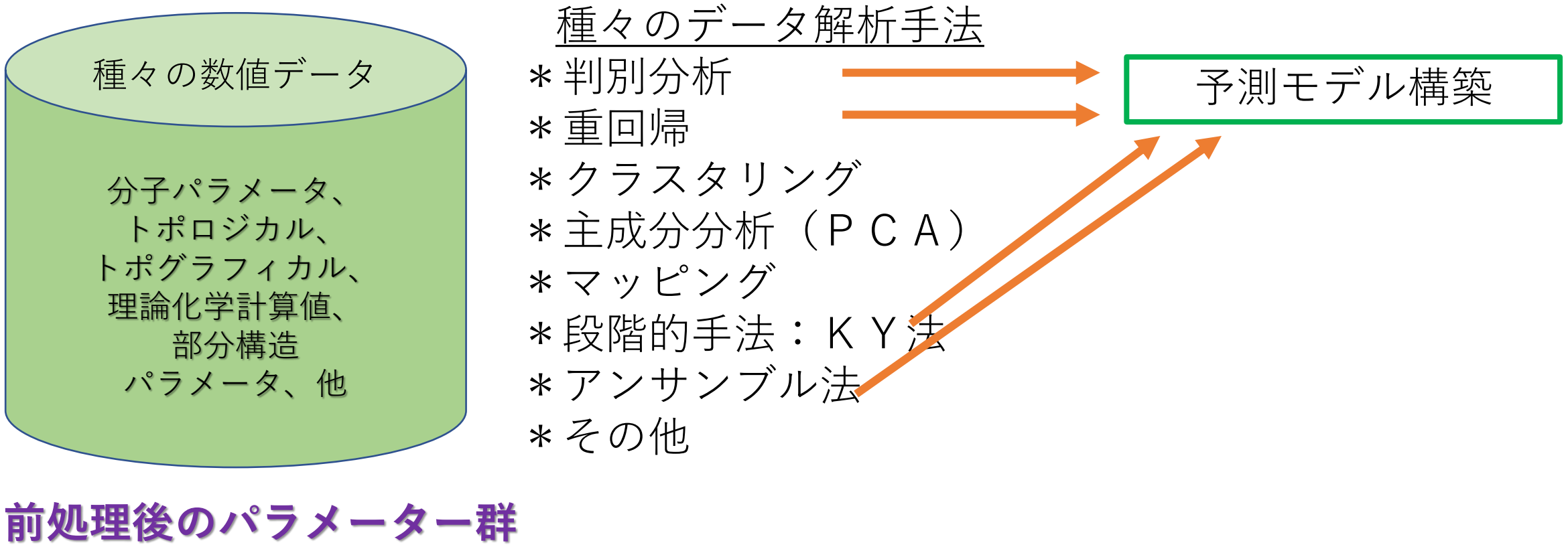
＊欠陥データ処理：
欠損データ、0値データ、同値データ、他

＊統計的処理：
単相関、重相関、Ｆｉｓｈｅｒ比、他

＊データ解析手法による個別特徴抽出
主成分解析、パーセプトロン、遺伝的アルゴリズム、他

＊パラメーター桁の調整
オートスケーリング

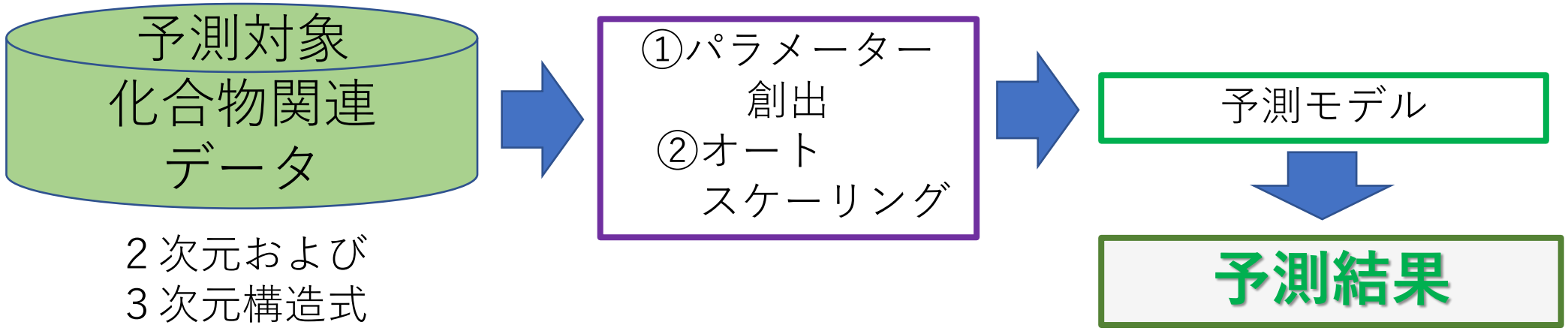
◇汎用的なデータ解析の流れ図：多変量解析／パターン認識関連



前処理後のパラメーター群

◇汎用的なデータ解析の流れ図：予測の実施

＊予測モデルを構成するパラメーターを化合物構造式から創出
＊構造式から発生できないパラメーターは外部パラメーターとして導入
＊オートスケーリングを実施



□データサイエンスに使う種々の化学パラメーター

◆化合物関連パラメーター

■ トポロジカルデータ
分子構造インデックス：原子数（原子種）、結合数（結合種）、リング数、その他
様々なインデックス値：ＨＯＳＯＹＡインデックス、分子結合インデックスＭＣ値
パス値インデックス、

■ トポグラフィカルデータ
化合物の３次元的形状に関するパラメータ
化合物全体構造　：ボックスパラメータ、対称パラメータ、
立体格子パラメータ、その他
化合物部分構造　：ステリモルパラメータ、

■ 物理化学データ
分子に関する様々な物性データ　：　分子屈折率、分子量、ＬＯＧＰ、融点、沸点
分子容積、分子表面積、その他
分子軌道法より得られる様々なパラメータ：電子密度、ＨＯＭＯ、ＬＵＭＯ、他
分子力学計算から得られるパラメータ：　種々歪みエネルギー
種々スペクトルより得られるデータ：　種々スペクトルデータ

■ その他のデータ
部分構造パラメータ：　部分構造の有無、部分構造数、
部分構造単位の様々なパラメータ値計算、
演算パラメータ１　：　記述子間の演算により得られるパラメータ（＋－ｘ÷ Log）
演算パラメータ２　：　他の解析手法より算出されたパラメータ
ダミーパラメータ　：　有るパターン存在の有無（１／０）に関するパラメータ

□AIに使う種々の化学パラメーター

深層学習による人工知能での化学パラメータ

1．フィンガープリント
2．フィーチュアベクトル
3．グラフコンボリューション

深層学習が発表され、化合物を扱う研究分野への適用初期は上記パラメータが利用されていた。

これらのパラメータ（特に２，３番）は、化合物をトポロジーと考える人工知能研究者が持ち出したパラメータであり、あまり化学的な意味はないといえる。

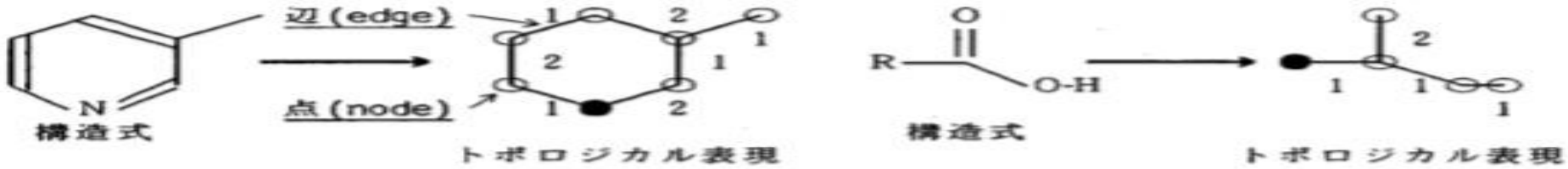
構造ー活性相関やケモメトリックスの分野では、古くから様々な化学パラメータが開発されているので、そちらを用いる方が化学的な解析には適している。

□解析に使うパラメーター

◆化合物関連パラメーター：トポロジカルパラメーター

□　トポロジカルデータの特徴
トポロジカルデータは化合物を構成する原子と結合とを、それぞれノードとエッジとに

定義する。　トポロジー的な問題として化合物構造式を捕らえ、化合物の原子（ノード）間の相互的な関係（つながり状態）を数値データに変換したものである。
このトポロジカルデータの特徴を簡単にまとめると以下のようになる。
長所：　化合物の複雑な結合情報を数値データに変換できるので、通常の数値データでは説明出来ないような複雑な情報を扱う事が可能となる。　この結果、分類能が飛躍的に向上することが期待される。
短所：　数値データの変換ルールと化合物構造式との関係が不明な時が多い。　アルゴリズムが数値データへの変換の為のルールとなっている事が多く、最終目的である目的変数に対する情報の説明や解釈が困難な事が多い。　即ち、分類の為のデータに陥り易く、分類だけが目的の時は強力なパラメータとなりうるが、そのパラメータの持つ意味（情報）を解釈する事が重要となる解析には不向きである。



このトポロジカルデータは現在様々なものが提唱されている。　特に有名なものとして化合物の物性予測に用いられる事の多いＨＯＳＯＹＡ　ＩＮＤＥＸ　と、構造活性相関分野で利用実績の多い分子結合インデックス（M. C.）（Molecular Connectivity Index）等が有名である。

□解析に使うパラメーター

◆化合物関連パラメーター：トポロジカルパラメーター

□　ＭＣＩ値の算出法

まず化合物を構成している個々の結合についてＣｋ値を求める。　続いて、このＣｋ値を化合物中の総ての結合について総和した値が分子に対するＭＣＩ値となる。

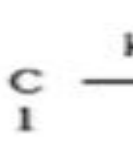
$$MCI = \sum_{k=1}^{m} C_k = \sum_{k=1}^{m} \frac{1}{[L_i \cdot L_j]^{1/2}}$$

k：　ある一つの結合のＩＤ番号
i：　結合ｋを形成する原子２個のうちの一つの原子に関するＩＤ
j：　結合ｋを形成する原子２個のうちｉ以外の原子に関するＩＤ

上式中、L_iは原子ｉの結合の多重度であり、L_jは原子ｊの多重度を示している。この多重度とは現在注目している原子から飛び出している結合の数を意味し、この時水素原子とつながっている結合の数は無視して計算する。

例）　Ｃｋ値の求め方

$$C_k = \frac{1}{[3 \cdot 3]^{1/2}} = \frac{1}{3}$$

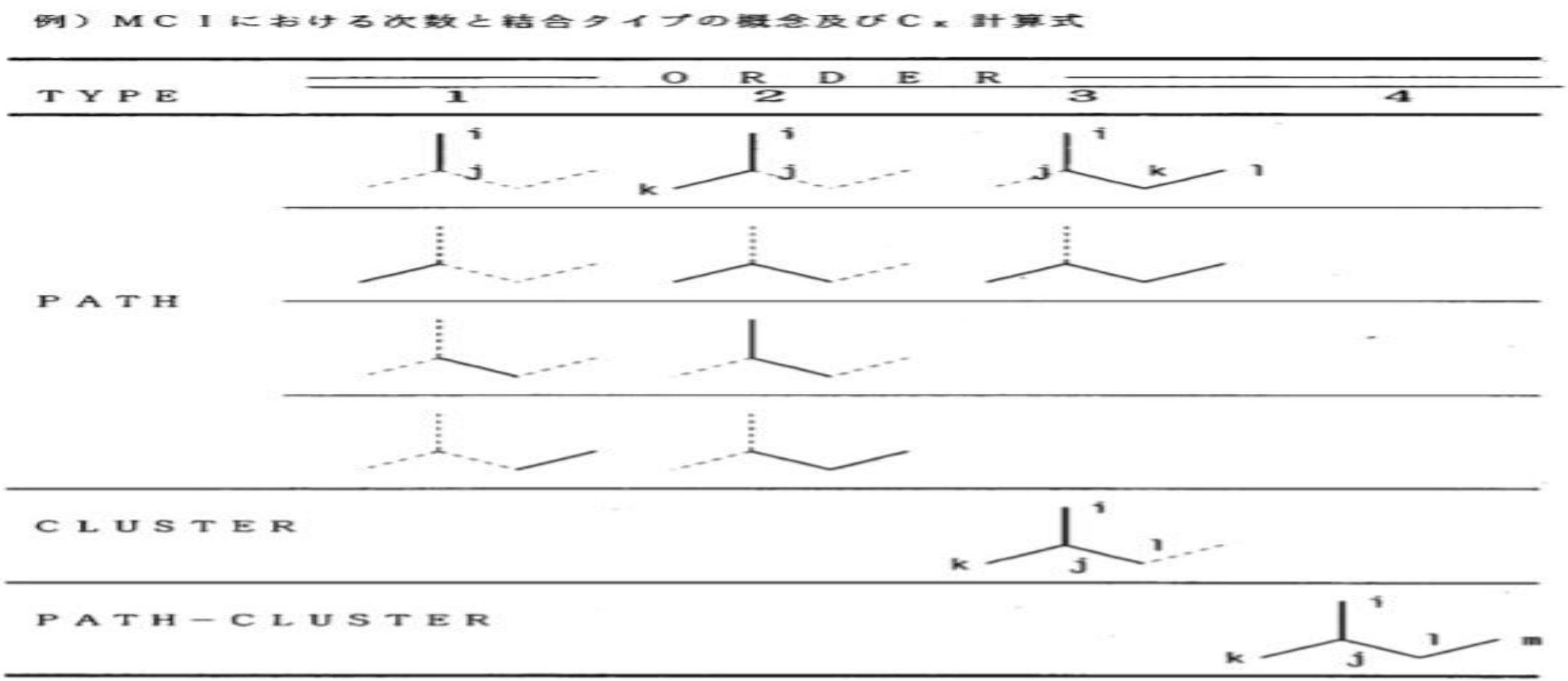


$$C_k = \frac{1}{[1 \cdot 4]^{1/2}} = \frac{1}{2}$$

□解析に使うパラメーター

◆化合物関連パラメーター：トポロジカルパラメーター

例）ＭＣＩにおける次数と結合タイプの概念及びＣｘ計算式



□解析に使うパラメーター

◆化合物関連パラメーター：トポロジカルパラメーター

□　ＭＣＩへの結合次数及び結合タイプの導入

Ｃｋ値を求め、この値を基準としてＭＣＩを求める時、化合物構造式の複雑さを情報として取り入れるべく結合次数（ＢＯＮＤ　ＯＲＤＥＲ）という概念と結合タイプ（ＢＯＮＤ　ＴＹＰＥ）という２つの概念を導入する。

・結合次数（ＢＯＮＤ　ＯＲＤＥＲ）はＣｋを求める時の対象となる結合と、その結合を形成する原子の数を拡大してゆくものである。
・結合タイプ（ＢＯＮＤ　ＴＹＰＥ）とは、結合が複数集まって一つのＣｋを形成する時の集合形態に関する情報である。

□　結合次数（ＢＯＮＤ　ＯＲＤＥＲ）について

結合次数は基本となるＣｋ値を求める時に対象とする結合や原子数を規定するものである。　次数が小さければＭＣＩの値は大きく、次数が増大するにつれてＭＣＩの値は小さくなる。

□　結合タイプ（ＢＯＮＤ　ＴＹＰＥ）について

ＴＹＰＥはＣｋとしてまとまった単位（特に次数が大きくなった時）の形を規制するものである。
・ＰＡＴＨは最も単純な形をしており、結合が直線上に繋がっているものを意味する。この時、次数が１のものは直線であり、ＰＡＴＨとみなす。
・ＣＬＵＳＴＥＲは分岐した形状を持つＣｋとなる。　従って、次数が３以上で現れる
・ＰＡＴＨ－ＣＬＵＳＴＥＲはＣｋ内部にＰＡＴＨ部分とＣＬＵＳＴＥＲ部分を持つ。

□解析に使うパラメーター

◆化合物関連パラメーター：トポロジカルパラメーター

例）

'χ p : 結合次数1、PATHタイプのCχ値を基本として求めたMCI値

'χ pc: 結合次数4、PATH−CLUSTERタイプのCχ値を基本として求めたMCI値

'χ p : 結合次数1、PATHタイプのCχ値を基本として求めたMCI値にリング補正を加えた値

'χ p v: 結合次数1、PATHタイプのCχ値の計算にヘテロ原子を考慮して求めたMCI値

□ 次数（ORDER）が異なる時のCχの計算式（次数1～4について）

$$\text{次数 } 1 = \sum_{s=1}^{N_a} (\delta_i \cdot \delta_j)_s^{1/2}$$

$$\text{次数 } 2 = \sum_{s=1}^{N_a} (\delta_i \cdot \delta_j \cdot \delta_k)_s^{1/2}$$

$$\text{次数 } 3 = \sum_{s=1}^{N_a} (\delta_i \cdot \delta_j \cdot \delta_k \cdot \delta_l)_s^{1/2}$$

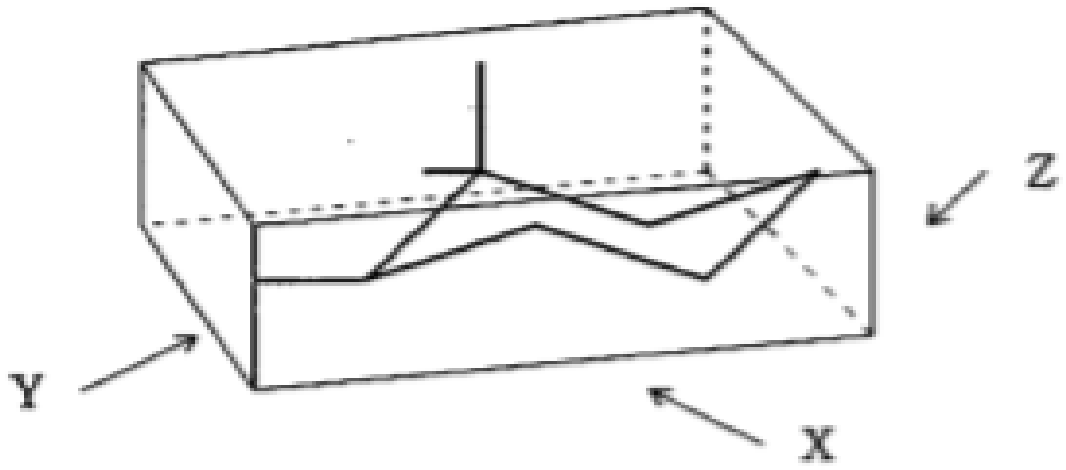
$$\text{次数 } 4 = \sum_{s=1}^{N_a} (\delta_i \cdot \delta_j \cdot \delta_k \cdot \delta_l \cdot \delta_m)_s^{1/2}$$

□解析に使うパラメーター

◆化合物関連パラメーター：**トポグラフィカル**パラメーター

② 分子全体の形状に関する幾何学的情報（ボックスパラメータ）

化合物の３次元構造式をそのまま長方形のボックスにいれる。 このボックスの各軸の長さとその比とをパラメータとする。



パラメータ１＝	X
パラメータ２＝	Y
パラメータ３＝	Z
パラメータ４＝	X／Y
パラメータ５＝	X／Z
パラメータ６＝	Y／Z

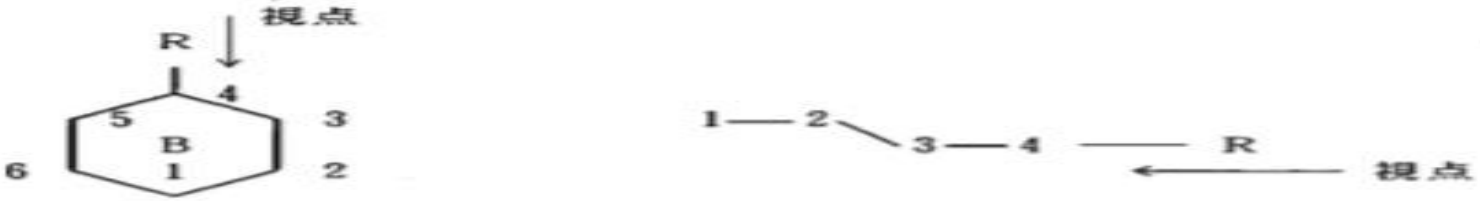
このパラメータにより、分子全体の立体的な形状についての情報がえられる。 例えば、分子が平面に近い、細長い、立方体に近い等の情報である。

□解析に使うパラメーター

◆化合物関連パラメーター：**トポグラフィカル**パラメーター

①ＳＴＥＲＩＭＯＬ　ＰＡＲＡＭＥＴＥＲ

このパラメータは化合物の置換基Ｒの３次元的立体情報を記述するのに用いられる。特に、重回帰手法によるＨＡＮＳＣＨ／ＦＵＪＩＴＡ法等に用いられて数多くの実績を有する構造活性相関には重要なパラメータである。



パラメータは化合物の基本構造部分（図中１～６で示されるＢ部分）と置換基Ｒ部分とに分けた時、基本構造部分と置換基Ｒ部分とが直結している結合をＲの方からＢに向かって見た時の置換基Ｒの占める空間上の領域をそれぞれの軸方向について分割した時の値を要素データとするものである。

