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

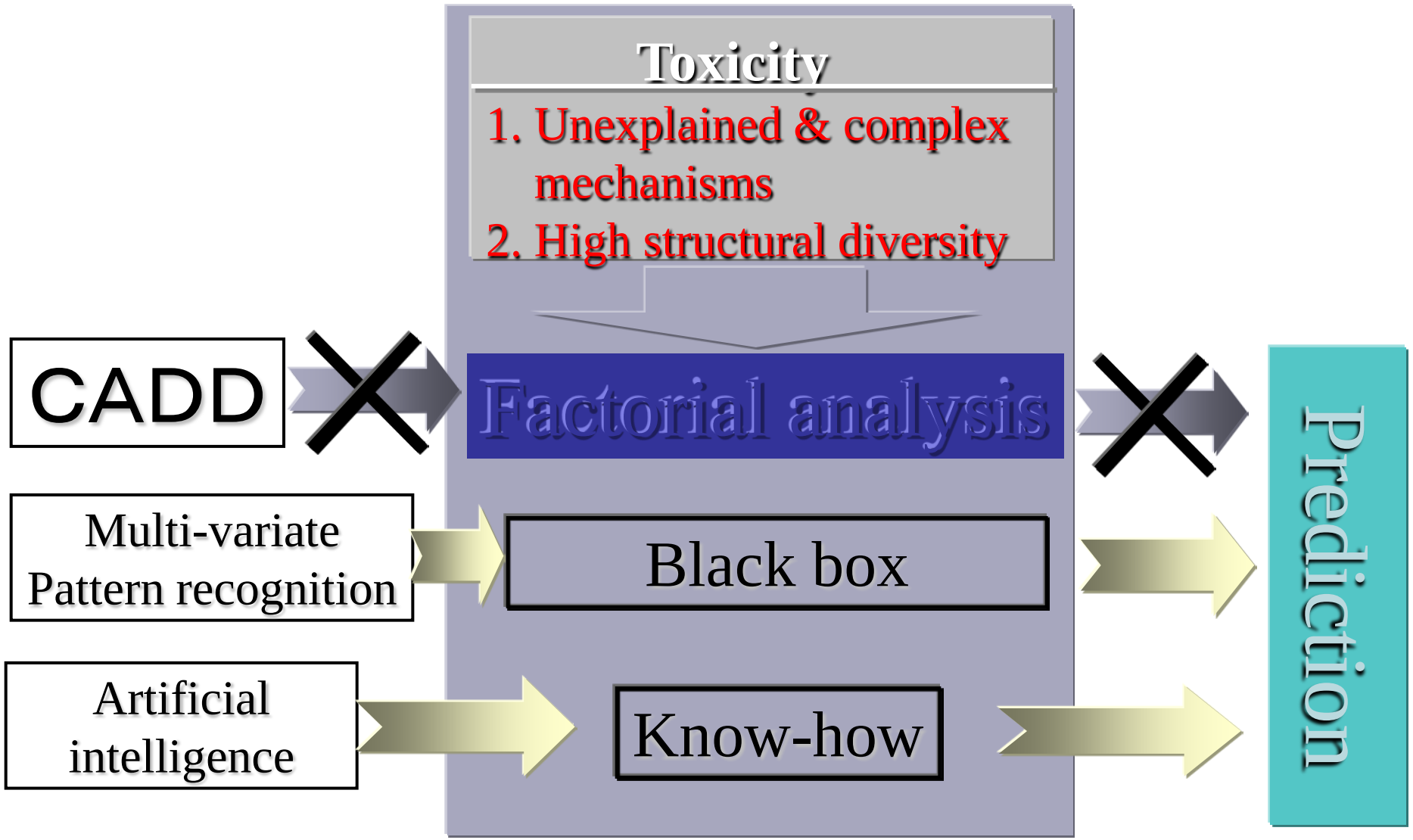
Kohtaro Yuta

In Silico Data, Ltd.

In Silico Data
Miracles by the KY-methods

1. Toxicity prediction and Pattern recognition (PR)
2. General features of data analysis by PR.
3. Building process to the features of “KY-method”
   *Step1 ;Yard sampling methods
   *Step2 ; K-step approach
   *Step3 ; Merge two approaches
   Yard sampling and K-step handling
4. Applicability statement of “KY-method”
   Classifying 7000 sample set of Ames test
5. Summary and conclusion

Approaches for toxicity screening



Problems of toxicity screening by pattern recognition

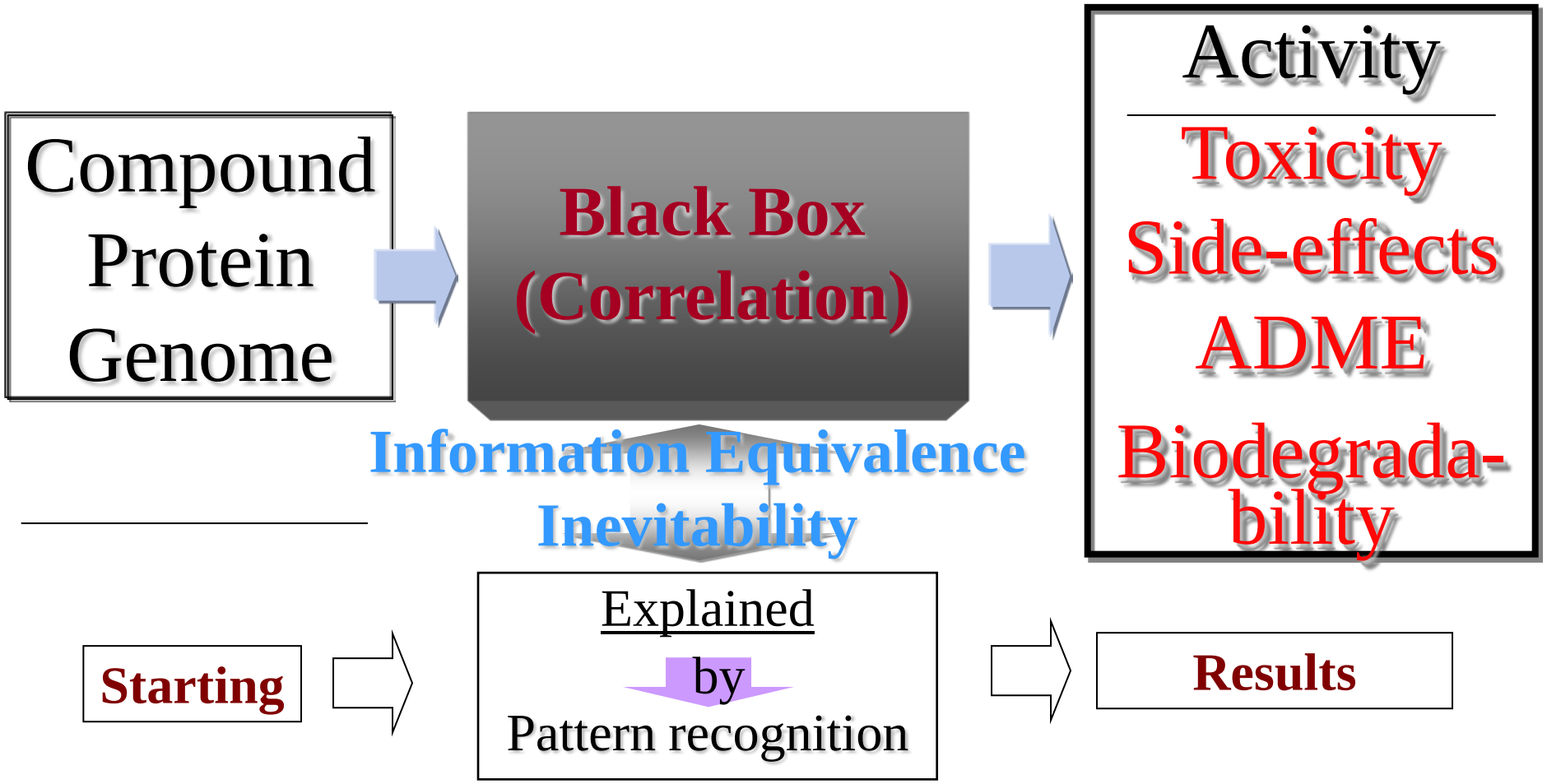
■ **Only a few methods can be applicable on toxicity screening**

Most of drug design methods can not be applied.