ＳＴＥＲＩＭＯＬ＝（Ｗｌ．Ｗｒ．Ｗｓ．Ｗｎ．Ｌ）
＝（1.5．2.5．2.0．1.5，4.0）

□解析に使うパラメーター

◆化合物関連パラメーター：物理化学パラメーター

◇種々の物性：

分子量、　融点、　沸点、分子屈折率、　LogP、　Hammet σ、　その他

◇分子軌道法関連パラメーター：

電子密度、　HOMO、　LUMO、　分極率、　双極子モーメント、その他

◇分子力学関連パラメーター：

結合エネルギー、トーションエネルギー、　水素結合エネルギー、　その他

□解析に使うパラメーター

◆化合物関連パラメーター：物理化学パラメーター　（LogP）

□ ＬＯＧＰパラメータの定義式

・ＨＡＮＳＣＨ－ＲＥＯによるフラグメント付加方式によるＬＯＧＰ値推算。

$$LOGP = LOG \frac{[\ C\ ]\ lipid}{[\ C\ ]\ aqueous}$$

[C] lipid　：　平衡状態における油層中の濃度
[C] aqueous ：　平衡状態における水層中の濃度

□解析に使うパラメーター

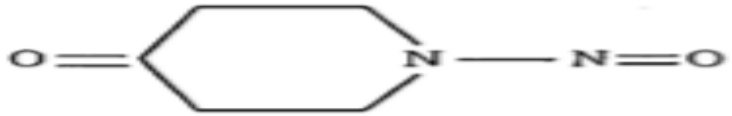
◆化合物関連パラメーター：物理化学パラメーター（LogP）

① フラグメント付加方式によるＬＯＧＰ推算式

$$LOGP = \sum_{i=1}^{n} a_i f_i + \sum_{j=1}^{m} b_j F_j$$

a_i ： i番目のフラグメントの出現回数
f_i ： i番目のフラグメントに対するフラグメント定数値
b_j ： j番目の修正因子の出現回数
F_j ： j番目の修正因子の修正定数値

ＬＯＧＰ値計算例）



4-KETO-N-NITROSO-PIPERIDINE

フラグメント定数

フラグメント	出現回数	フラグメント定数	総　和
― CH_2 ―	4	0．66	2．64
ケトン	1	－1．90	－1．90
N－N＝O	1	－2．45	－2．45

修　正　定　数

リングボンド	（n－1）（－0．09）		
	＝5（－0．09）		－0．45
極性基修正	2X［－（0．20）（$f_1 + f_2$）］		
	2X［－（0．20）（−2.45 − 1.90）］		
	2X［0．87］		1．74

$LOGP_{CALC}$ ＝ －0．42
$LOGP_{OBS}$ ＝ －0．47

□解析に使うパラメーター

◇　機器スペクトルパラメーター

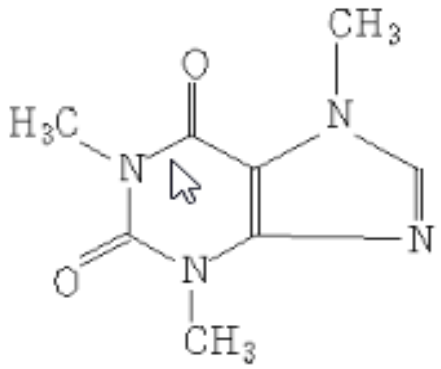
■有機化合物のスペクトルデータベース **SDBS**

SDBS No: 1898 **CAS Registry No.:** 58-08-2

DOI:

Molecular Formula: $C_8H_{10}N_4O_2$ **Molecular Weight:** 194.2

SDBS-NO= 1898
CAFFEINE



Compound Name:

caffeine
1,3,7-trimethyl-3,7-dihydro-1H-purine-2,6-dione
1,3,7-trimethyl-3,7-dihydro-purin-2,6-dion, kaffein
1,3,7-trimethyl-3,7-dihydro-purine-2,6-dione
1,3,7-trimethylxanthine
1H-purine-2,6-dione, 3,7-dihydro-1,3,7-trimethyl-
3,7-dihydro-1,3,7-trimethyl-1H-purine-2,6-dione
theine

InChI:

InChI=1S/C8H10N4O2/c1-10-4-9-6-5(10)7(13)12(3)8(14)11(6)2/h4H,1-3H3

InChIKey:

RYYVLZVUVIJVGH-UHFFFAOYSA-N

Publisher:

National Institute of Advanced Industrial Science and Technology (AIST)

https://sdbs.db.aist.go.jp/sdbs/cgi-bin/direct_frame_top.cgi

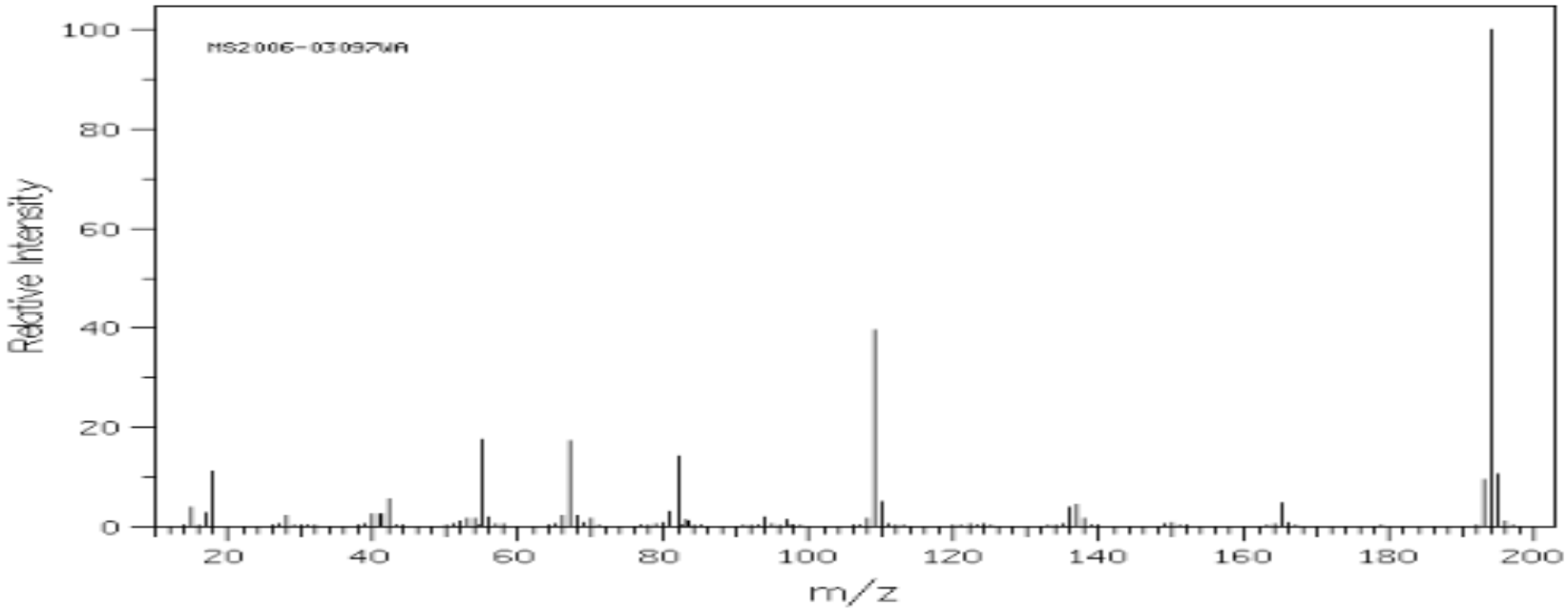
□解析に使うパラメーター

◇　機器スペクトルパラメーター

SDBS-Mass

MS2006-03097WA　　SDBS NO.　1898
caffeine
C8H10N4O2　　(Mass of molecular ion:　194)

Ｍａｓｓ



https://sdbs.db.aist.go.jp/sdbs/cgi-bin/direct_frame_top.cgi

□解析に使うパラメーター

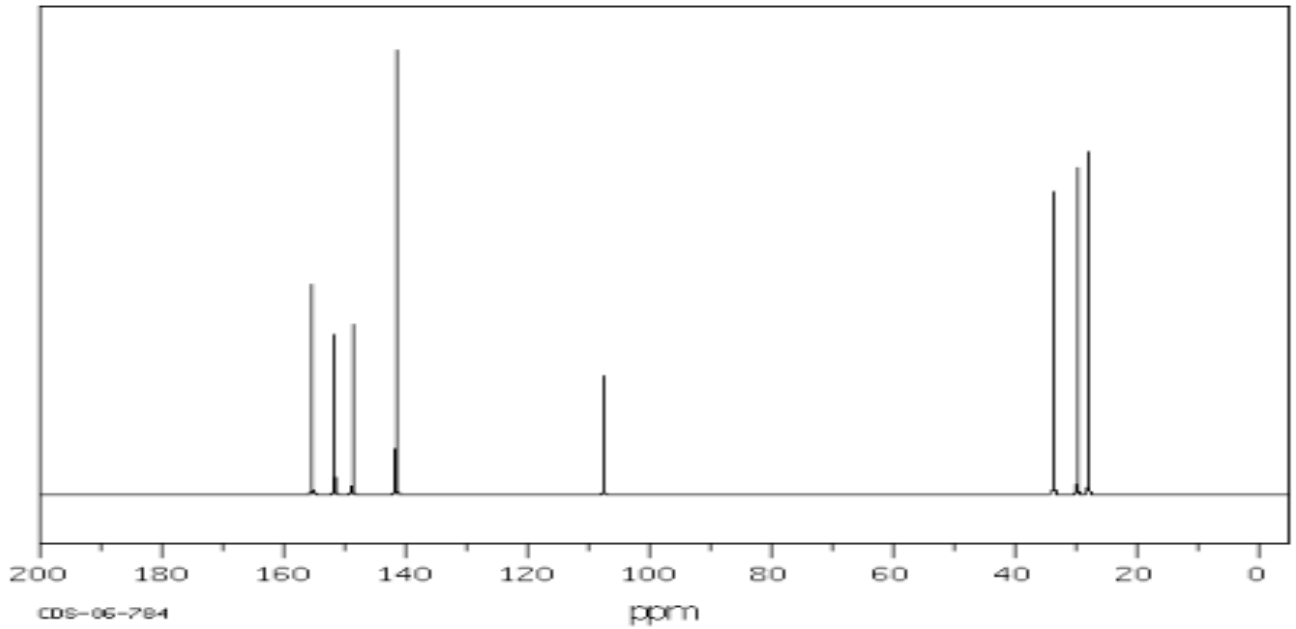
◇　機器スペクトルパラメーター：

SDBS-^{13}C NMR SDBS No. 1898CDS-06-784

$C_8H_{10}N_4O_2$

caffeine

^{13}C NMR : in $CDCl_3$

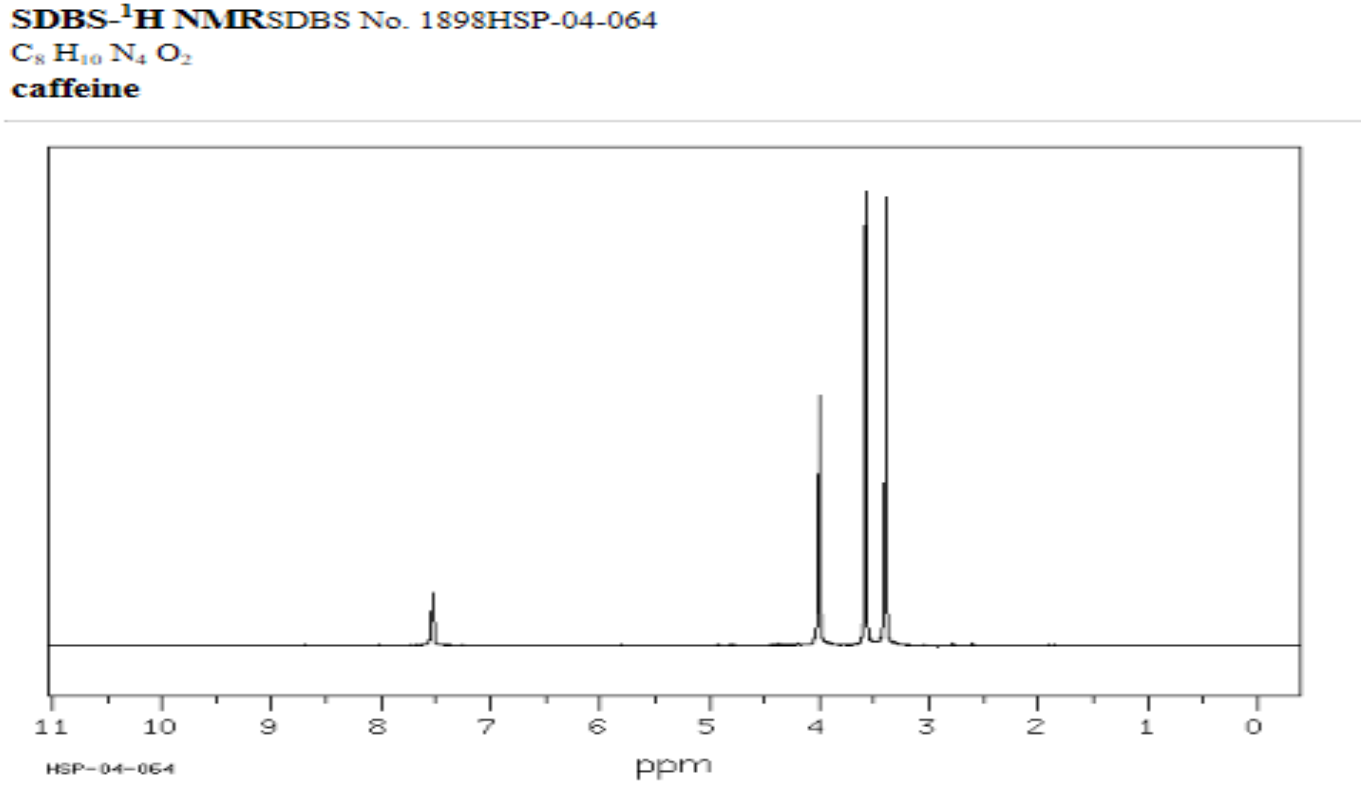


https://sdbs.db.aist.go.jp/sdbs/cgi-bin/direct_frame_top.cgi

□解析に使うパラメーター

◇　機器スペクトルパラメーター

^{1}H NMR : 90 MHz in $CDCl_3$

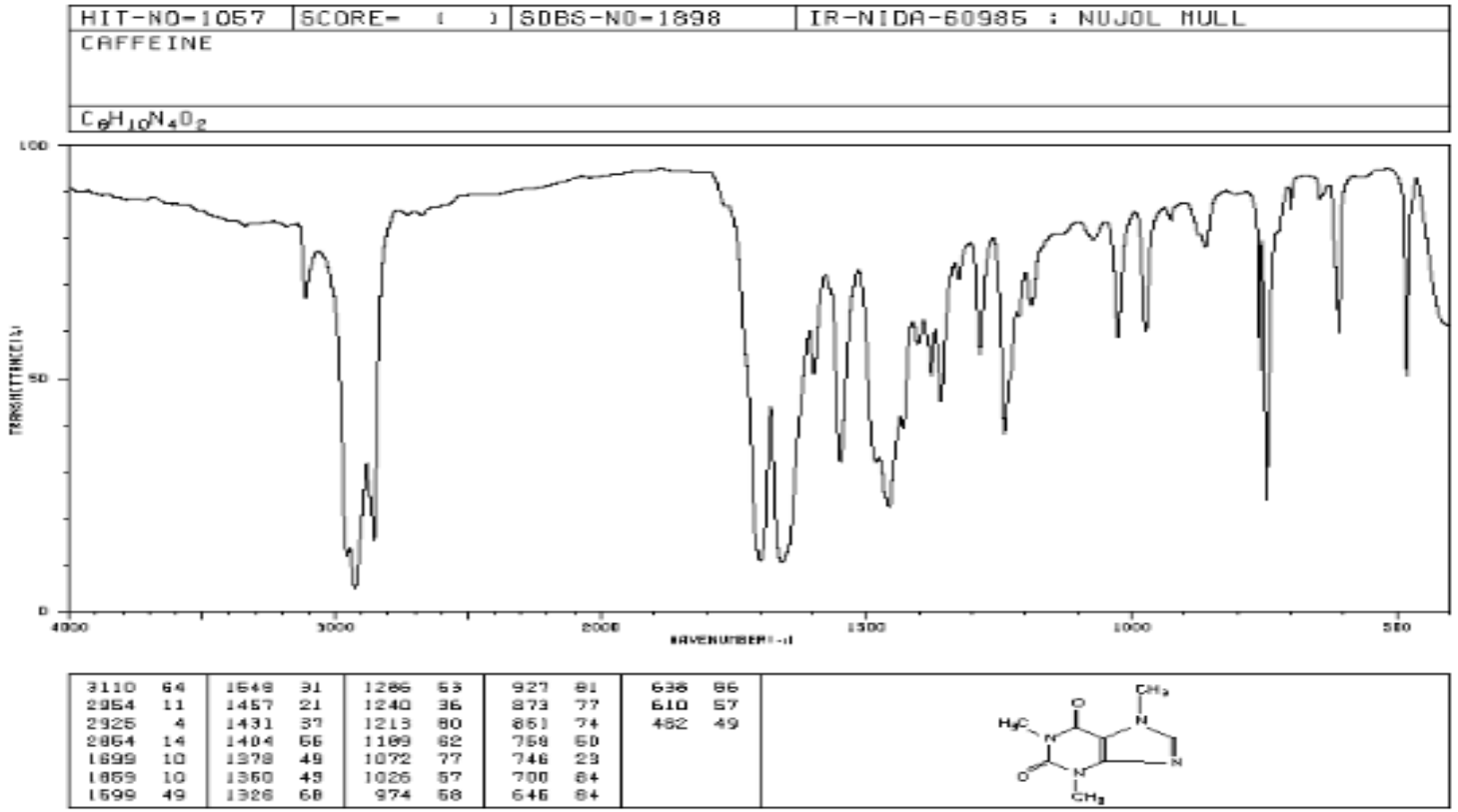


https://sdbs.db.aist.go.jp/sdbs/cgi-bin/direct_frame_top.cgi

□解析に使うパラメーター

◇　機器スペクトルパラメーター

IR：nujol null



https://sdbs.db.aist.go.jp/sdbs/cgi-bin/direct_frame_top.cgi

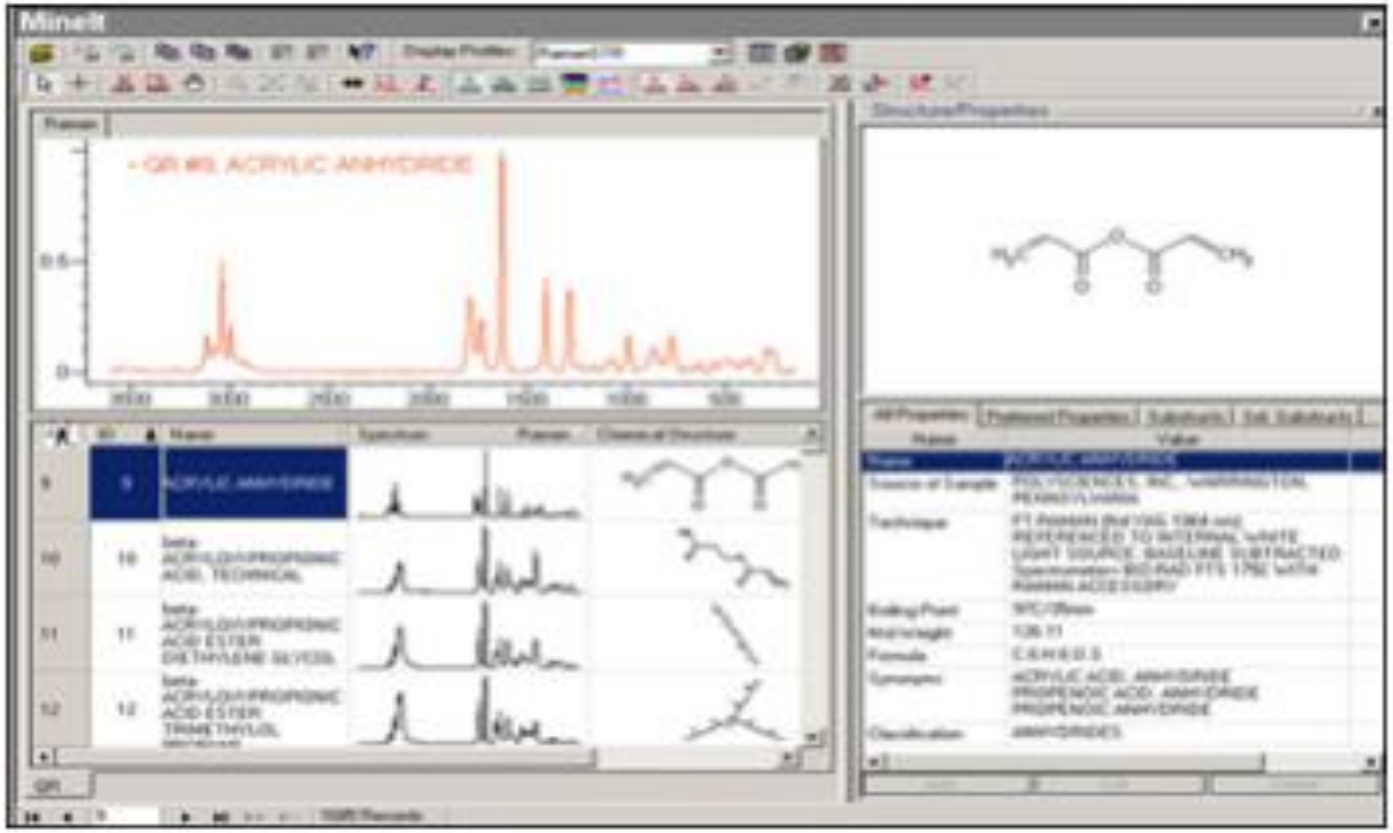
□解析に使うパラメーター

◇　機器スペクトルパラメーター

Ｒａｍａｎ
データ

BIO-RADの
スペクトルデータベースより

http://www.bio-rad.com/



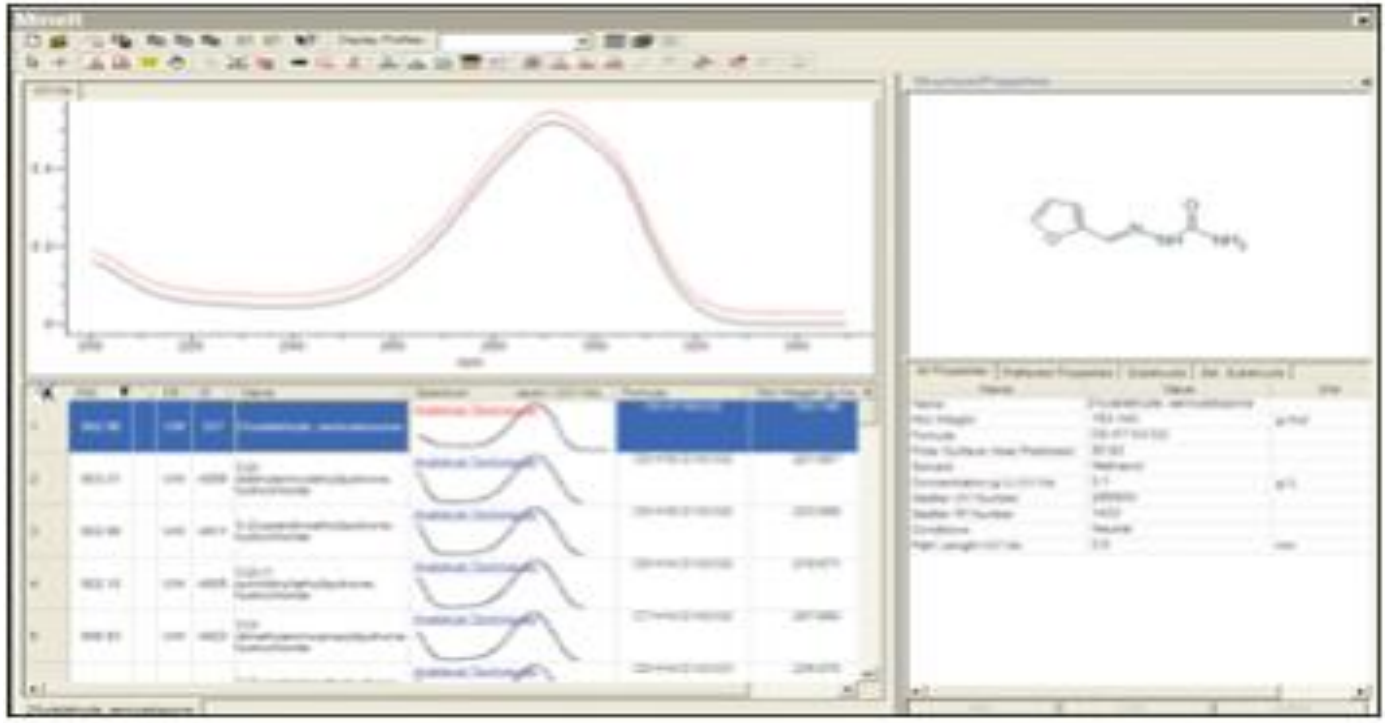
http://www.bio-rad.com/ja-jp/product/raman-spectral-databases?ID=N0ZXPS4VY