- **Un-known mechanisms →**
  Inability of "Hypothesis testing" method
- **Extremely high compounds diversity →**
  From methane to macrolide
- **Large sample population →**
  Normal D.D. approach handle small samples

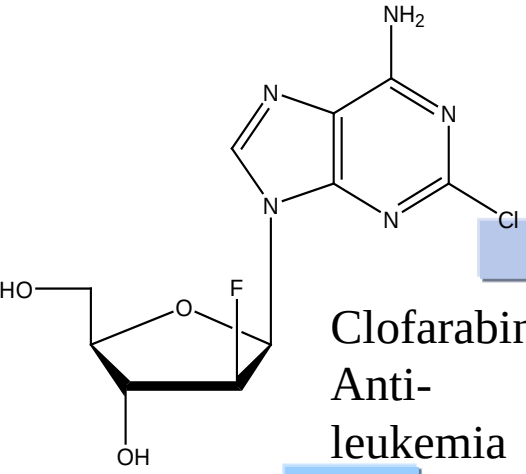
Basic concept of prediction by Pattern recognition

Principle of Information Equivalence



Basic concept of prediction by Pattern recognition

Principle of Information Equivalence



Clofarabin:
Anti-
leukemia

Black Box (Correlation)

Activity

Toxicity
Side-effects
ADME
Biodegrada-
bility

Information Equivalence
Inevitability

Initial Descriptors

MW
Atoms／Bonds
HOMO／LUMO
CPSS
Others

Feature Selection

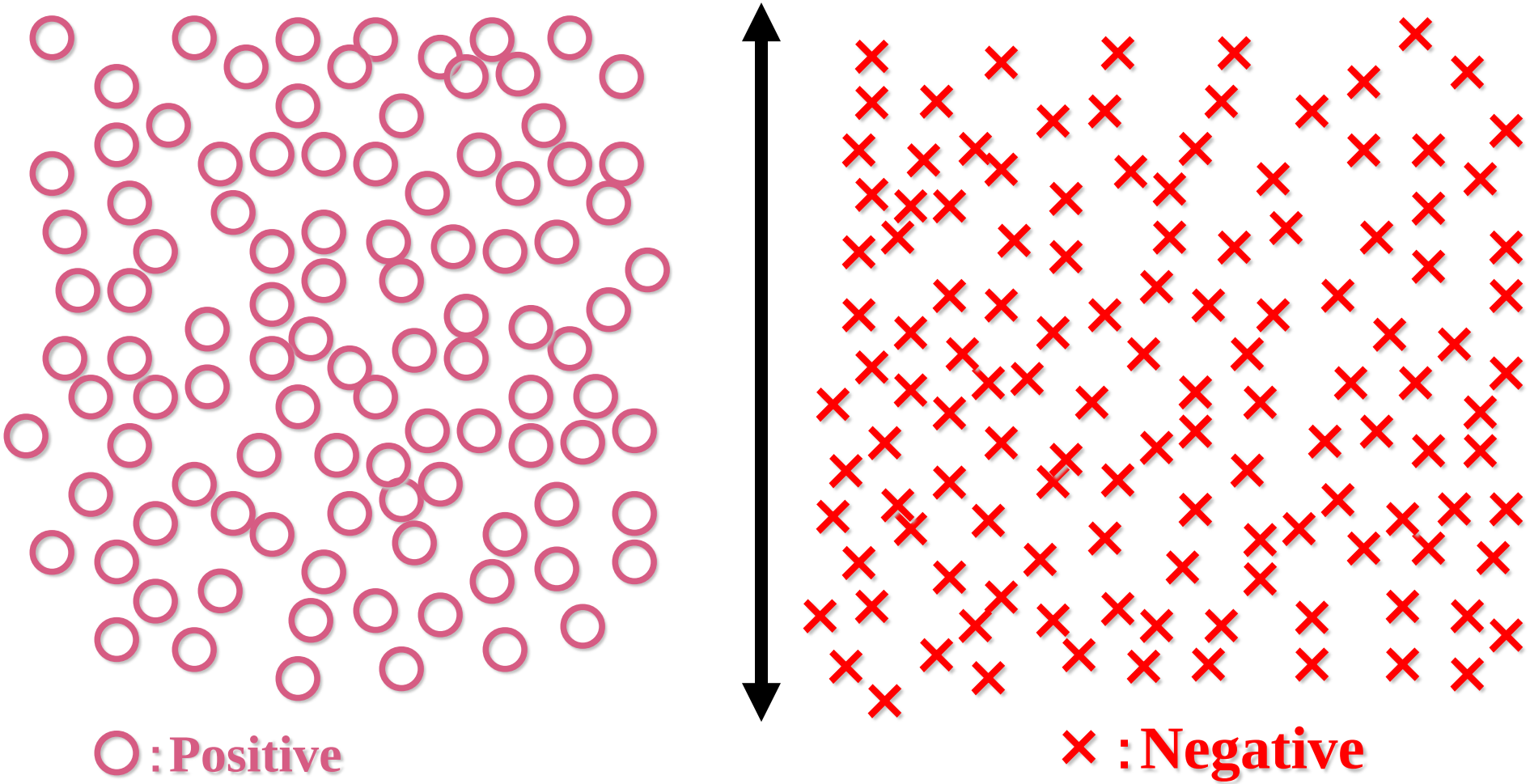
Final Descriptor Set