□解析に使うパラメーター

◇　機器スペクトルパラメーター

UV-Visスペクトル

紫外可視データベース



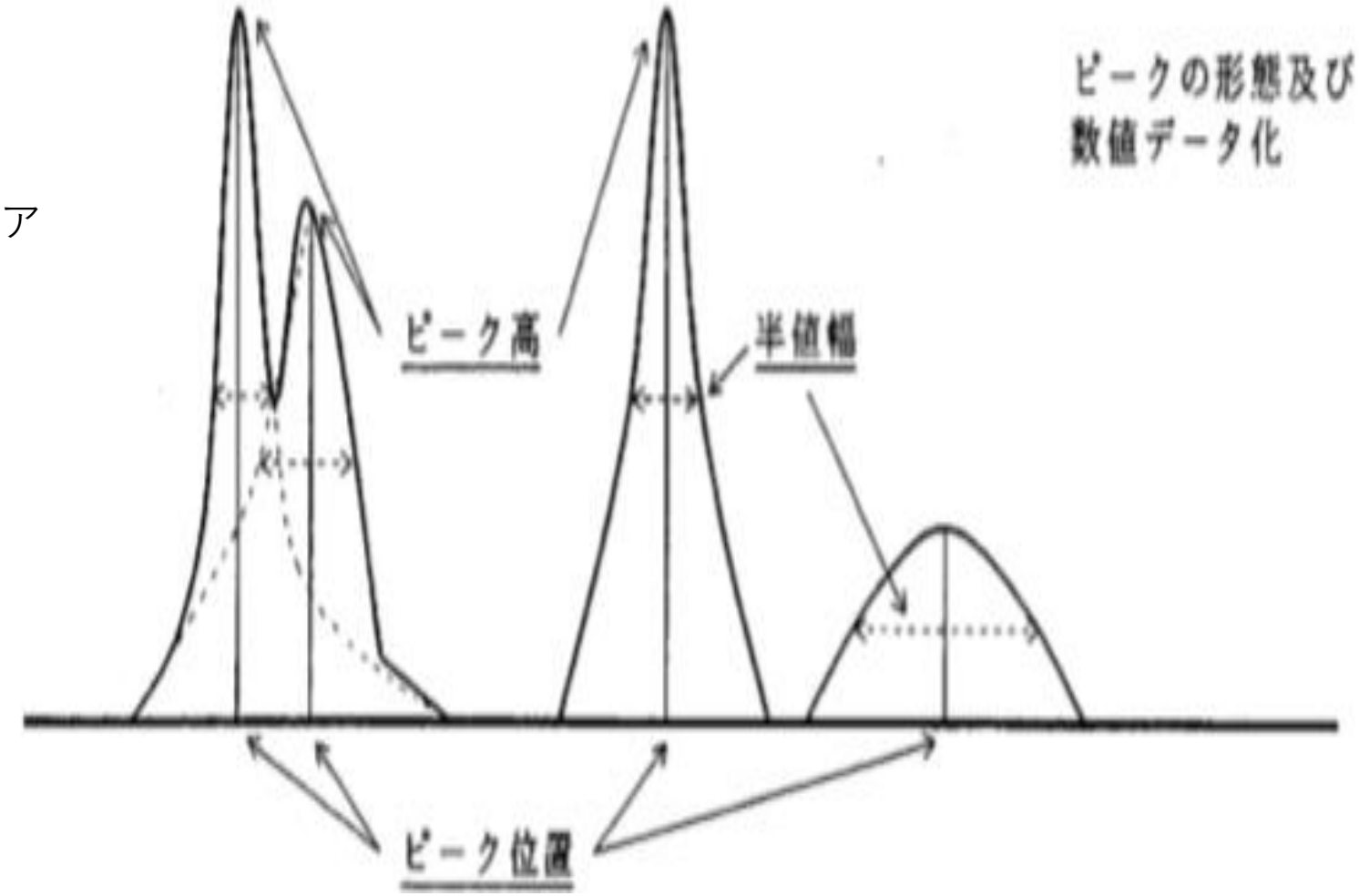
BIO-RADの
スペクトルデータベースより
http://www.bio-rad.com/

http://www.bio-rad.com/ja-jp/product/uv-vis-spectral-databases?ID=NH262L4VY

□解析に使うパラメーター

◆　機器スペクトルパラメーター

＊機器スペクトルデータは殆どの場合アナログデータなので、デジタルの数値データに変換することが必要である。



□解析に使うパラメーター

InSilico Data
Miracles by the KY-methods

◆　機器スペクトルパラメーター

特徴：

①機器があれば数値データとして簡単に蓄積できる
②スペクトルチャートは様々な実験過程で種々蓄積される

留意点：

①一般的にパラメーター数が極めて大きくなりやすい
　結果として、データ解析手法が限定、適用不可となる可能性が高くなる
②スペクトルデータは多重共線性が極めて高い
　①と②の特徴により、データ解析実施においては次元圧縮・統合等が必要で、
　解析手法もPLSやPCAと制限されることが多い
③スペクトルチャートの測定条件等統一が必要
　出来れば測定機器メーカーや機種も統一
例：６０MのH-NMRデータと９０MのH-NMRデータを混在してのデータ解析は無意味
・測定条件等が統一されないとデータ解析の精度が保証されにくくなる

In Silico Data
Miracles by the KY-methods

□本日の解説内容と順番

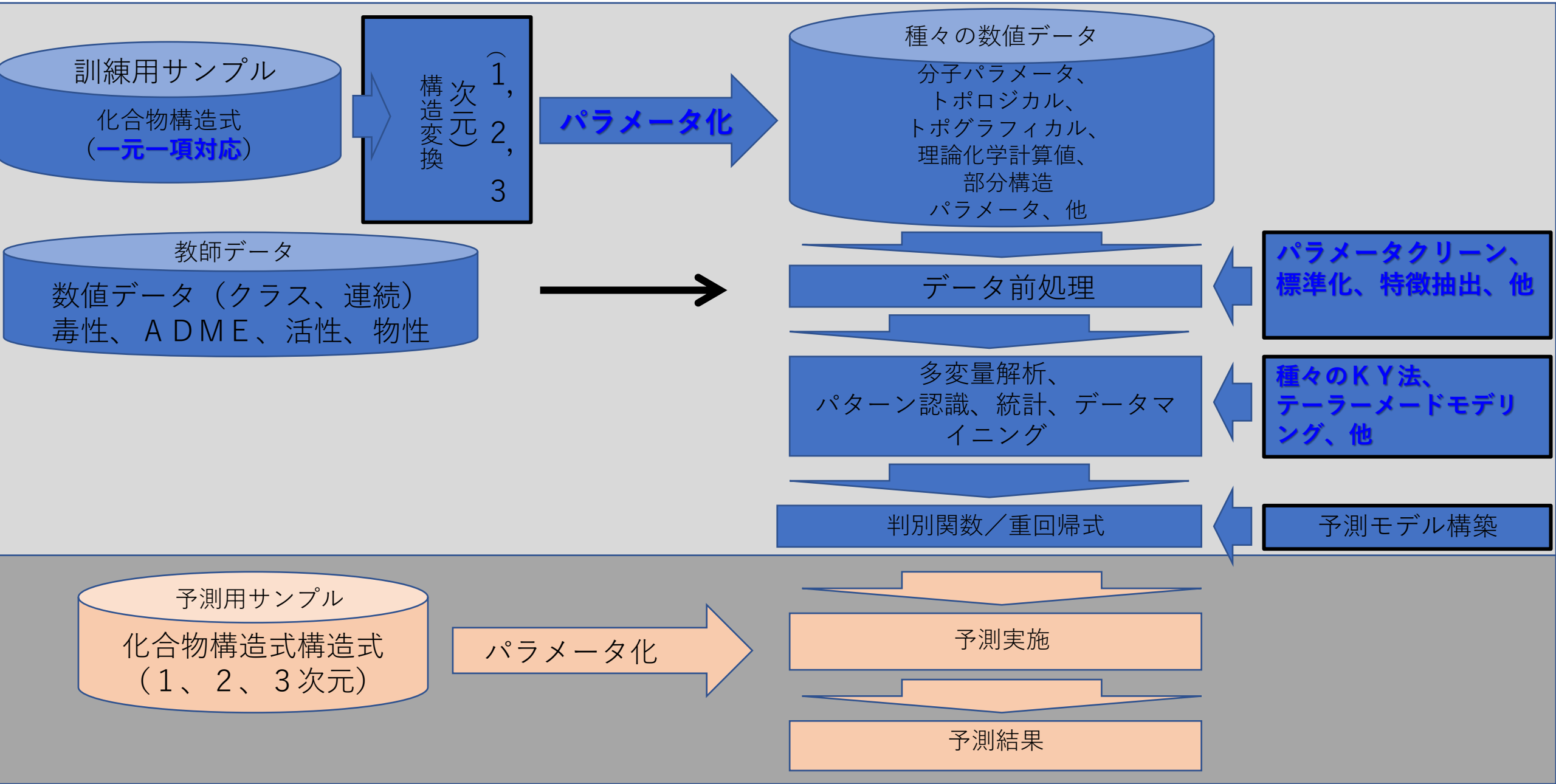
１．AI（人工知能）に関する概要

２．データサイエンスで利用される化学パラメータ

３．データサイエンス手法の概要

４．KY法の説明

◇汎用的なデータ解析の流れ図：解析全体の流れ



□ データ解析手法の機能的分類

統計関連情報の扱い方によりデータ解析手法は大きく二種類に分類される

1. パラメトリック手法

解析に用いるパラメーターの母集団分布情報を用いて解析する手法
分布（正規分布等）等の情報、平均値、分散、サンプル数は大きいものが良い
統計の検定等を実施するときに利用される

---

2. ノンパラメトリック手法

解析に用いるパラメーターの母集団分布情報が不明な環境で解析する手法
分布型は問わない、サンプル数も小さくても実施可能、他
制限事項が少ないので、広範囲にわたって適用できる
基本的に多変量解析／パターン認識として展開される手法はノンパラメトリック

◇ データ解析手法の機能的分類

解析アプローチの差異に伴い大きく以下の解析手法に分類される
個々の手法での代表的な手法をリストする

1. 判別分析：二クラス分類、多クラス分類
2. 重回帰（フィッテイング）：単回帰、重回帰
3. クラスター分析：階層型クラスタリング、非階層型クラスタリング
4. 種々マッピング：主成分分析（PCA）、非線形写像（NLM）
5. チャート表示：手法⇒レーダーチャート、顔チャート、ラインチャート、
6. 決定木：手法⇒C5.0、CART
7. アンサンブル学習法：手法⇒AdaBoost、ランダムフォレスト
8. 段階的手法：手法⇒KY法
9. 最適化手法：最小二乗法、シンプレクス法、遺伝的アルゴリズム

□ニクラス分類の基本概念

判別関数とドラグデザイン的情報解析

$$Y = a_1 x_1 \pm a_2 x_2 \pm \cdots \pm a_n x_n \pm const.$$

Y ： 目的薬理活性／毒性／他

Y ≧ 0	Y ＜ 0
・活性あり ・毒性あり	・活性無し ・毒性無し

□構造－活性／毒性／他との相関解析

係数 $a_i \geqq 0$ の時
パラメータ **X i** の持つ情報は
・活性向上、・毒性強化要因

係数 $a_i < 0$ の時
パラメータ **X i** の持つ情報は
・活性低下、毒性低下要因

構造－活性相関、構造－毒性相関

□ニクラス分類の基本概念

判別関数、重回帰式からの情報取り出し

□パラメータとウェイトベクトルの利用

1．要因の抽出および明確化：パラメータ利用

最終判別関数や重回帰式中で用いられているパラメータが持つ情報内容の評価

2．寄与の方向性：ウエイトベクトル利用

①係数　$a_i \geqq 0$　の時 ⇒　パラメータ X i の持つ情報は

・活性向上、あるいは毒性強化要因となる

②係数　$a_i < 0$　の時 ⇒　パラメータ X i の持つ情報は

・活性低下、毒性低下要因

3．寄与の相対的な貢献度：ウエイトベクトル利用

係数の絶対値比較による貢献度の評価

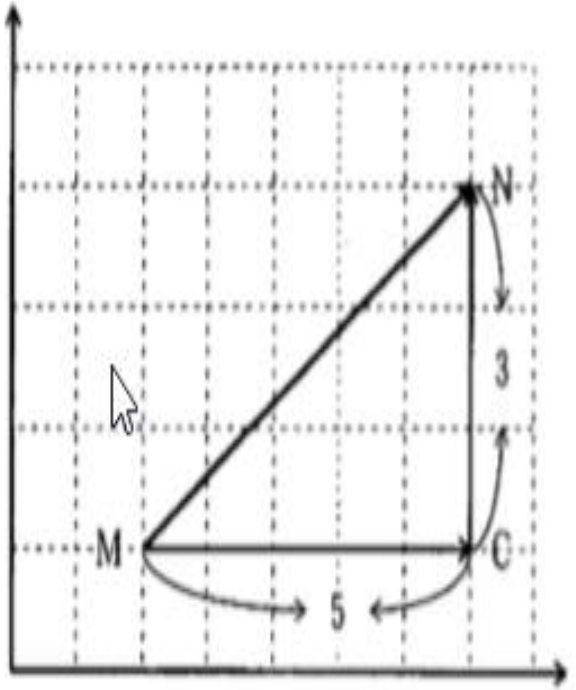
□データサイエンス手法の基本事項

◇サンプル間の**距離**

Ｎ次元空間上でのサンプル間の距離の算出は様々な多変量解析／パターン認識手法の基本となる。

サンプル間の距離算出に様々なメトリック手法が展開されている。

右は、ユークリッド距離とシテイブロック距離による距離計算事例である



２次元空間中でＡ点とＢ点との距離を計る時、ユークリッド距離では距離D_{MN}を算出し、シテイブロック距離では距離D_{MC}と距離D_{CN}との距離を合わせたものをＡ点とＢ点との距離とする。

M＝（2，1）
N＝（7，4）

ユークリッド距離 ＝ $D_{MN} = (5^2 + 3^2)^{1/2} = 5.83$

シテイブロック距離 ＝ $D_{MC} + D_{CN} = 5 + 3 = 8$

図１．　ユークリッド距離とシテイブロック距離

◇サンプル間の距離：**距離基準**

(1) ミンコフスキー (MINKOWSKI)距離

$$D = \left[\sum_{i=1}^{d} (X_{Mi} - X_{Ni})^{k} \right]^{1/k}$$

(2) ユークリッド(EUCLIDEAN) 距離

ユークリッド距離はミンコフスキー距離の式において、kが2の時にあたる。

$$D = \left[\sum_{i=1}^{d} (X_{Mi} - X_{Ni})^{2} \right]^{1/2}$$

(3) シテイブロック(CITY BLOCK)距離

シテイブロック距離は2パターン間の最短距離をとるのではなく、直交する2線の距離の総和をとるものである。

$$D = \sum_{i=1}^{d} \left[X_{Mi} - X_{Ni} \right]$$

(4) キャンベラ (CANBERA)距離

$$D = \frac{\sum_{i=1}^{d} \left[X_{Mi} - X_{Ni} \right]}{\sum_{i=1}^{d} \left[X_{Mi} + X_{Ni} \right]}$$

(5) ハミング(HAMMING) 距離

ハミング距離は1／0のバイナリデータで利用される事が多いが、ORとANDの識別を効率良く行う事が出来る。

$$D = \sum_{i=1}^{d} \left[X_{Mi} + X_{Ni} - 2X_{Mi}X_{Ni} \right]$$

尚、1／0のバイナリーデータを用いた時、ハミング距離とシテイブロック距離は同じものとなる。

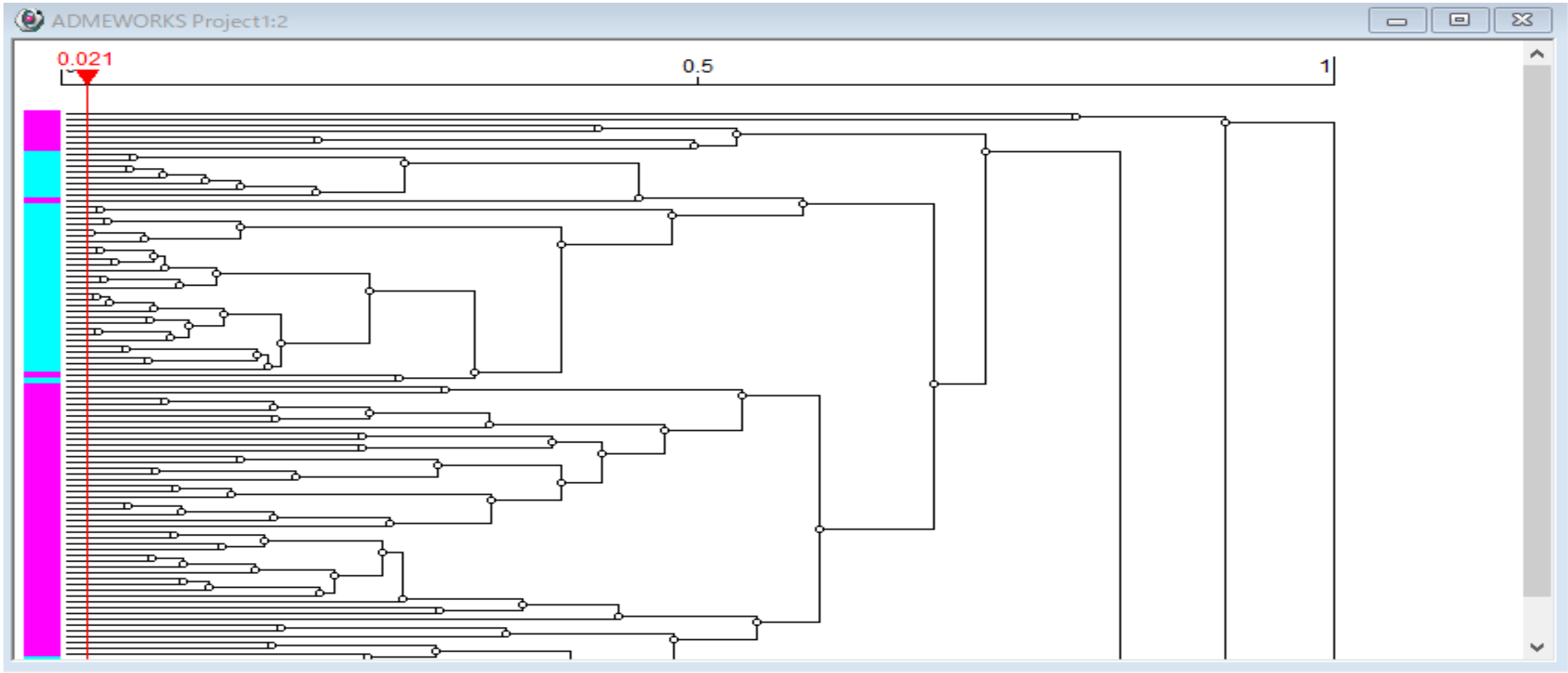
(6) 谷本距離

谷本距離はハミング距離がその性格上、1が少ないデータは不利に評価されるという欠点を改良したものである。

$$D = \frac{\sum_{i=1}^{d} \left[X_{Mi} + X_{Ni} - 2X_{Mi}X_{Ni} \right]}{\sum_{i=1}^{d} \left[X_{Mi} + X_{Ni} - X_{Mi}X_{Ni} \right]}$$

◇個別手法の特徴と適用内容：クラスター分析

ModelBuilderの画面より



◇個別手法の特徴と適用内容：クラスター分析
　クロスター分析は以下の基準にて様々な手法に分類される

クラスター化
アプローチ

■階層的クラスタリング
　解析結果はデンドログラム
　として出力される

■非階層的クラスタリング
　解析結果は単に
　クラスターの数とメンバー

クラスター化
アルゴリズム

①Division Method
　（分割法）
② Aggregative Method
　（凝集法）

融合法の種類

① 最近隣法 (NEAREST NEIGHBOR METHOD)
② 最遠隣法 (FURTHEST NEIGHBOR METHOD)
③ 群平均法 (GROUP-AVERAGE METHOD)
④ 重心法 (CENTROID METHOD)
⑤ メジアン法 (MEDIAN METHOD)
⑥ ワード法 (WARD METHOD)
⑦ 可変法 (FLEXIBLE METHOD)

◇個別手法の特徴と適用内容：クラスター分析
クラスタリングの融合手法

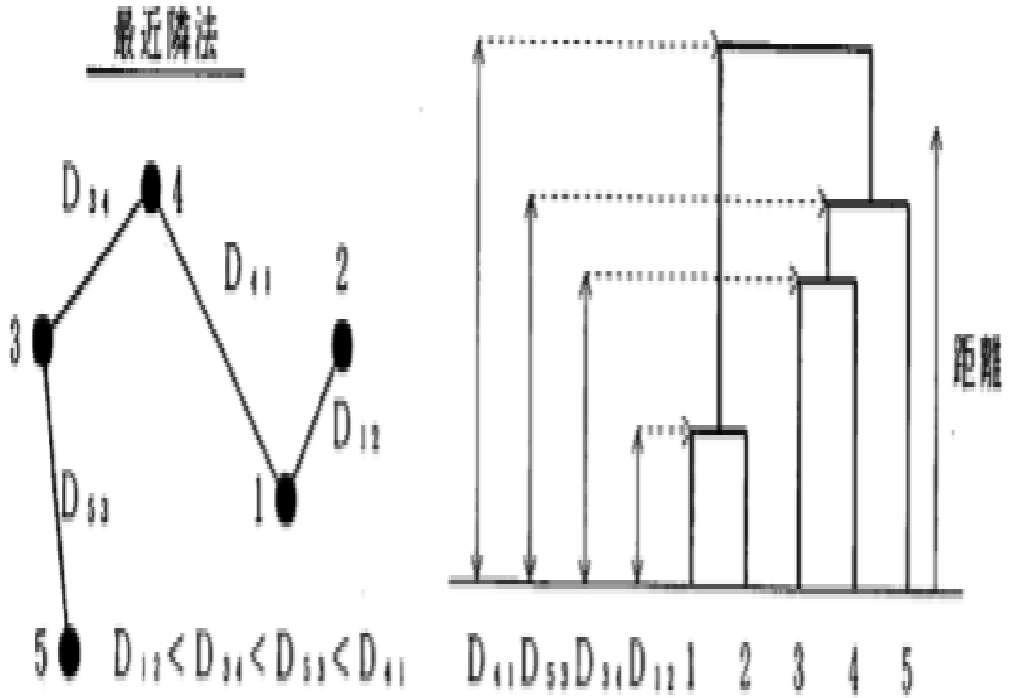


図2．　最近隣法によるクラスタリング手続きとデンドログラム

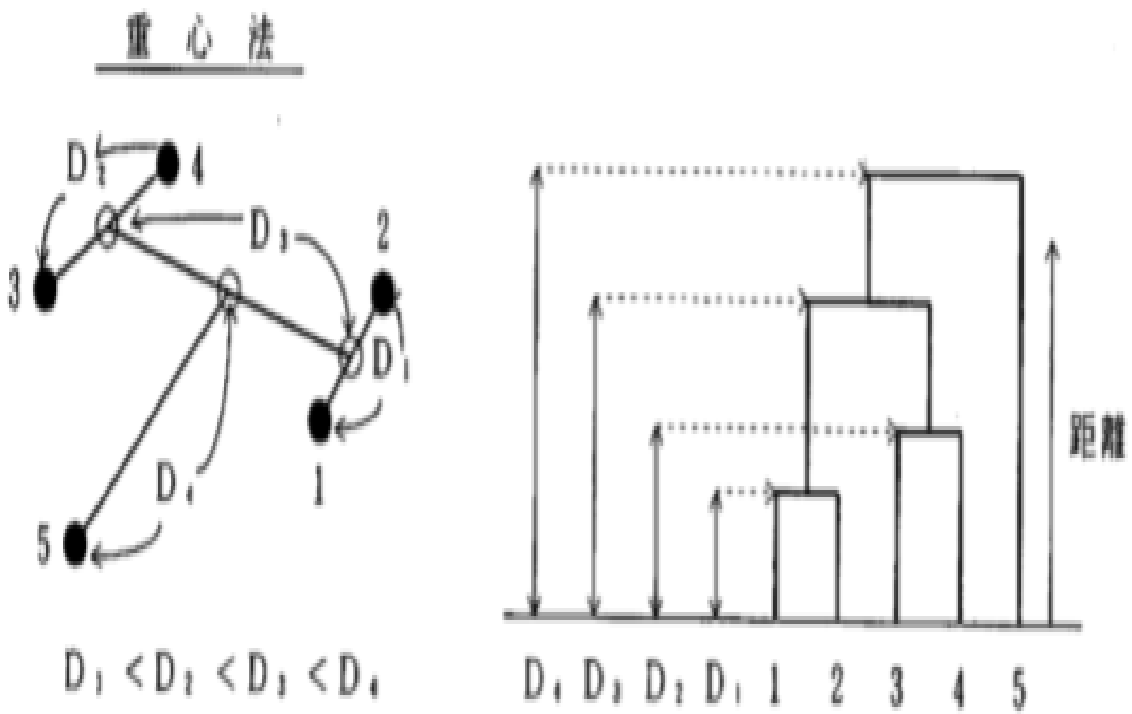
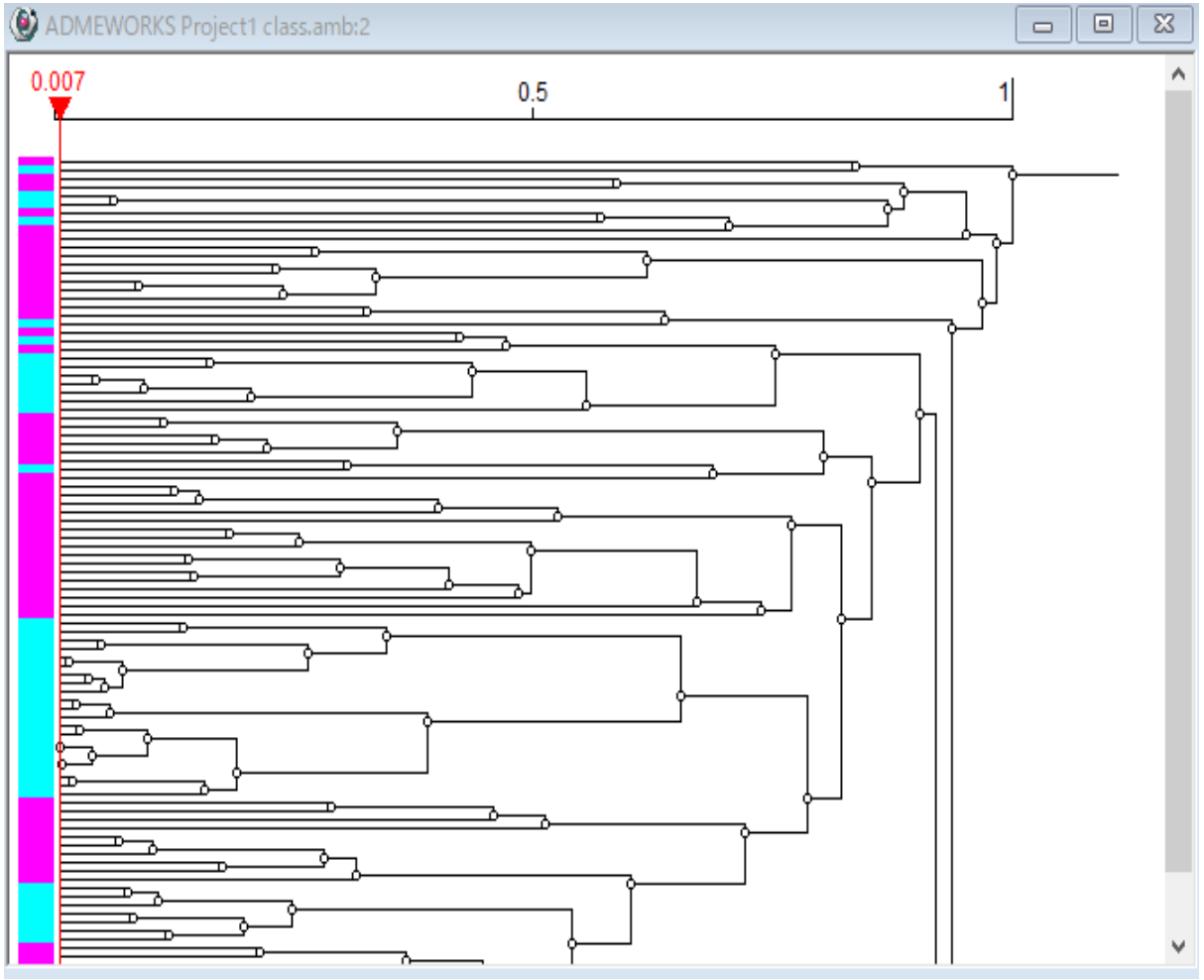


図3．重心法によるクラスタリング手続きとデンドログラム

◇個別手法の特徴と適用内容：クラスター分析

＊クラスタリングはサンプル同士の相対的な位置関係や近隣関係を俯瞰して見ることが出来る
＊メトリックや融合手法の差異によりクラスタリング結果が大きく変化するので要因解析は注意

ModelBuilderの画面より



◇個別手法の特徴と適用内容：マッピング

■主成分分析（PCA）
・データ解析的にはリニアプロジェクション手法
・サンプル空間中のサンプルの位置関係を変えない
・オリジナルのサンプル空間を俯瞰する方向性を変えて空間を見直す手法
・PCA適用前と後とで次元（パラメーター）数の変化はない

■非線形写像（NLM）
・データ解析的にはマッピング（写像）手法
・サンプルを人間が可視できる二次元/三次元上に分散
・サンプル空間を強制的に2/3次元に圧縮する
・非線形写像実施後は、サンプル空間の次元は2/3次元となる

◇個別手法の特徴と適用内容：Non　Linear　Mapping（NLM）

＊NLMにより、最初のN次元空間が
可視できる2次元空間に変換された。
＊NLMでは、元のN次元空間における
サンプル間の位置関係を保ちつつ
新しい2次元空間上に再配置される。
＊2次元空間上の第6サンプルは、
その他のサンプル（1～5、7～9）との相
互位置関係（空間上の距離）をN次元空間
上における関係と略同じとなる

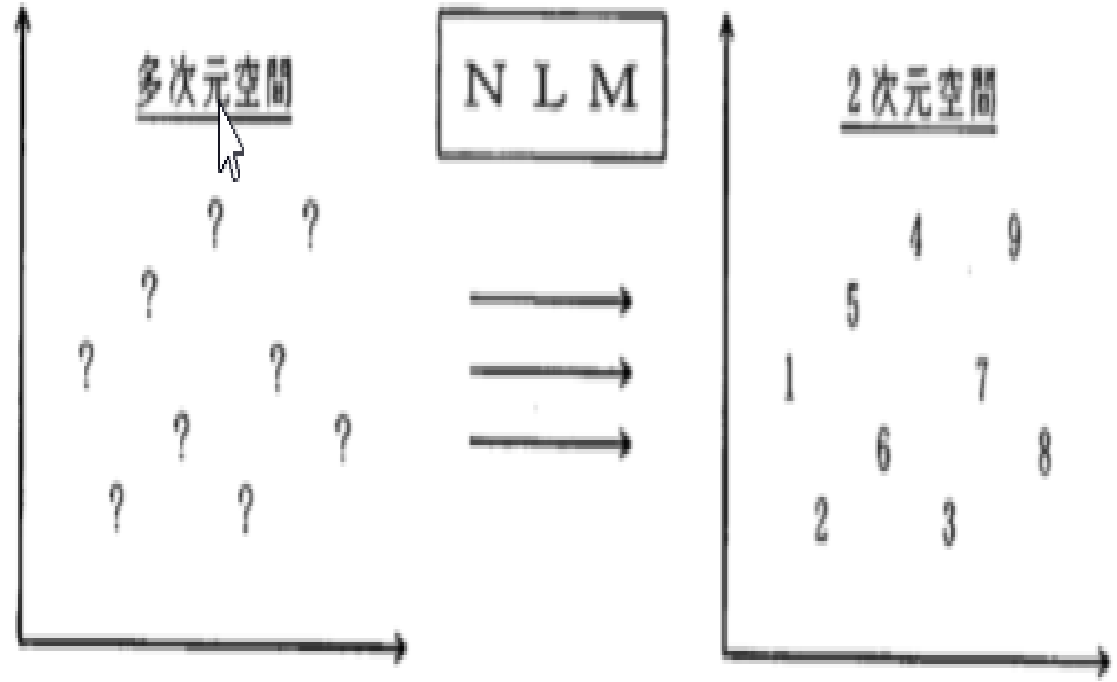


図5－6　ノンリニアマッピングによる多次元空間の2次元空間への写像

◇個別手法の特徴と適用内容：Non　Linear　Mapping（NLM）

多次元（d次元）空間上のパターンA，B

$$PA = (X_{A1},\ X_{A2},\ X_{A3},\ \cdots\cdots,\ X_{A,d-1},\ X_{A,d})$$

$$PB = (X_{B1},\ X_{B2},\ X_{B3},\ \cdots\cdots,\ X_{B,d-1},\ X_{B,d})$$

2次元空間上のパターンM，N

$$P2A = (Y_{A1},\ Y_{A2})$$

$$P2B = (Y_{B1},\ Y_{B2})$$

$$DX_{AB} = \sum_{i=1}^{d} \left\{ (X_{Ai} - X_{Bi})^2 \right\}^{1/2} \quad \text{(　)}$$

$$D2_{AB} = \sum_{i=1}^{2} \left\{ (Y_{Ai} - Y_{Bi})^2 \right\}^{1/2} \quad \text{(　)}$$

$$Ei = \frac{1}{\sum_{A\neq B}^{n} DX_{AB}} \sum_{A\neq B}^{n} \frac{(DX_{AB} - D2_{AB})^2}{DX_{AB}} \qquad (i = 1,\ 2,\ \sim,\ n)$$

ここで、nはパターンの数を示す。

このエラー関数の極小化は（Y_{i1}，Y_{i2}）（i＝1～n）の初期チャートを適当に与え

$$\frac{\partial E(Y_{ij})}{\partial Y_{ij}} = 0 \quad (i = 1 \sim n;\ \ j = 1,\ 2) \quad \text{(　)}$$

$$Y_{ij}(m+1) = Y_{ij}(m) - k\Delta_{ij}(m) \quad \text{(　)}$$

ここで $\Delta_{ij}(m) = \dfrac{\partial E(m)}{\partial Y_{ij}(m)} \Big/ \left| \dfrac{\partial^2 E(m)}{\partial Y_{ij}(m)} \right|$，で、$k \fallingdotseq 0.3 \sim 0.4$

◇個別手法の特徴と適用内容：

ニューラルネットワークによる次元圧縮

＊入力層のN個のパラメーターから中間層のユニット数（図は3）に次元を変換する手法

＊NLMと異なり、サンプル間の相互位置関係はキープされない

＊新たな次元は、ニューラルネットワークが解析目標とするクラスデータ等を説明する情報を含んでいる

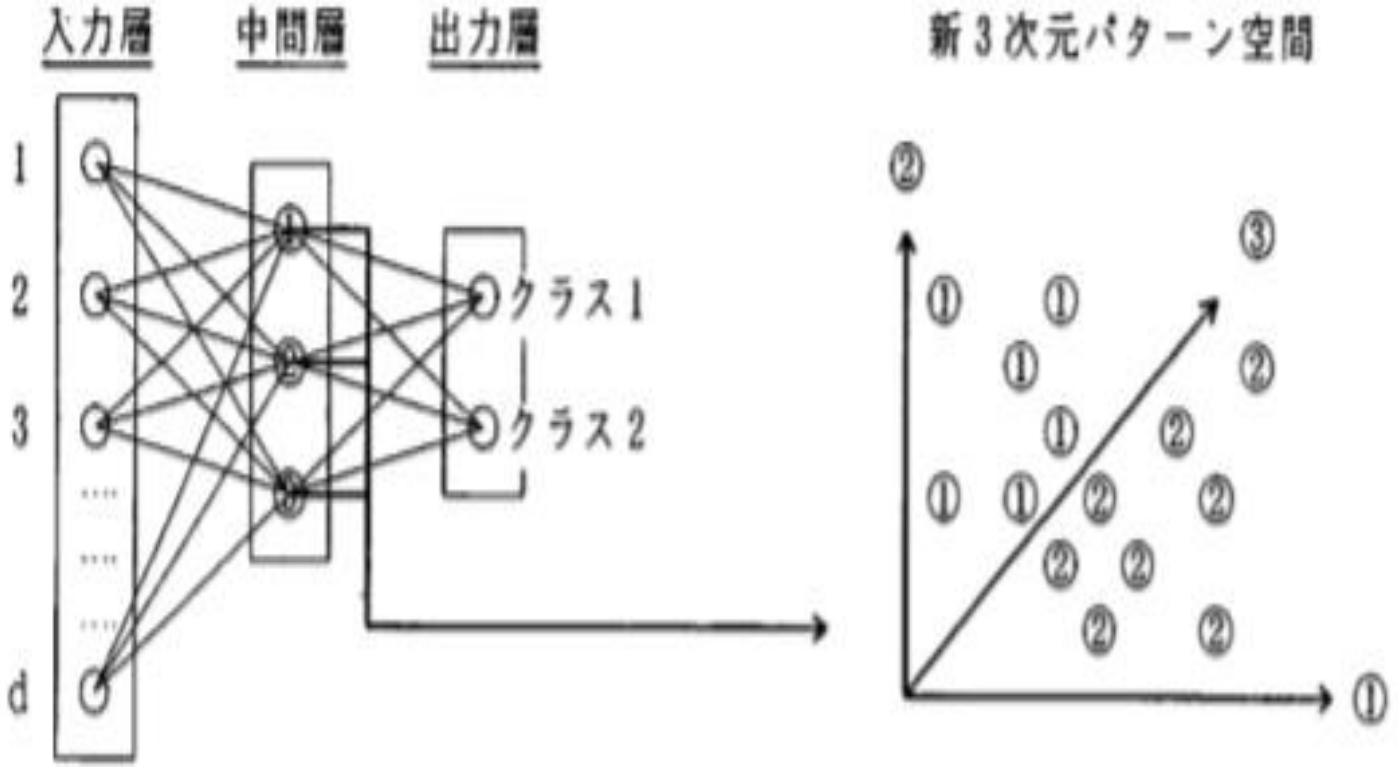


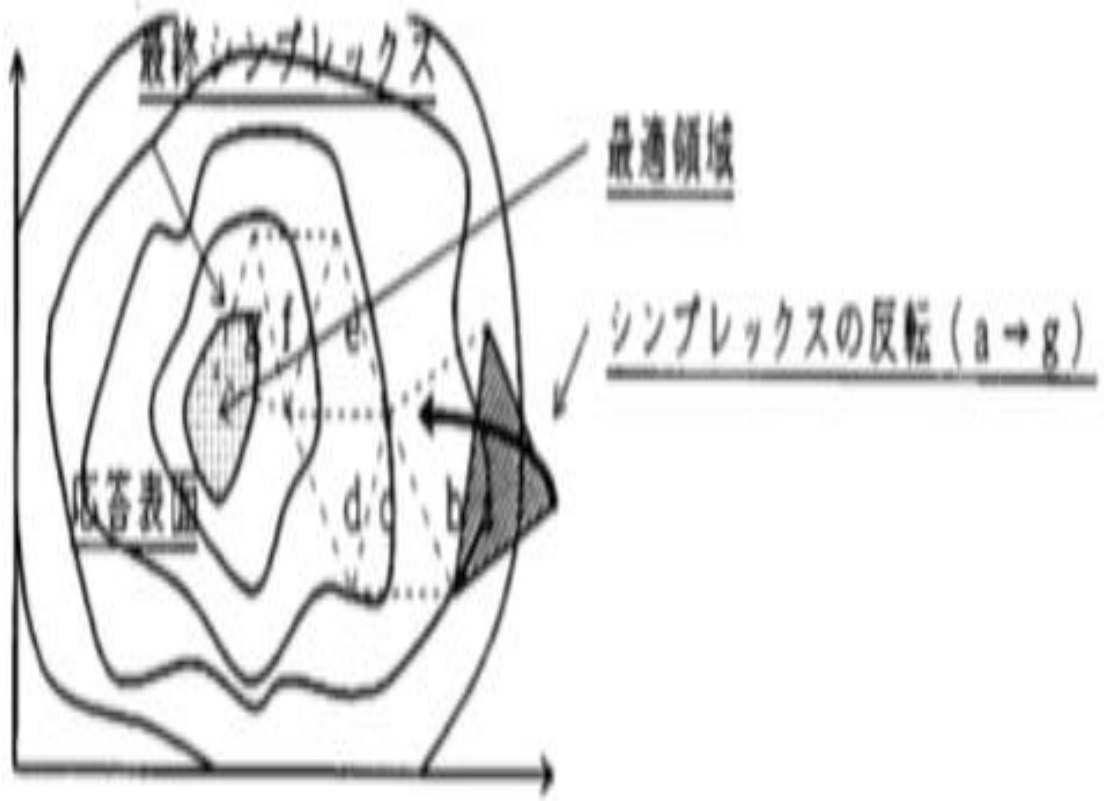
図５－７．バックプロパゲーションによる写像。 d次元⇨３次元
（新たな次元は各パターンがクラス１とクラス２とに分類されやすいように３次元空間上に分布していることに注意）

□アンサンブル学習法

AdaBoostやランダムフォレスト等で採用されている解析手法
特徴：
　弱分類機を条件を変えつつ多数用い、最終的な分類結果は
　個々の分類結果データを統合して最終結果とする。

　従来の分類手法を組み合わせて使う「メタ解析手法」である。
＊「**KY法**」もメタ解析手法となる

◇最適化手法：シンプレックス法



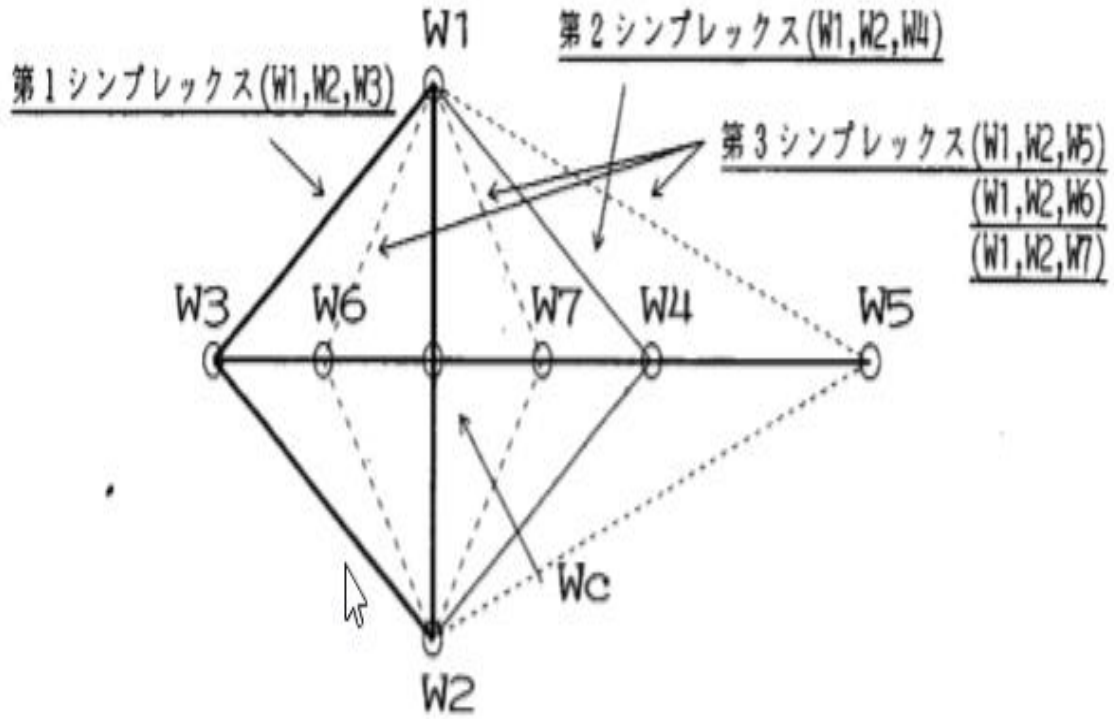


図6－2　シンプレックス反転に関するルール

◇最適化手法：遺伝的アルゴリズム

遺伝子の分裂／増殖／突然変異等の動きをシミュレーションするアプローチ

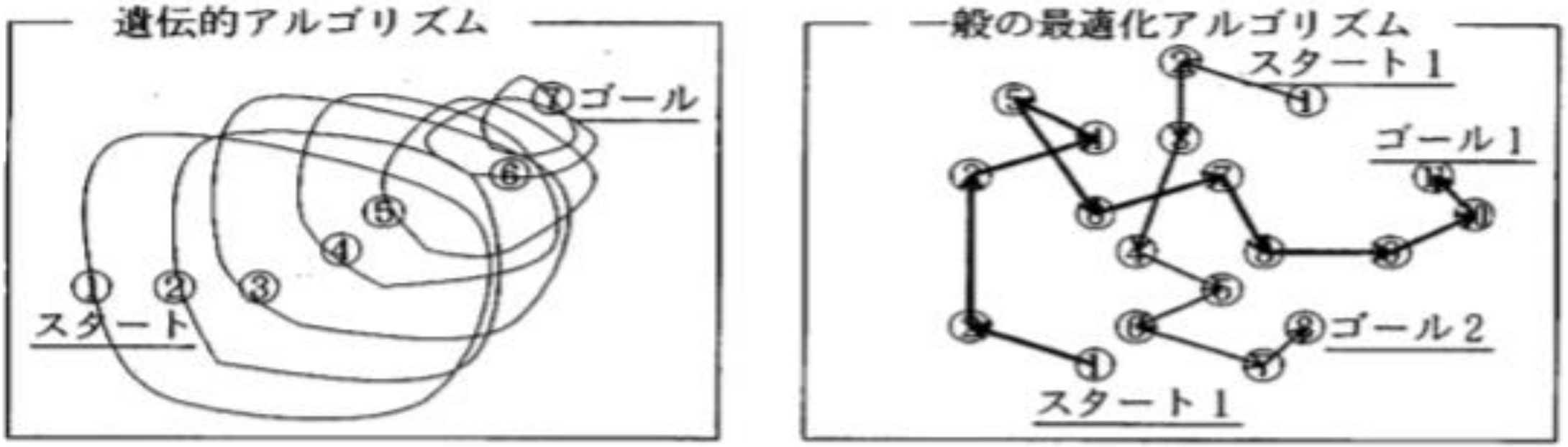


図３．遺伝的アルゴリズムと一般の最適化アルゴリズムによる最適領域の探索過程

1．増殖（MULTIPLICATION）
2．交叉（CROSSING-OVER）
3．突然変異（MUTATION）
4．淘汰／選択（SELECTION）

◇最適化手法：遺伝的アルゴリズム

1. 増殖（MULTIPLICATION）

2. 交叉（CROSSING-OVER）
情報の取替により、情報の変換を果たすものである。

```
101101001011101100100101111   遺伝子A

100111001100001100100101111   新遺伝子BA

101101001011101011100100011   新遺伝子AB

100111001100001011101000011   遺伝子B
```

異なる遺伝子パターンをもつ2本の遺伝子がある時、これら2本の遺伝子を交叉させる。この結果、新たに2本の遺伝子が誕生するがこれらの遺伝子は親となるAおよびBの遺伝子の形質を交叉させた点を中心としてそれぞれあわせ持つことがわかる。

3. 突然変異（MUTATION）
この突然変異では遺伝子の持つ情報がある部位で変化して対立遺伝子となる。

```
101101001011011001001011111   遺伝子A
                ↓ 突然変異
101101001010011001001011111   遺伝子A'
```

4. 淘汰／選択（SELECTION）
様々な変換パターンにより構築された遺伝子が淘汰により悪い形質（問題解決に取って望ましくない）を持つものが取り除かれてゆくことである。生物学的にはまわりの環境に適合した形質を持つ固体だけが生き残り（選択され）、次世代へと情報（形質）を伝えてゆくことを意味する。

```
                    淘汰
101 ……… 111  遺伝子A  ──→ ×
010 ……… 110  遺伝子B  ──→ 010 ……… 110  遺伝子B
001 ……… 001  遺伝子C  ──→ ×
……                    ──→ ……
100 ……… 011  遺伝子Z  ──→ ×
```

◇最適化手法：ニューラルネットワーク

閉じたネットワーク構造を有するニューラルネットワーク
ボルツマンマシンおよびホプフィールドネット

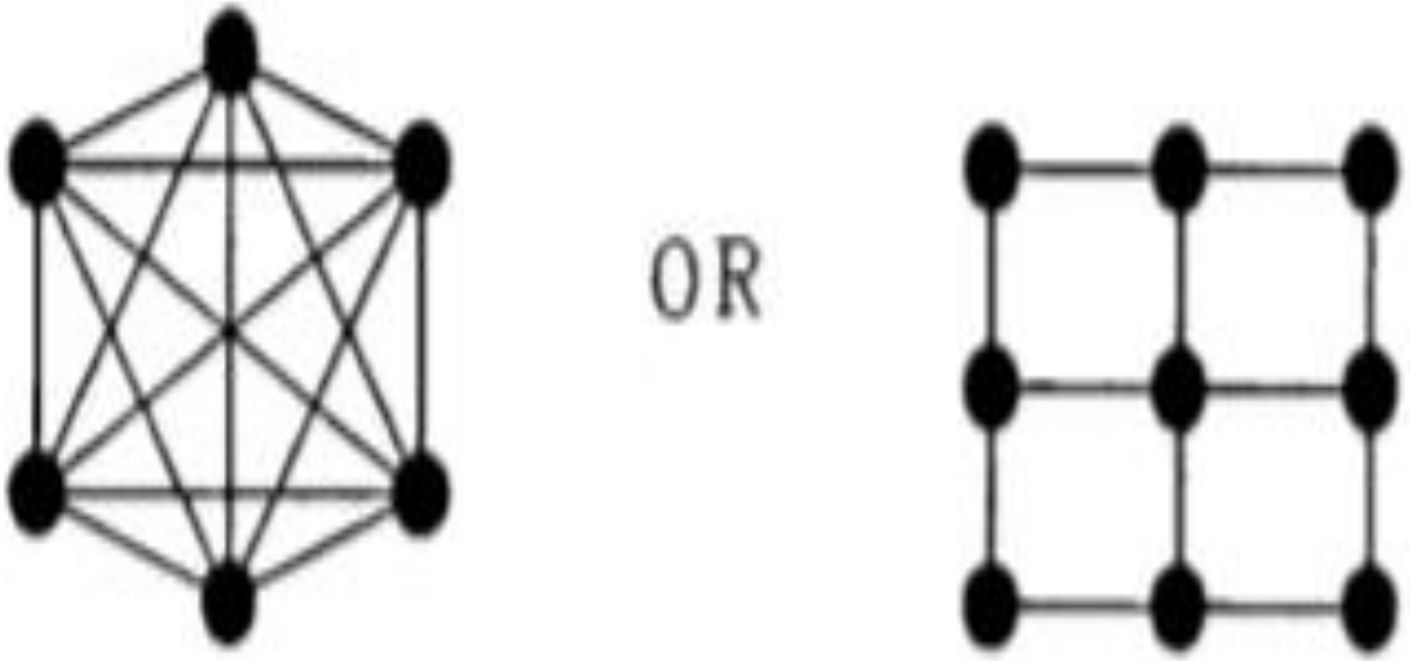


図1．ボルツマンマシン及びホップフィールドネットのネットワーク構造

□本日の解説内容と順番

１．AI（人工知能）に関する概要

２．データサイエンスで利用される化学パラメータ

３．データサイエンス手法の概要

４．KY法の説明

□湯田が開発した新規データ開発手法：

KY法（K-step Yard sampling methods）

KY法の基本技術：
多段階、リサンプリング、アンサンブル法をミックスして完成した最強力な**ニクラス分類手法**

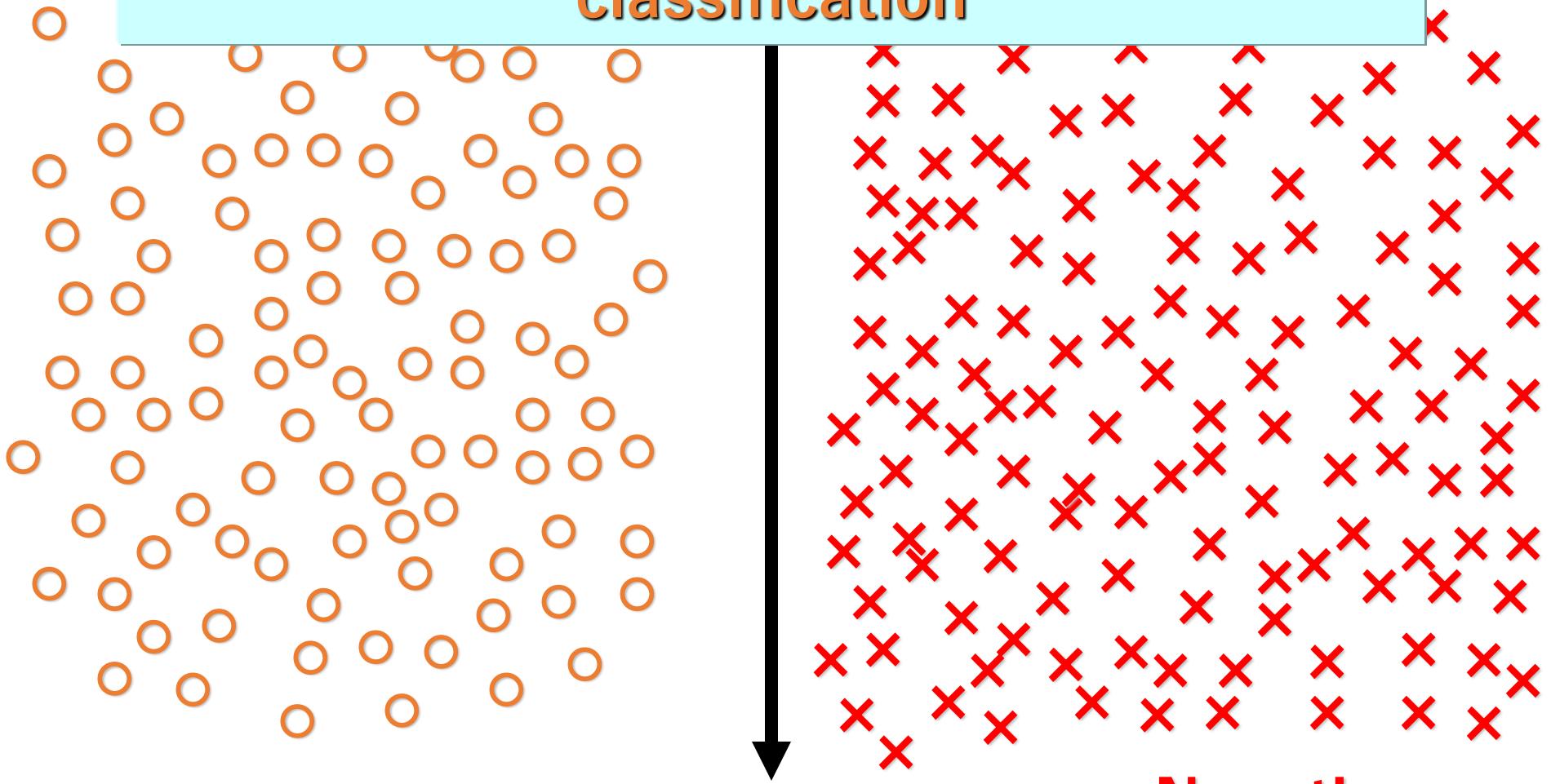
＊当初はニクラス分類手法のみであったが、現在はフィッテイング（重回帰）手法にも展開可能となっている。
＊判別関数や回帰式の作成様式の違いにより、ニクラス分類およびフィッテイング手法それぞれに３種類のKY法が創出された。
①２モデルKY法
②１モデルKY法
③モデルフリーKY法

Development of “K-step Yard sampling method” and Apply to the ADME-T In Silico Screening

Kohtaro Yuta

Sample space ：two cluster samples

Discriminant function for perfect classification



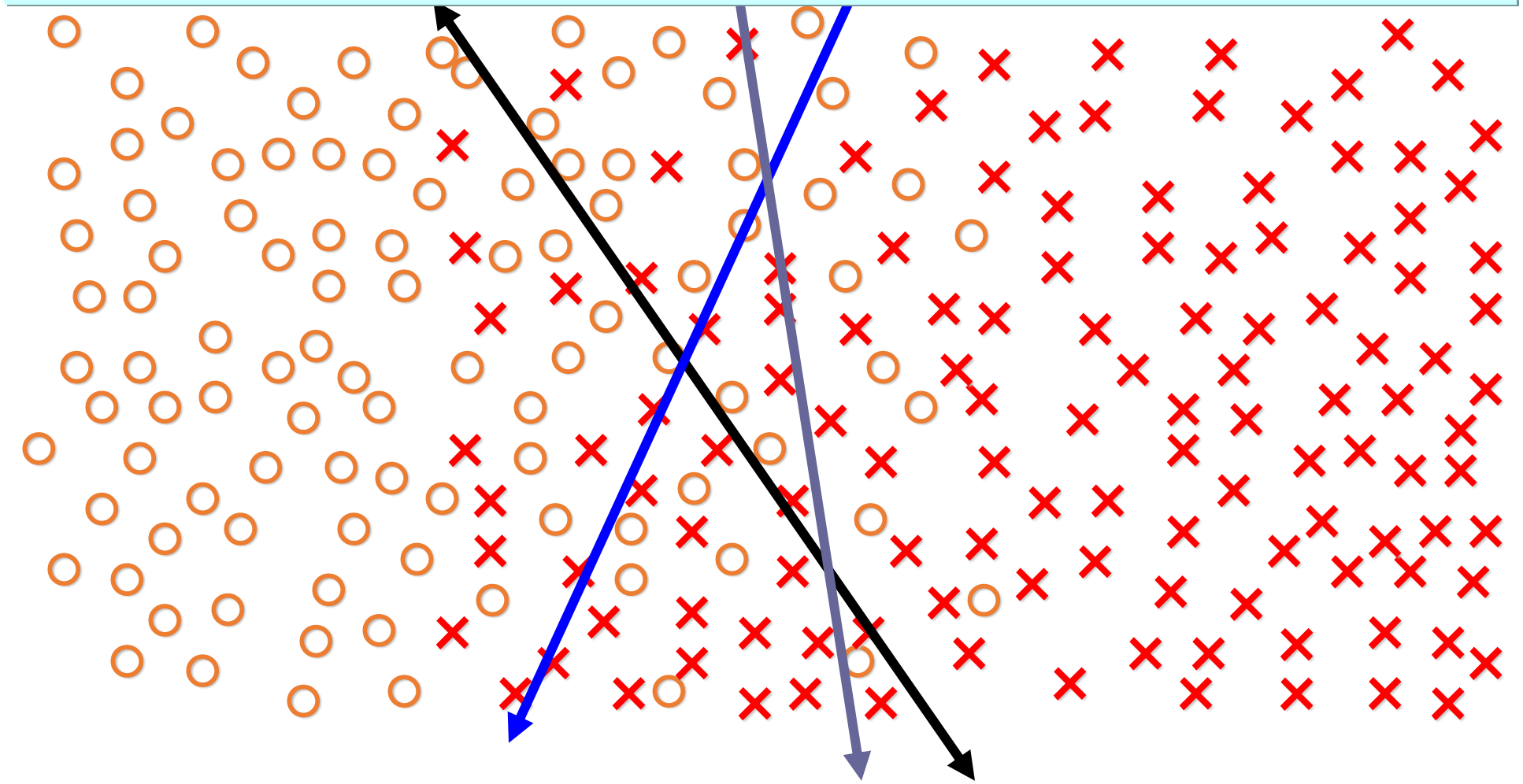
○：Positive

×：Negative

Sample space ：highly overlapped space

In Silico Data
Miracles by the KY-methods

Discriminant function generated by various methods



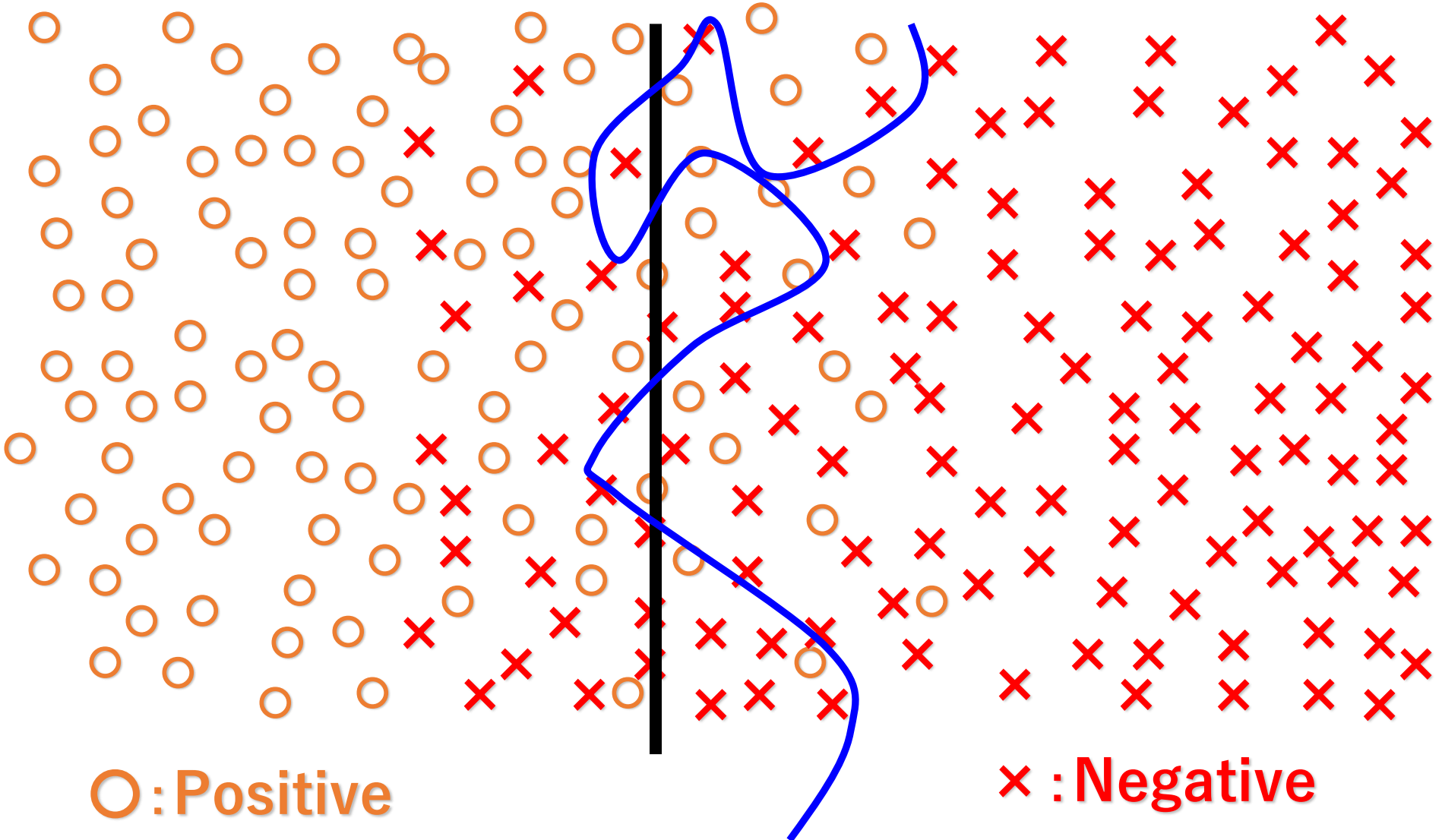
○：Positive

×：Negative

Sample space ：highly overlapped space

In Silico Data
Miracles by the KY-methods

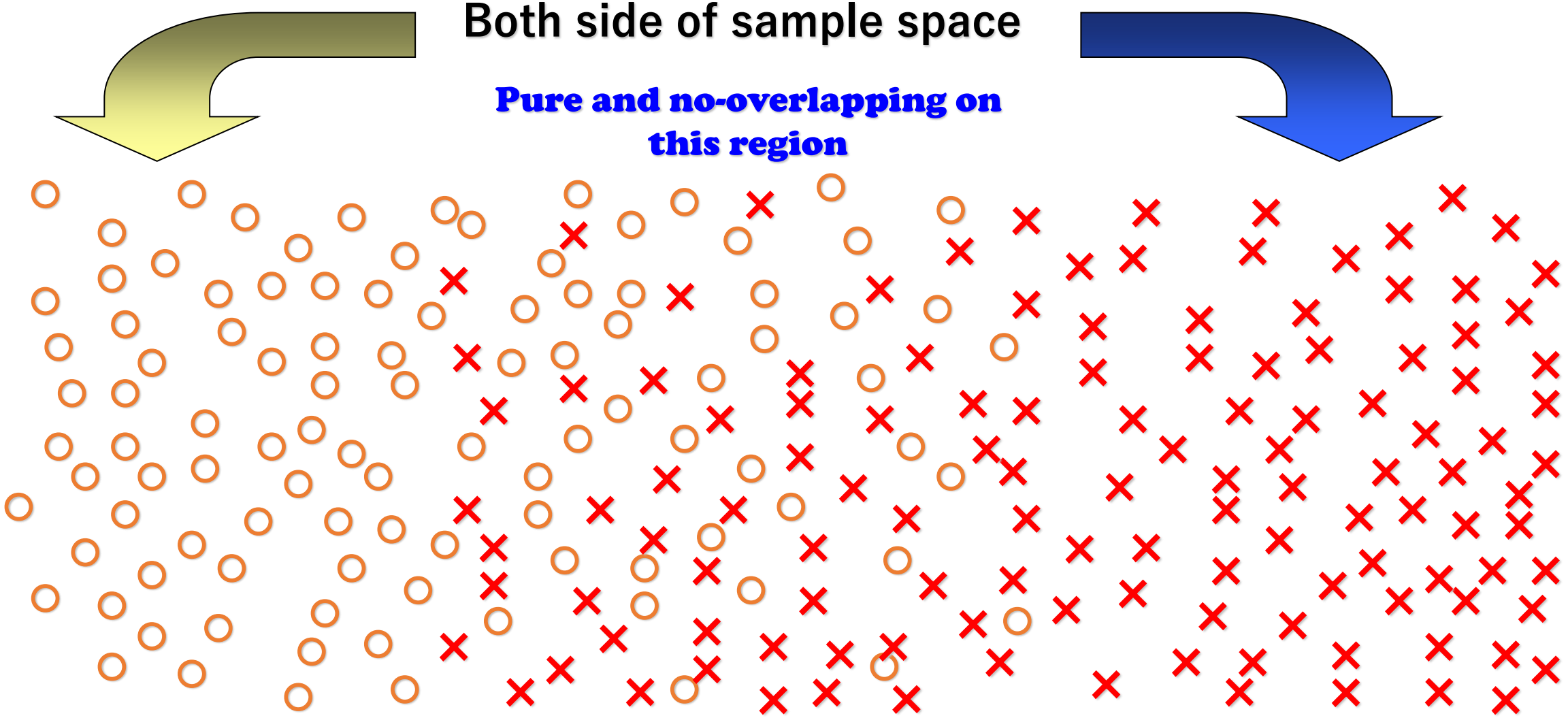
Discriminant function ： Linear and non-linear



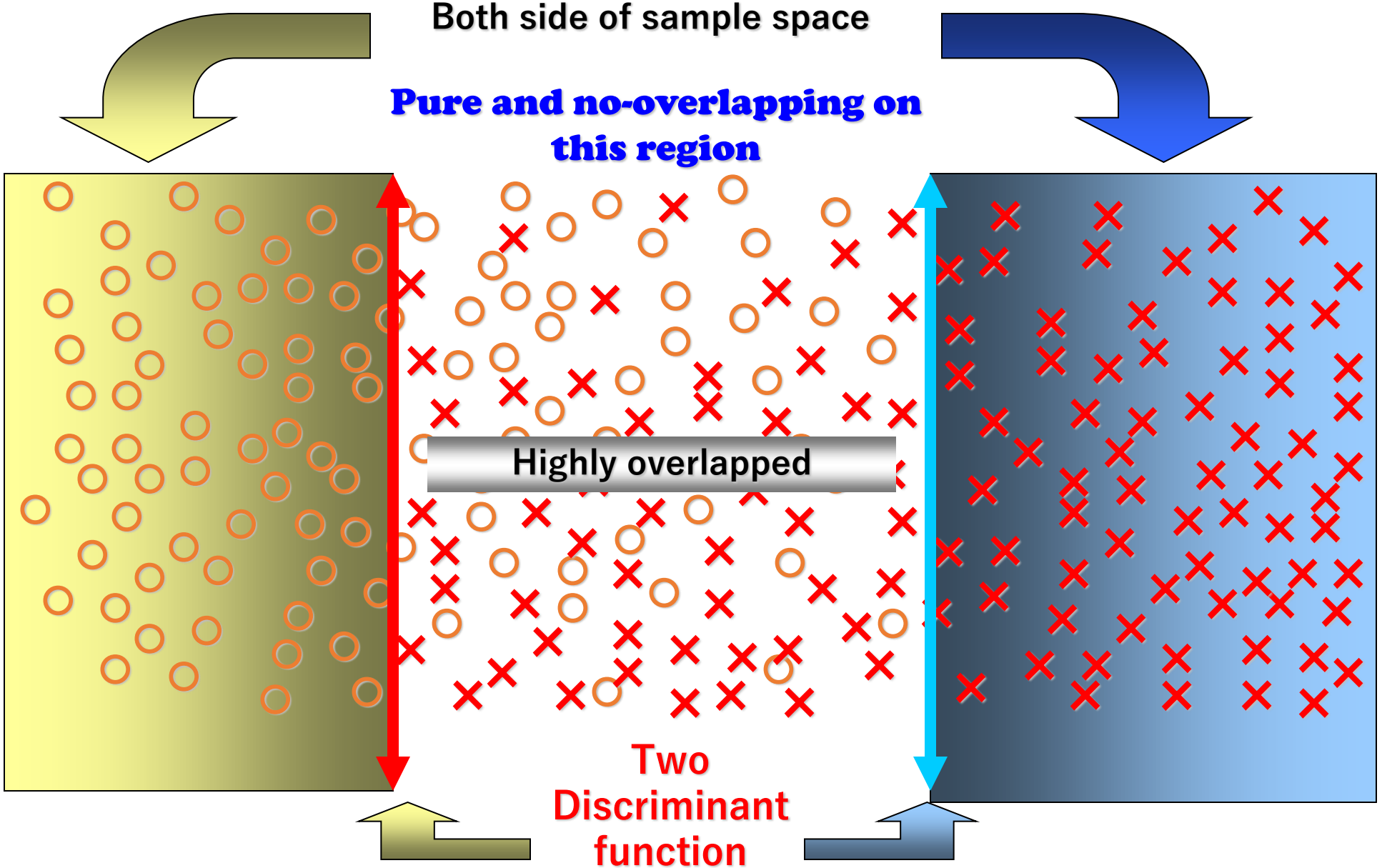
Building process to the features of “K-step Yard sampling method”

Step1: Yard sampling methods

Spatial region on sample space

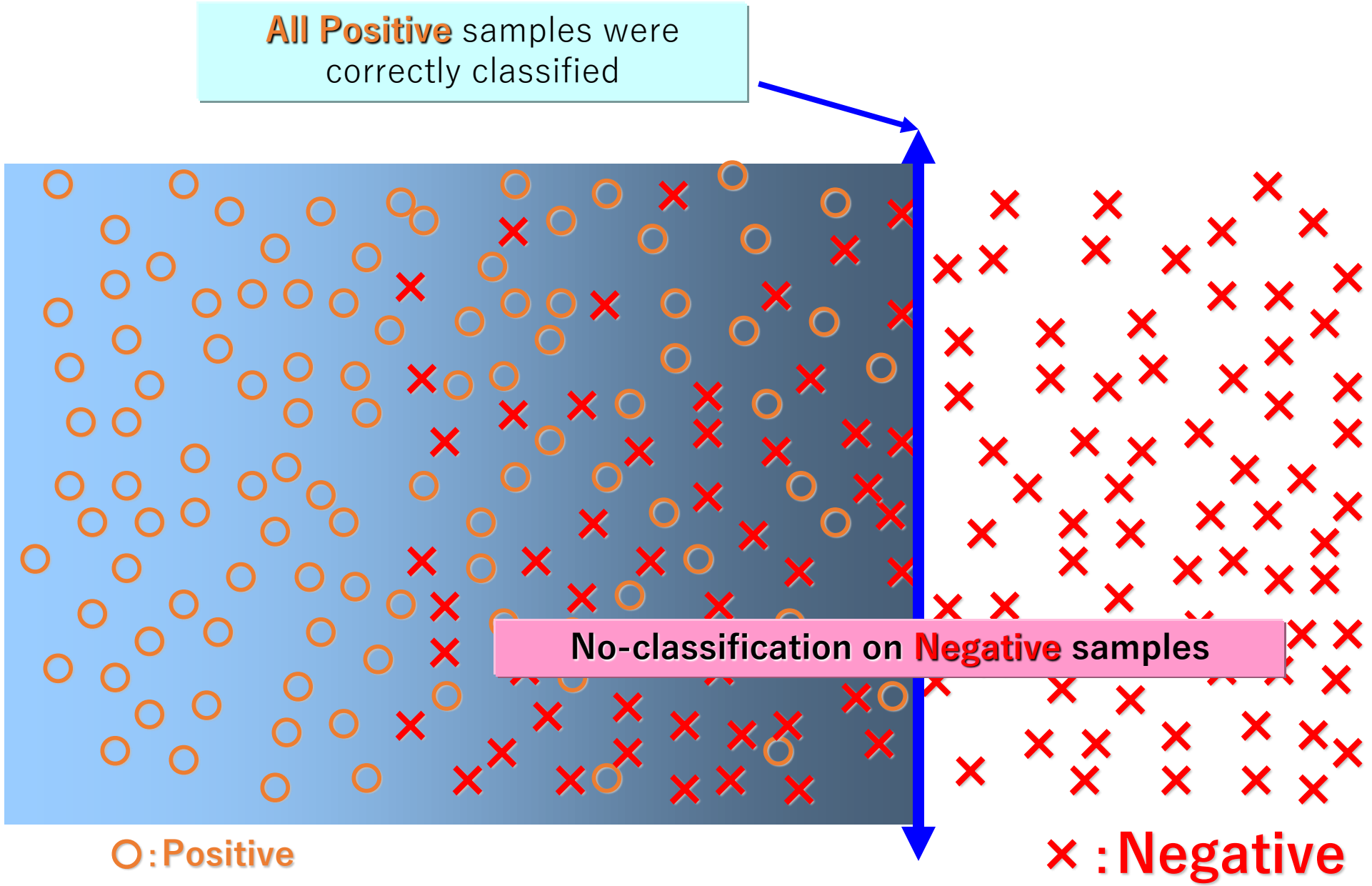


Spatial region on sample space



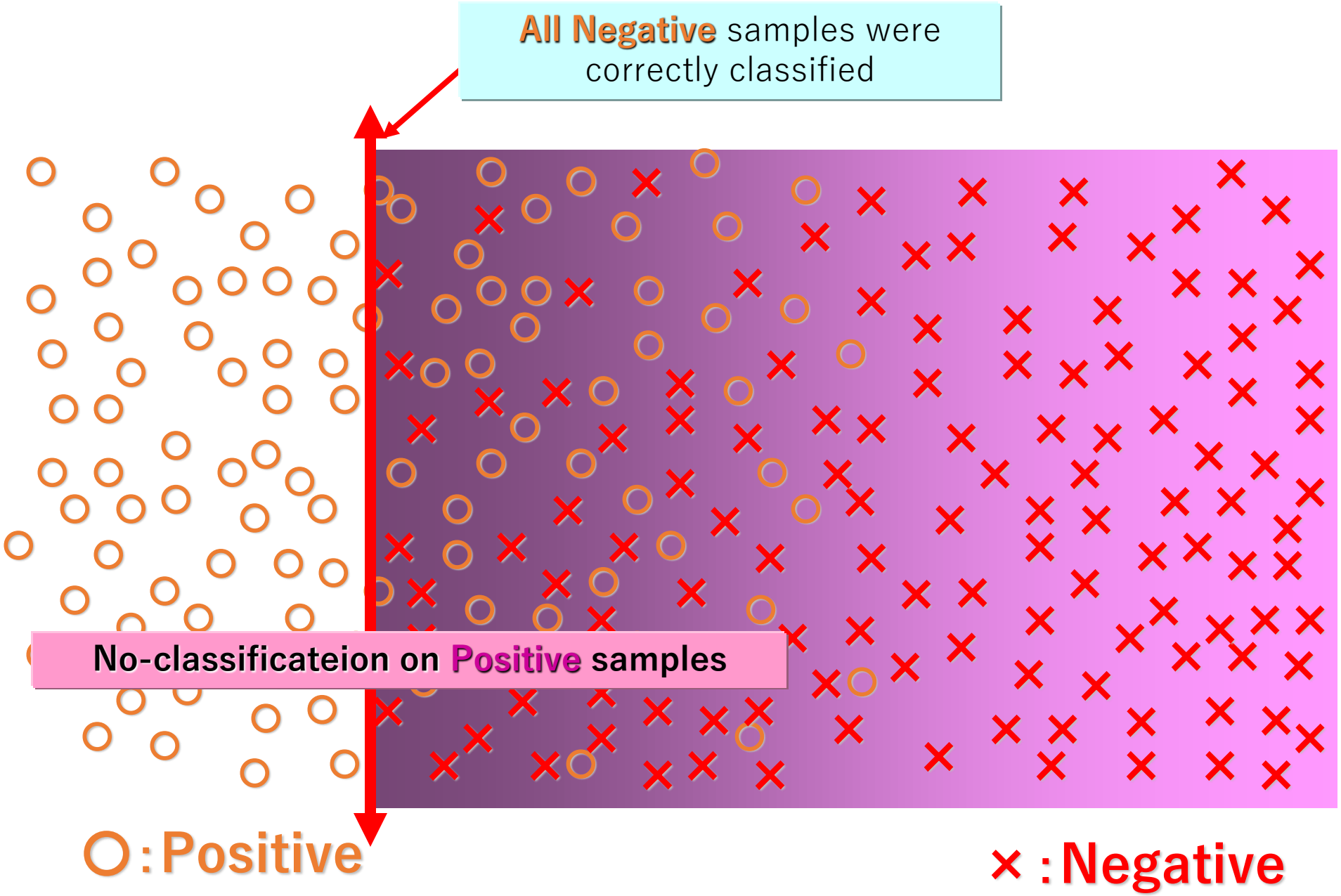
Property of AP(All Positive)model

In Silico Data
Miracles by the KY-methods



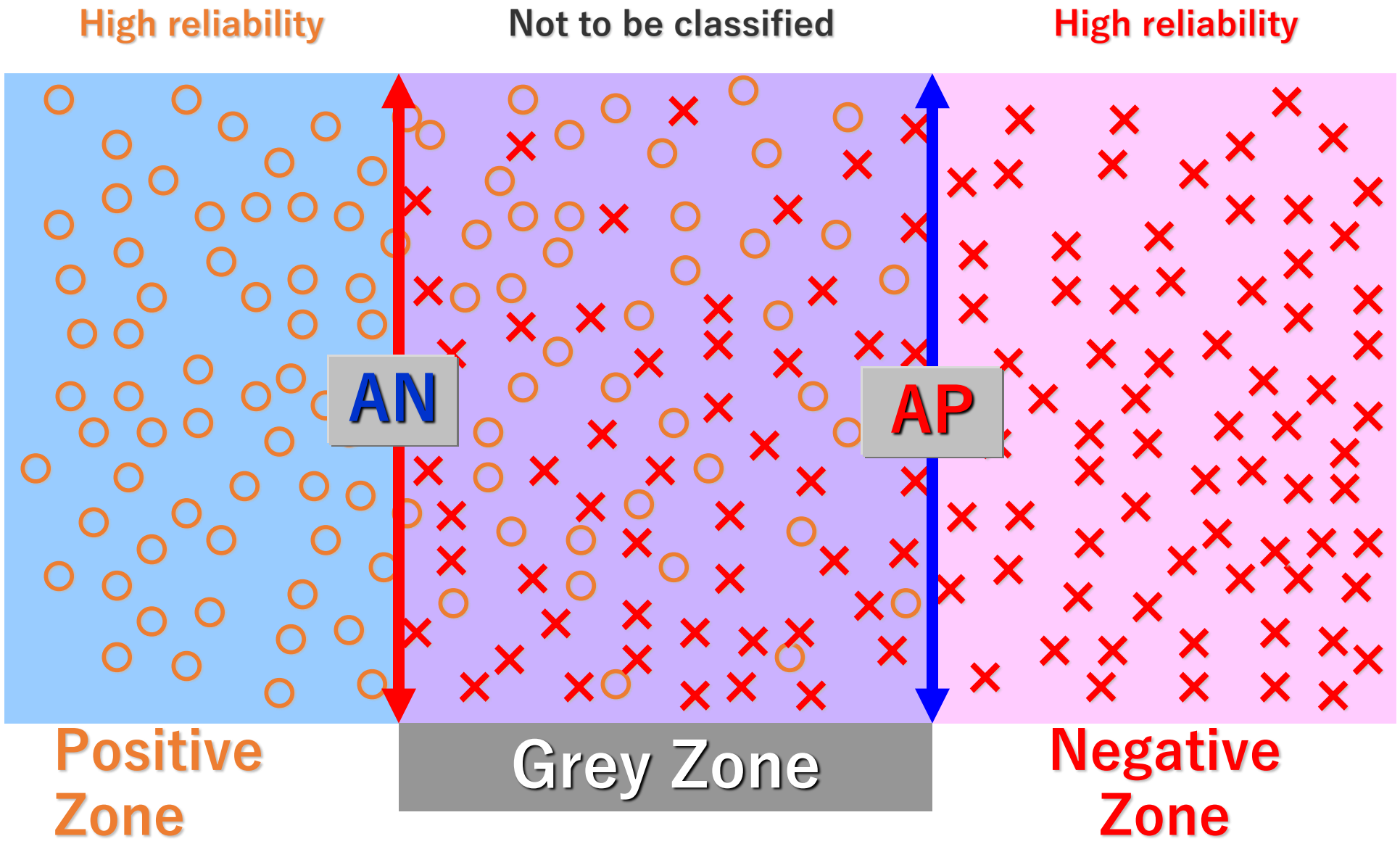
Property of AN(All Negative)model

In Silico Data
Miracles by the KY-methods

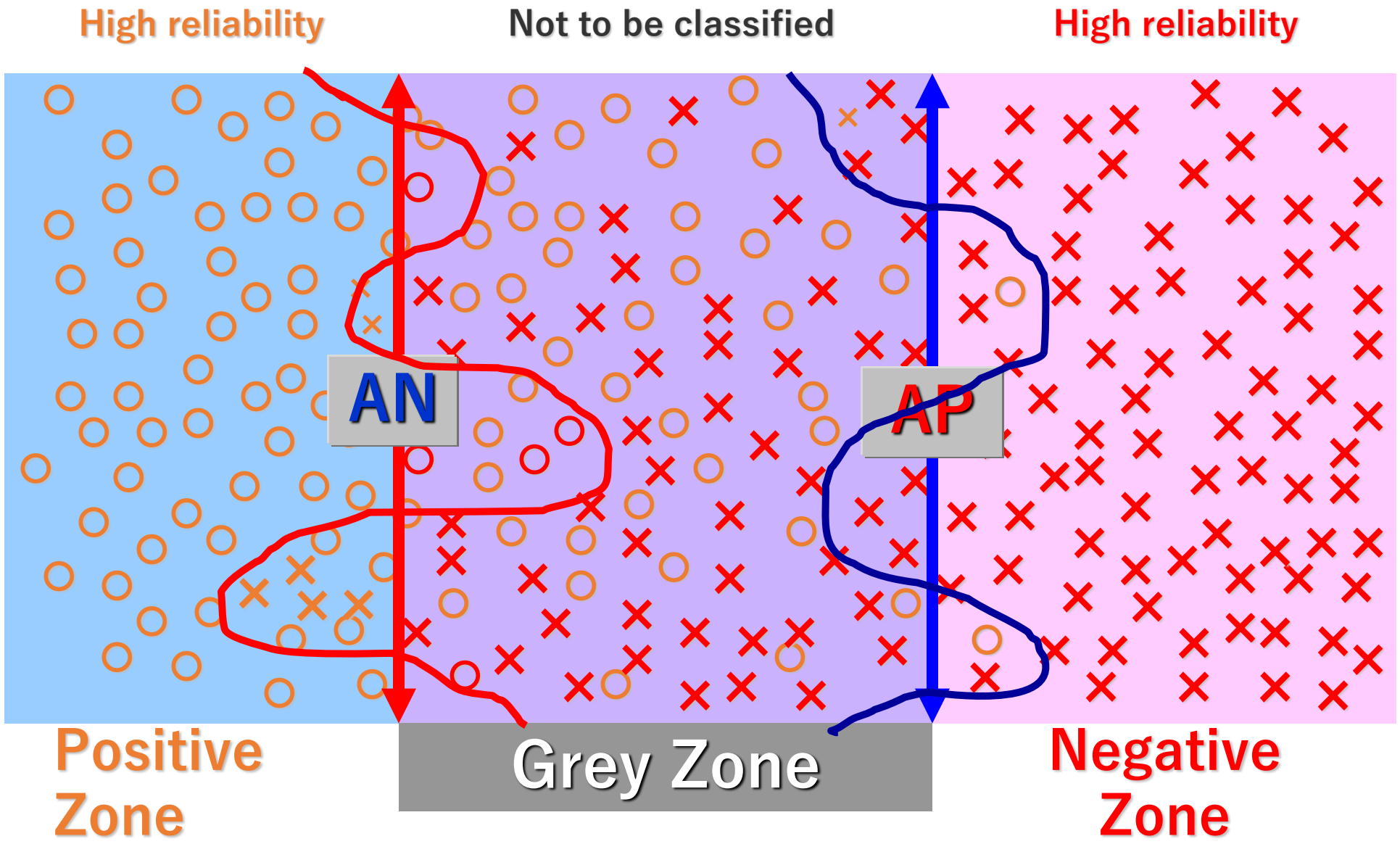


Combination of AN and AP models

In Silico Data
Miracles by the KY-methods

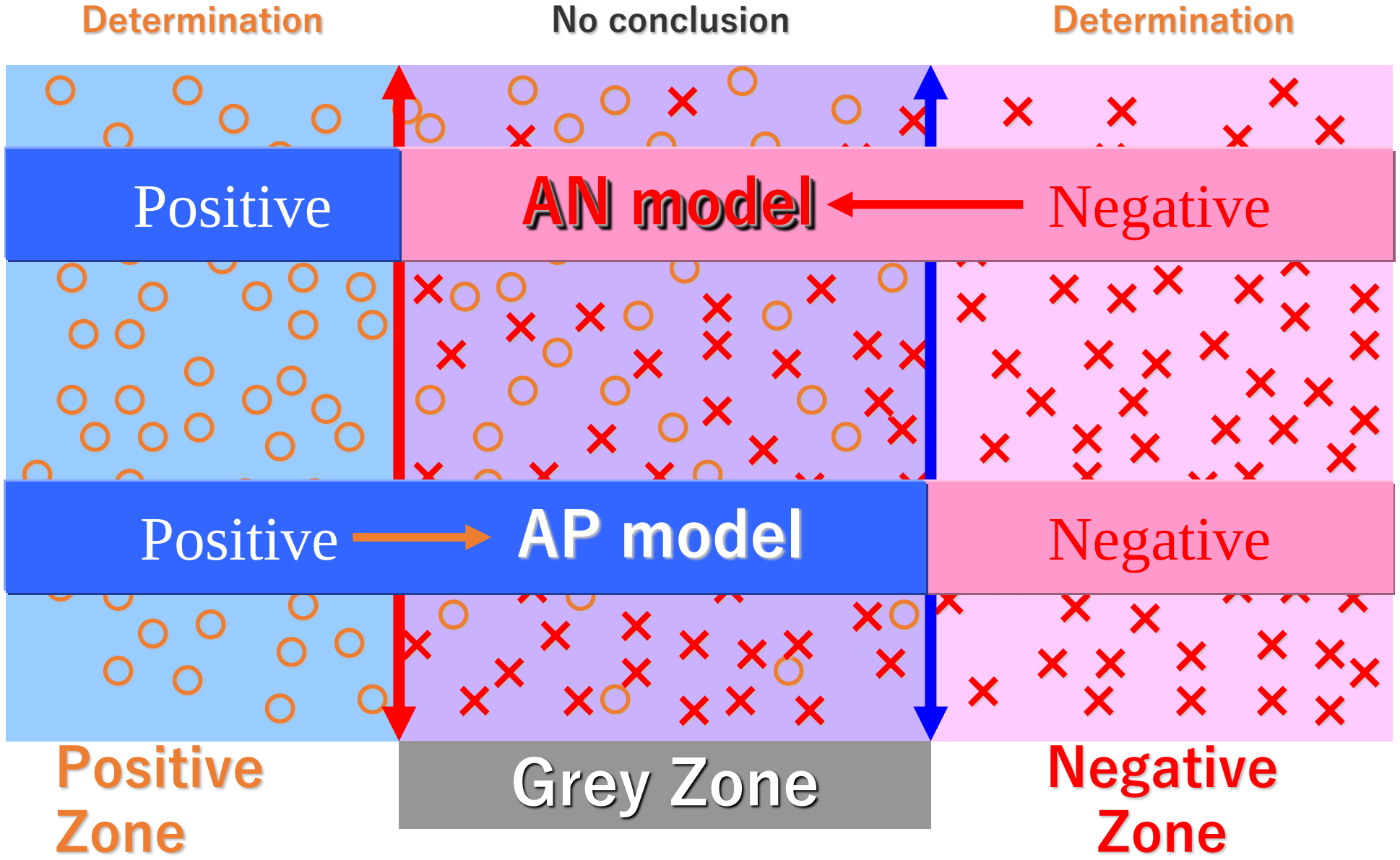


Linear and non-linear discriminant on AP and AN models



Relations between Sample space & AN and AP models

In Silico Data
Miracles by the KY-methods



Class determination by AN and AP models

In Silico Data
Miracles by the KY-methods

☐ Sample Classification and prediction must be done by Combination of the results of AP and AN models.

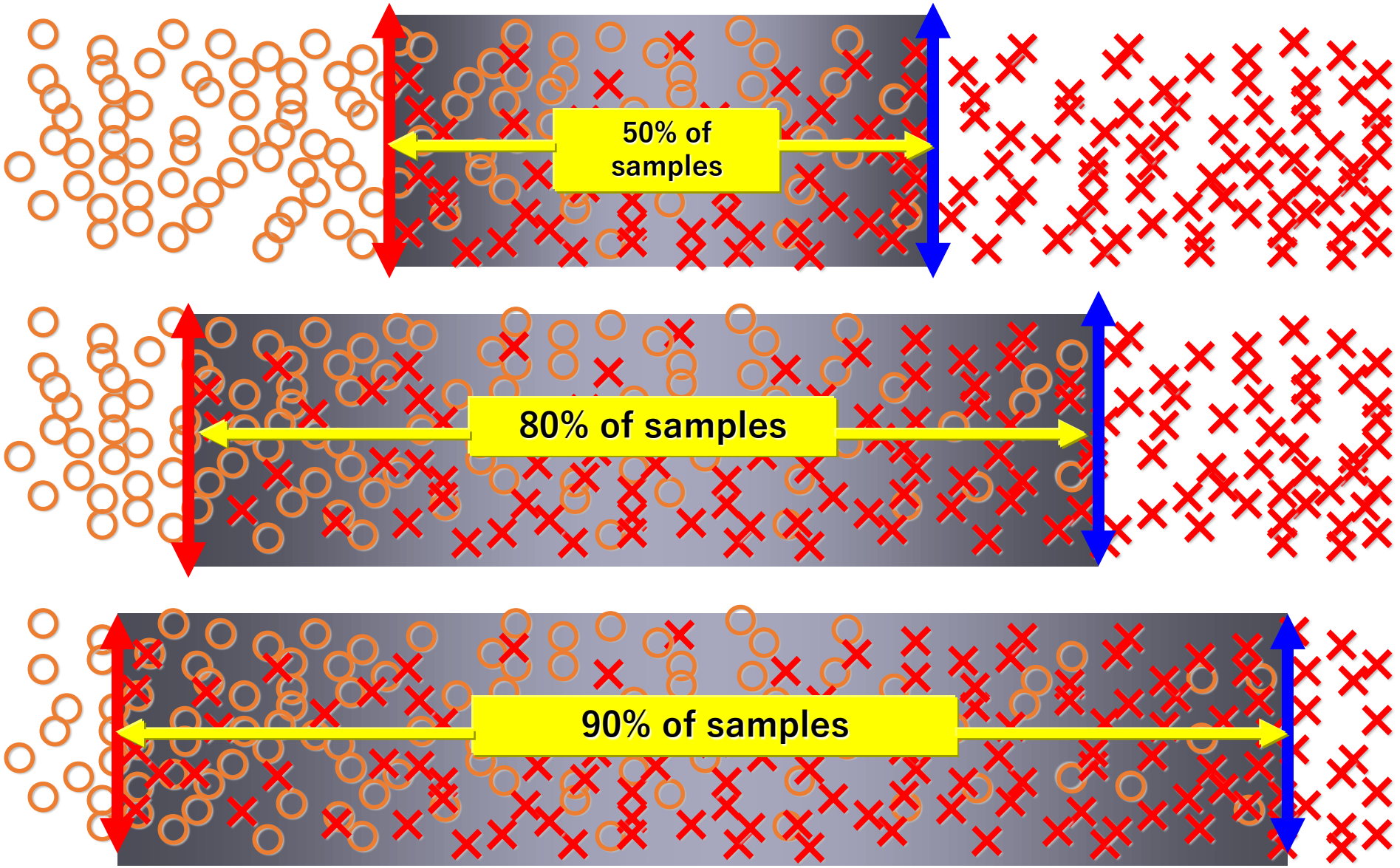
AP model	AN model		Results
①AP ; POSI,	AN ; POSI	→	POSI
②AP ; POSI,	AN ; NEGA	→	GREY
③AP ; NEGA,	AN ; POSI	→	GREY
④AP ; NEGA,	AN ; NEGA	→	NEGA

Building steps to the features of “K-step Yard sampling method”

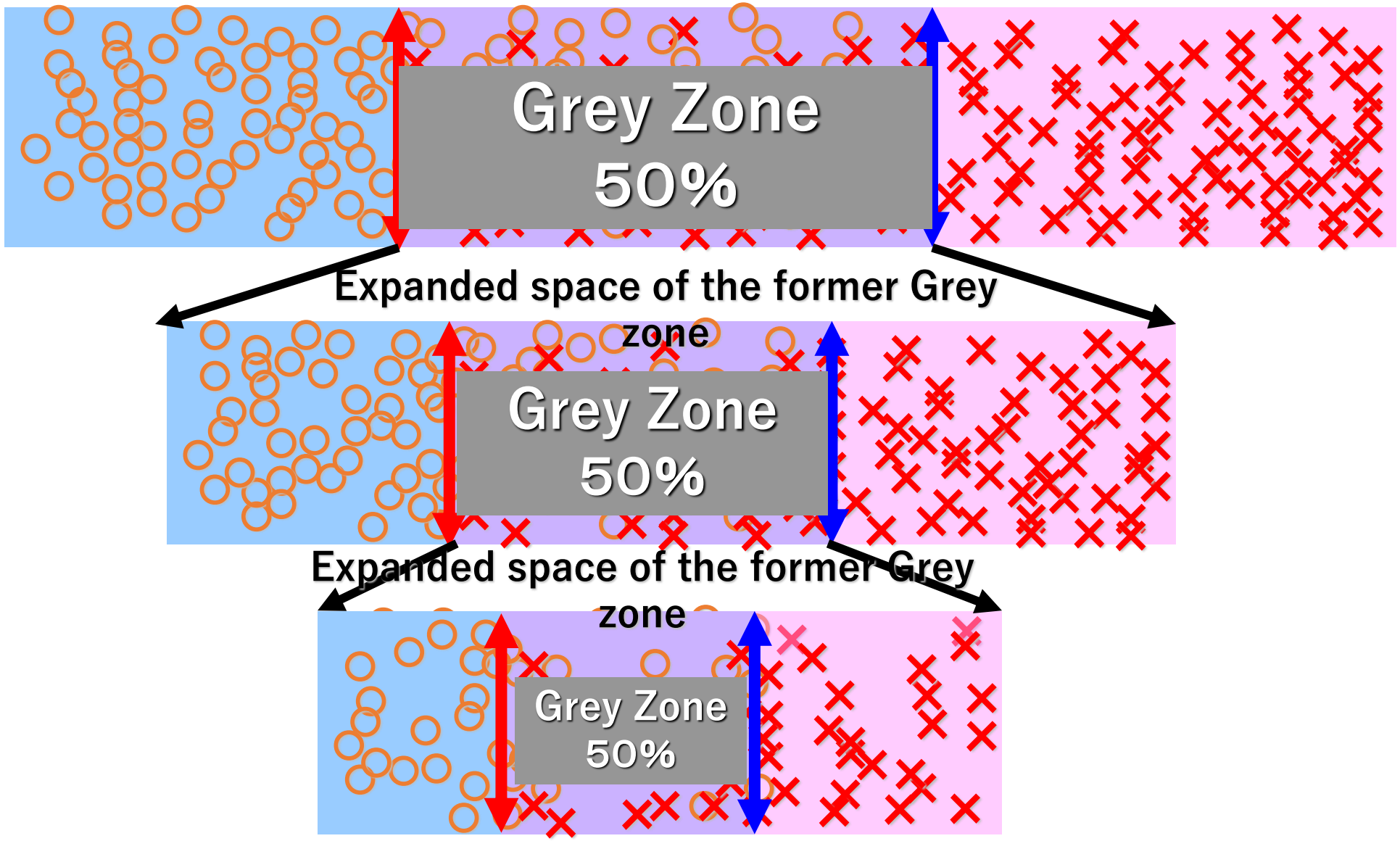
Step2: K-step approach

Problems of Yard sampling methods

The ratio of Grey zone:Highly overlapped sample space



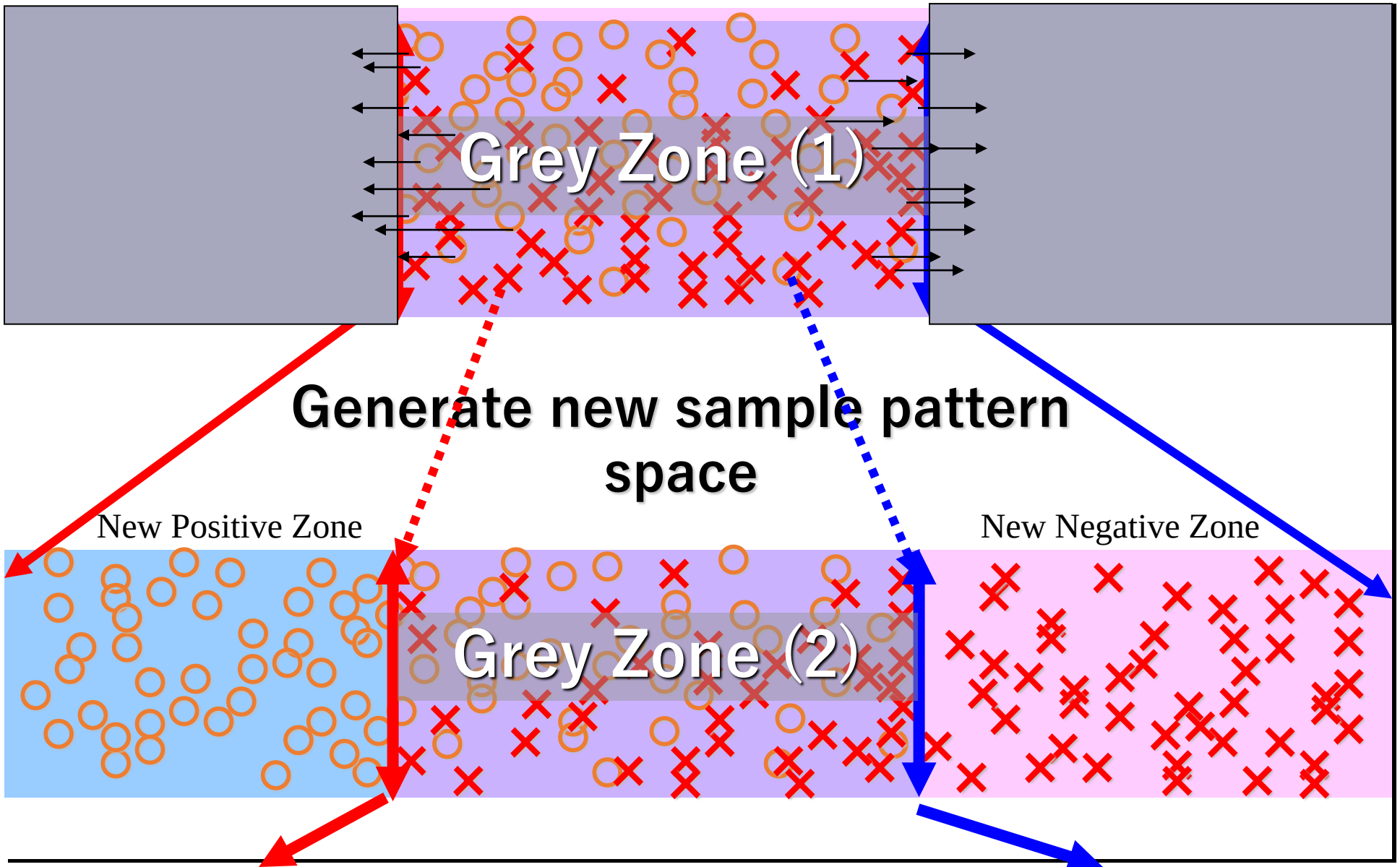
Steps to the K-step methods



“K-step Yard sampling (KY) Method”

In Silico Data
Miracles by the KY-methods

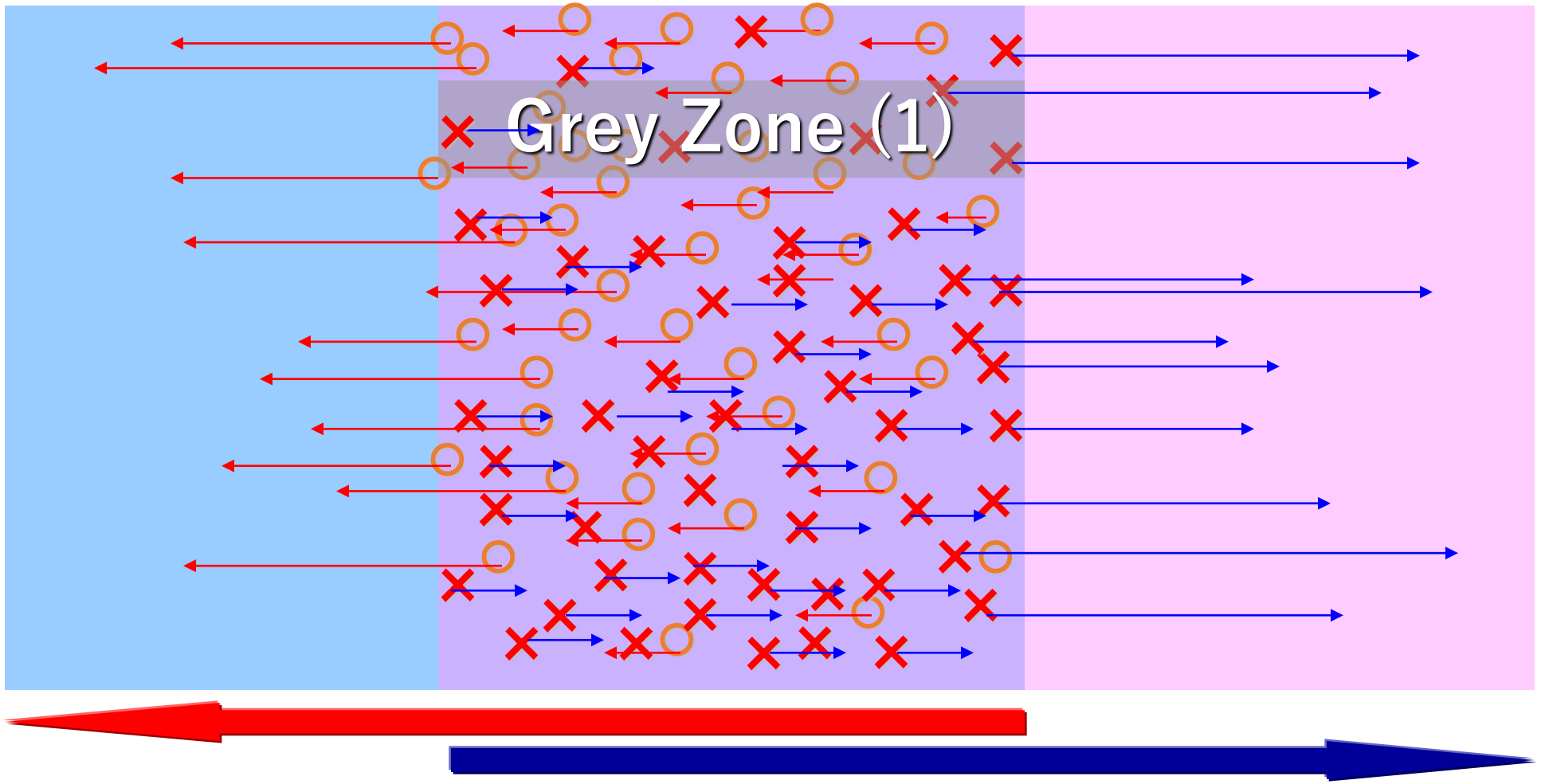
Improvement by repeated classification of Grey Zone samples



“K-step Yard sampling (KY) Method”

In Silico Data
Miracles by the KY-methods

□ Relocation of Grey Zone samples on new sample space



Applicability statement of "K-step Yard sampling method"

Classifying 7000 sample set of Ames test

Challenge for classification and prediction

In Silico Data
Miracles by the KY-methods

K-step Yard sampling method
KY-method

The most powerful and advanced data analysis method

The most difficult classification problem

6,965 sample of Ames test samples were,

Classified perfectly

Application test of “K-step Yard sampling”

In Silico Data
Miracles by the KY-methods

□Samples

1. Ames test data
2. Sample population
   total :6,965
   Mutagen; 2,932
   Non-mutagen; 4,033

□Result of KY-method

1. Number of steps : 23 steps ; 22 (2 models) + 1 (1 model)
2.Classification ratio : 100 %

□Used system

ADMEWORKS / ModelBuilder
V 3.0.22

□Used parameters (Initial condition)

Number of generated parameters : 838
Number of parameters for step 1 : 98
Confidency index (Samples(6965) / Parameters(98)) : 71.1 > 4.0

Application test results by various D.A. methods

1. Linear discriminant analysis with linear least-squares method

Classification ratio : total; 73.50(6965), Mutagen;73.02(2932), Non mutagen;73.84(4033)
Number of mis-classified : (1846), (791) (1055)

Prediction ratio (L100 out) 72.58% deviance(0.92%)
(L500 out) 73.32% deviance(0.18%)

---

2. SVM (Support Vector Machine with Kernel)

Classification ratio : total; 90.87(6965), Mutagen;86.83(2932) Non mutagen; 93.80(4033)
Number of mis-classified : (636), (386) (250)
Prediction ratio (L500 out) 80.99% deviance(9.88%)

---

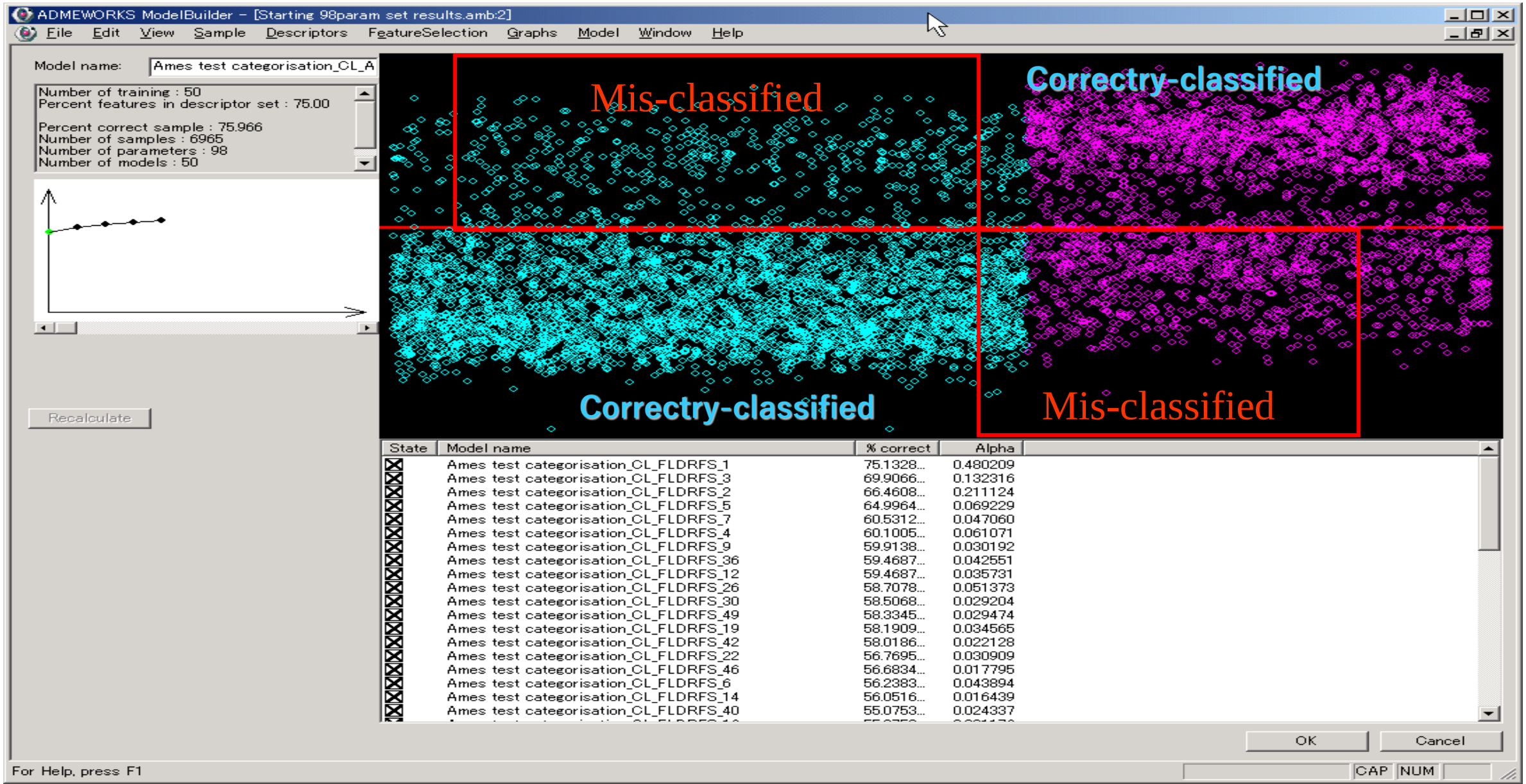
3. AdaBoost

Classification ratio : total; 77.24(6965), Mutagen;66.13(2932) Non-mutagen; 85.32(4033)
Number of mis-classified : (1585) (993) (592)
Prediction ratio (L500 out) 75.16% deviance(2.08%)

Classification results by AdaBoost

In Silico Data
Miracles by the KY-methods

Sample distribution of 6,965 of 77.24%



“K-step Yard sampling (KY) Method”

In Silico Data
Miracles by the KY-methods

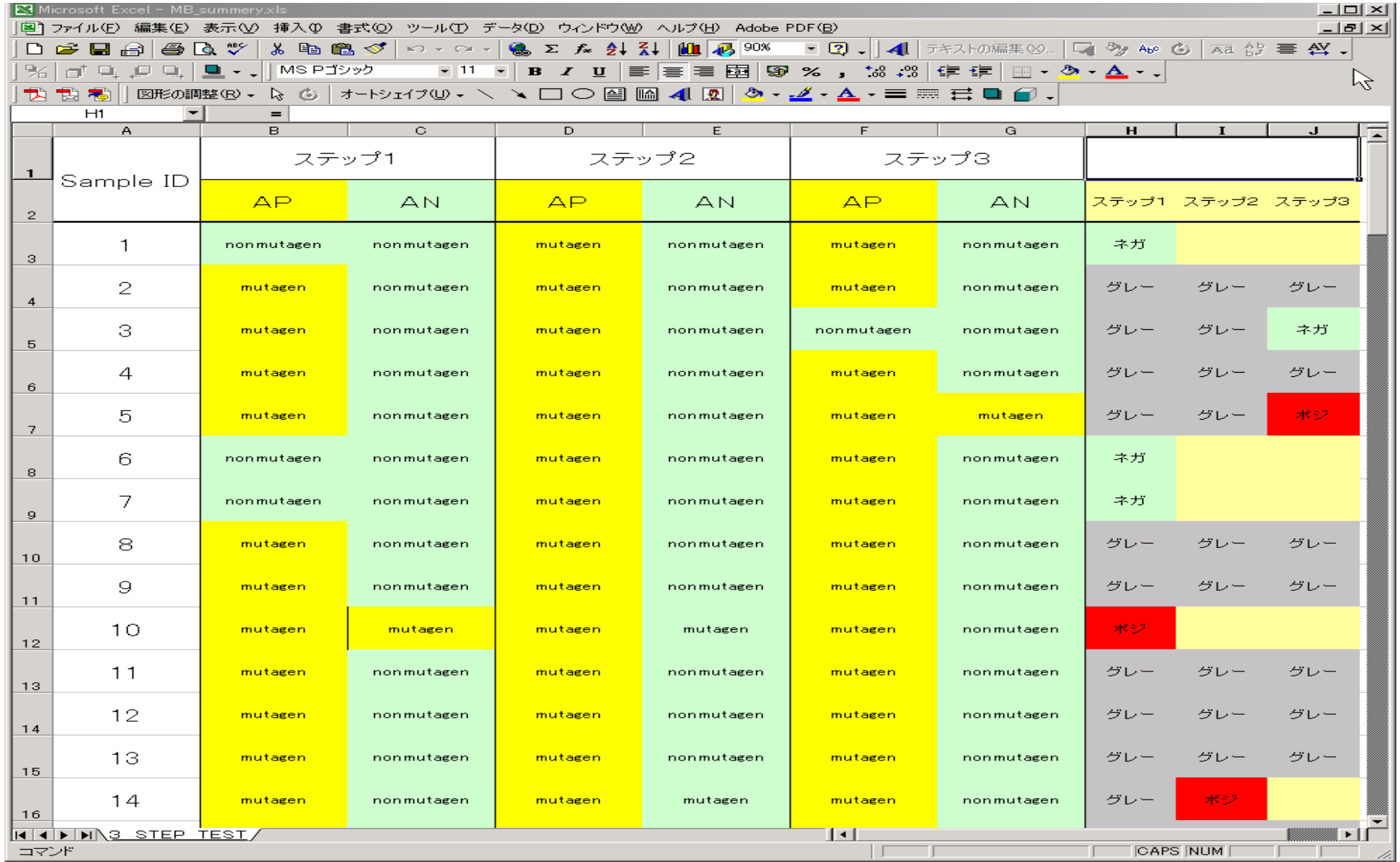
Total steps : 23 steps (2 models) + 1 step (1 model)

ステップID（KY法）	Starting samples(Total)	Mutagen (Initial)	Non-mutagen (Initial)	Grey sample (Initial)
	Final samples	Mutagen (Final)	Non-mutagen (Final)	Grey sample (Final)
	Determined samples(Total)	Determined samples(Mut.)	termined samples(Non-mu	Grey ratio(%) (Grey/Total)
1	6965	2932	4033	0
	5864	2413	3451	5864
	1101	519	582	84.19
2	5864	2413	3451	5864
	5108	2142	2966	5108
	756	271	485	87.11
3	5108	2142	2966	5108
	4486	1919	2567	4486
	622	223	399	87.82
4	4486	1919	2567	4486
	4133	1779	2354	4133
	353	140	213	92.13
5	4133	1779	2354	4133
	3794	1651	2143	3794
	339	128	211	91.8
6	3794	1651	2143	3794
	3462	1485	1977	3462
	332	166	166	91.25
7	3462	1485	1977	3462
	3090	1345	1745	3090
	372	140	232	89.25
8	3090	1345	1745	3090
	2826	1220	1606	2826
	264	125	139	91.46
9	2826	1220	1606	2826
	2592	1139	1453	2592
	234	81	153	90.63
10	2592	1139	1453	2592
	2384	1047	1337	2384
	208	92	116	91.98

“K-step Yard sampling (KY) Method”

In Silico Data
Miracles by the KY-methods

Classification results by 3 steps



Sample ID	ステップ1		ステップ2		ステップ3				
	AP	AN	AP	AN	AP	AN	ステップ1	ステップ2	ステップ3
1	nonmutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	ネガ		
2	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	グレー	グレー	グレー
3	mutagen	nonmutagen	mutagen	nonmutagen	nonmutagen	nonmutagen	グレー	グレー	ネガ
4	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	グレー	グレー	グレー
5	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	mutagen	グレー	グレー	ポジ
6	nonmutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	ネガ		
7	nonmutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	ネガ		
8	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	グレー	グレー	グレー
9	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	グレー	グレー	グレー
10	mutagen	mutagen	mutagen	mutagen	mutagen	nonmutagen	ポジ		
11	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	グレー	グレー	グレー
12	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	グレー	グレー	グレー
13	mutagen	nonmutagen	mutagen	nonmutagen	mutagen	nonmutagen	グレー	グレー	グレー
14	mutagen	nonmutagen	mutagen	mutagen	mutagen	nonmutagen	グレー	ポジ	

Spatial features of “K-step Yard sampling”

In Silico Data
Miracles by the KY-methods

□Summary

■Advantages

1. Sample number free approach

2. Sample distribution free approach

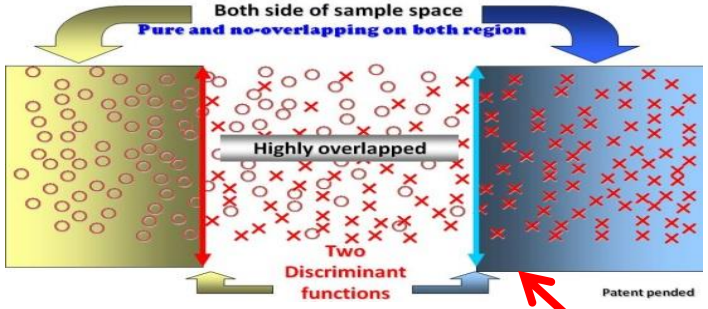
3. Perfect classification is achieved in any condition

■Disadvantages

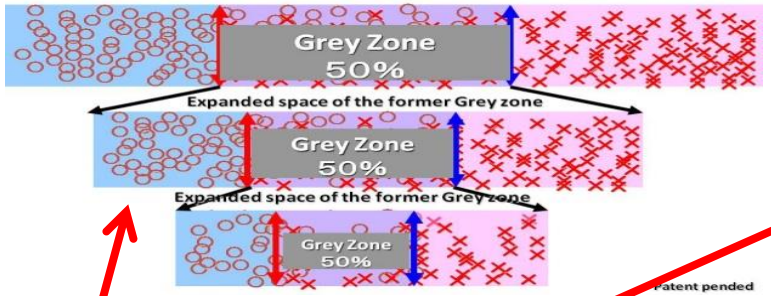
1. Relatively complex operation to generate discriminant functions

2. Need powerful computer power

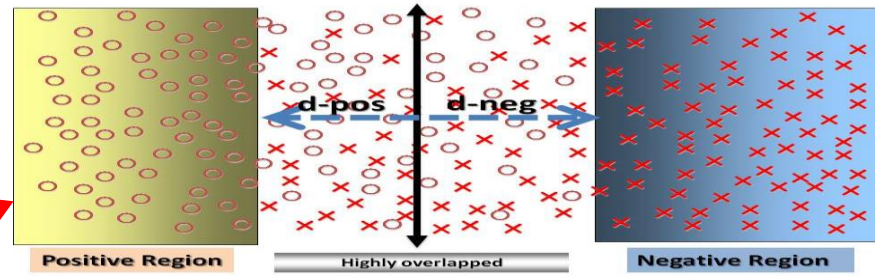
First basic concept of KY method
Spatial region on sample space



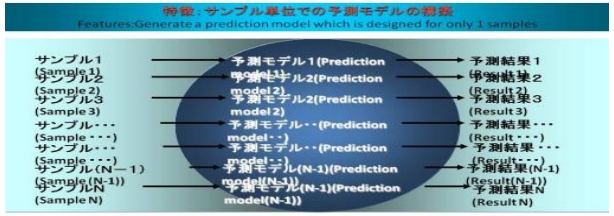
Second basic concept of KY method
Multi-steps for 100% classification



New approach to the "KY method" by one discriminant function



「テーラーメード・モデリング」による予測アプローチ
(Prediction approach by "Tailor-Made Modeling")



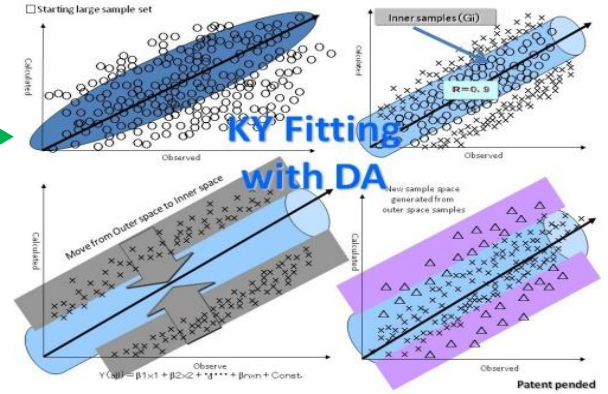
利点(Merit) ： 予測率が大幅に向上する(High prediction ratio will be achieved)
難点(Weakness)： 計算時間がかかる(Need large calculation time)

In Silico Data
Miracles by the KY-methods

A series of KY methods

Discriminant Analysis	Fitting
Two model KY	KY Fitting with DA
Single model KY	KY Fitting with no DA
Model free KY	Model free KY Fitting

*Always carry perfect classification　　*Always high coefficient of determination

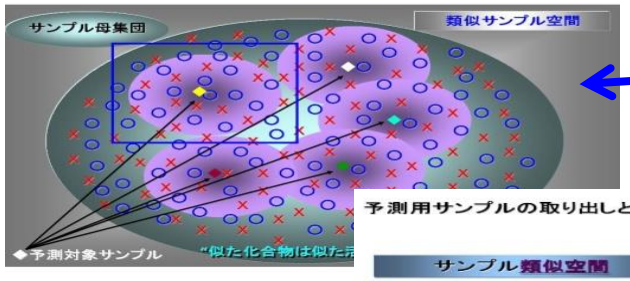


Discriminant Analysis	Fitting
Tailor-made Modeling for DA	Tailor-made Modeling for Fitting

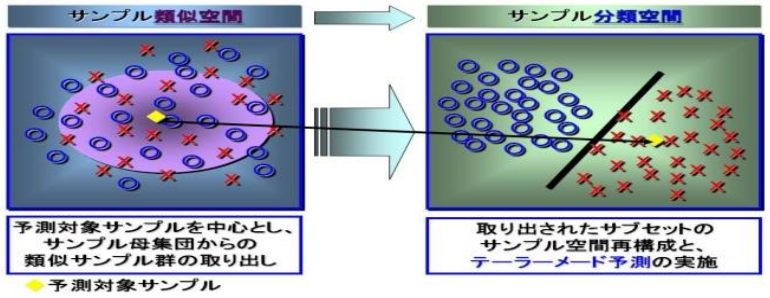
*Always carry high prediction ratio

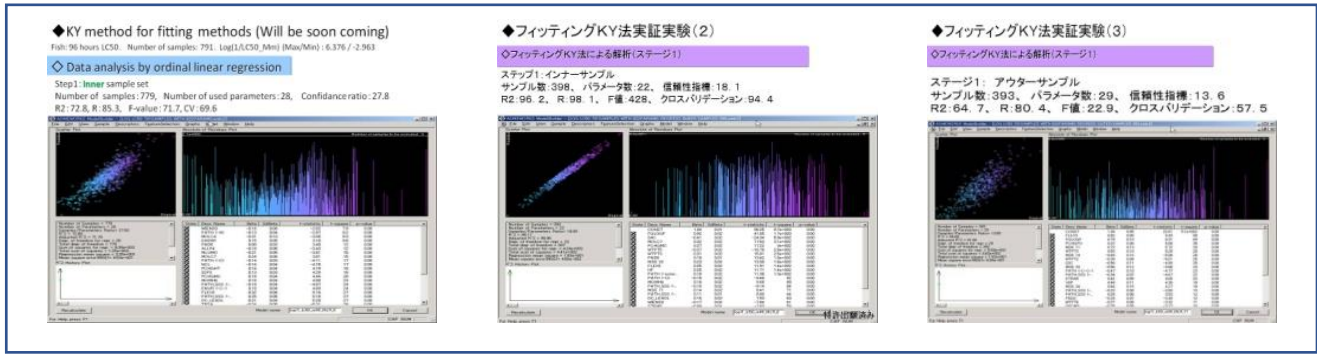
All methods were Patent Pended

サンプル母集団からの予測用サンプルの取り出し



予測用サンプルの取り出しと、テーラーメード予測





TS-02 第四章　適用留意事項に続く