General features of data analysis by Pattern recognition techniques

Linear / Non-linear and DA / Fitting

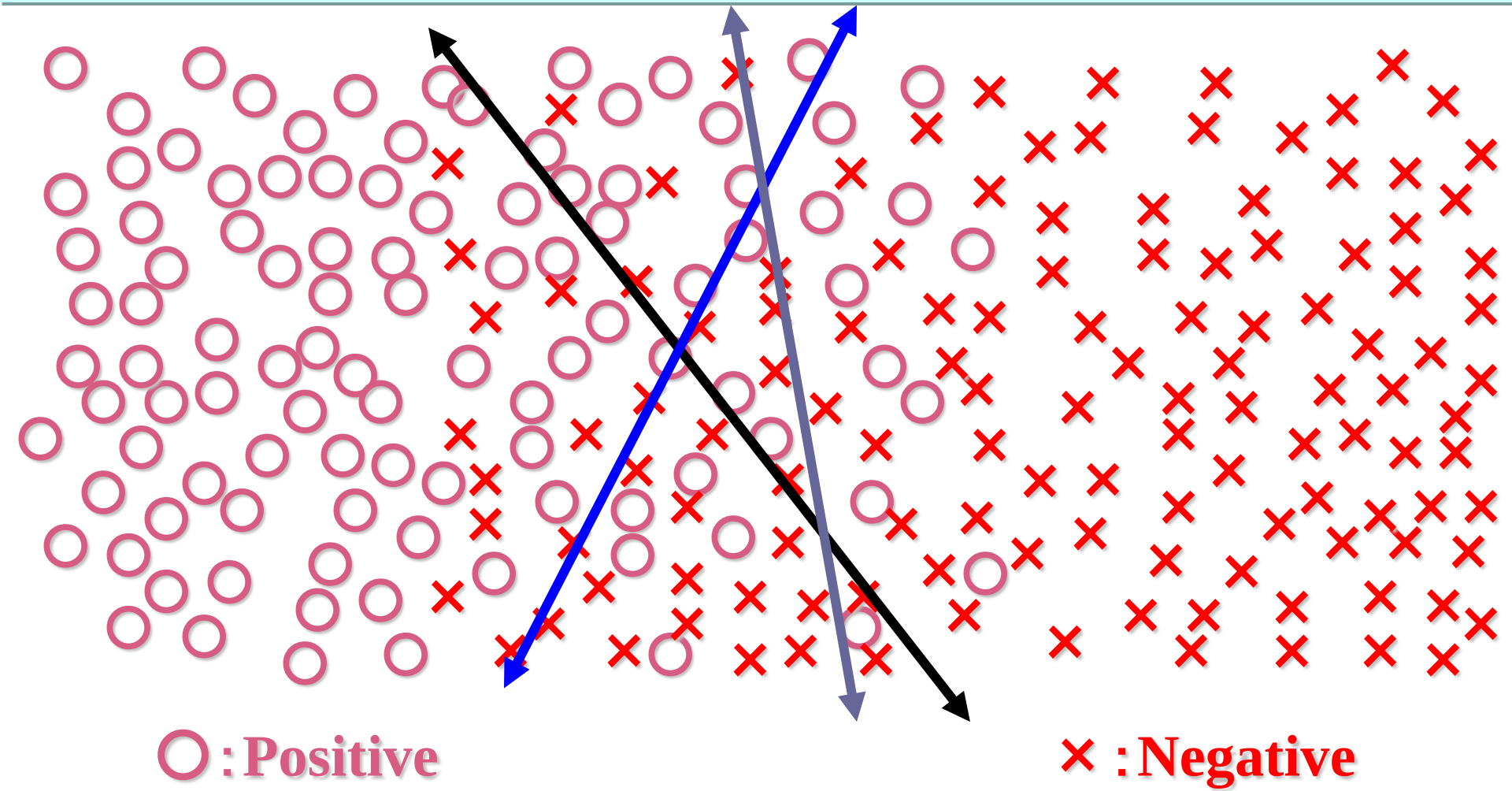
Sample space : two cluster samples

Discriminant function for perfect classification



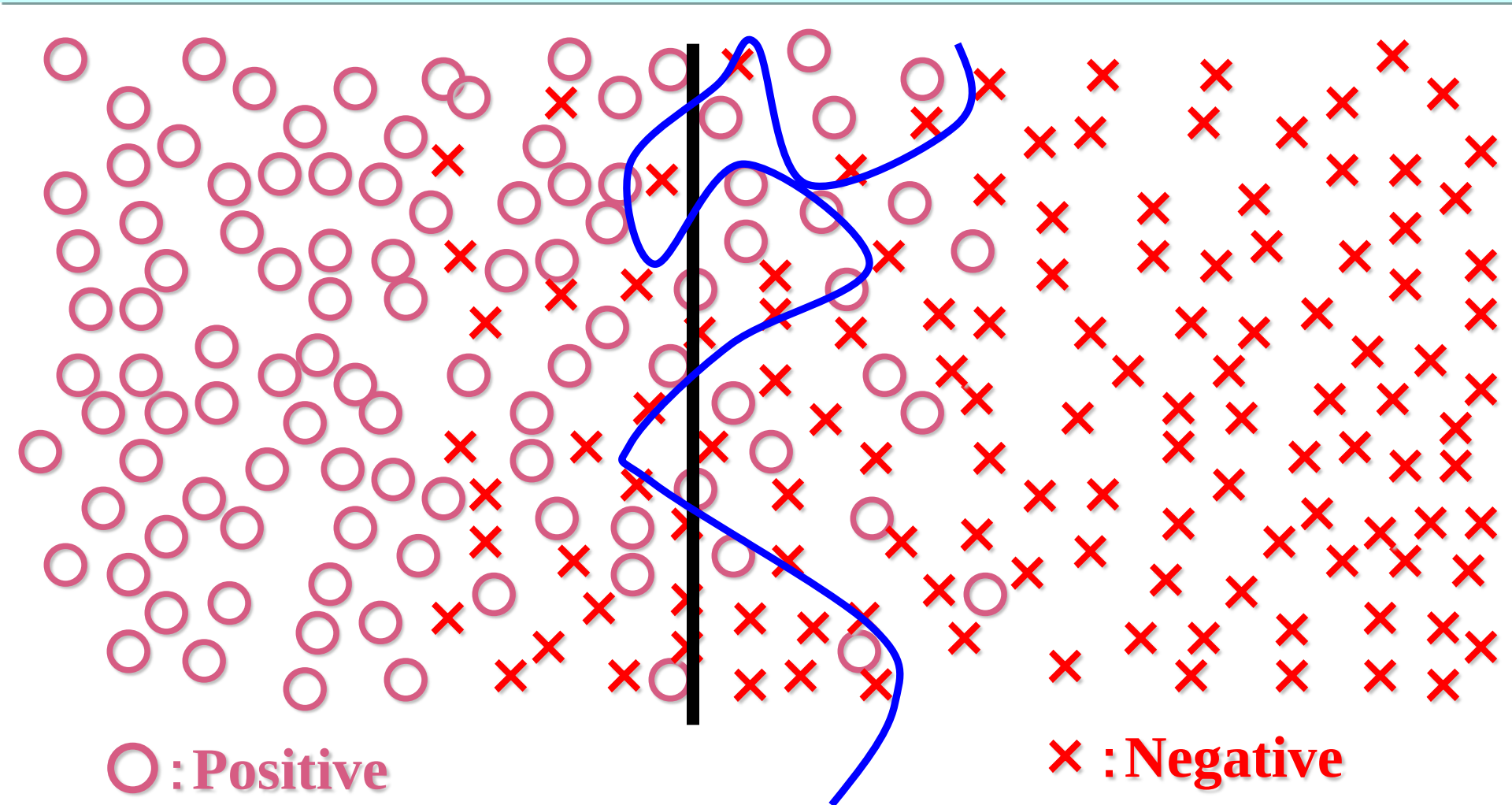
Sample space ：highly overlapped space

Discriminant function generated by various methods



Sample space ：highly overlapped space

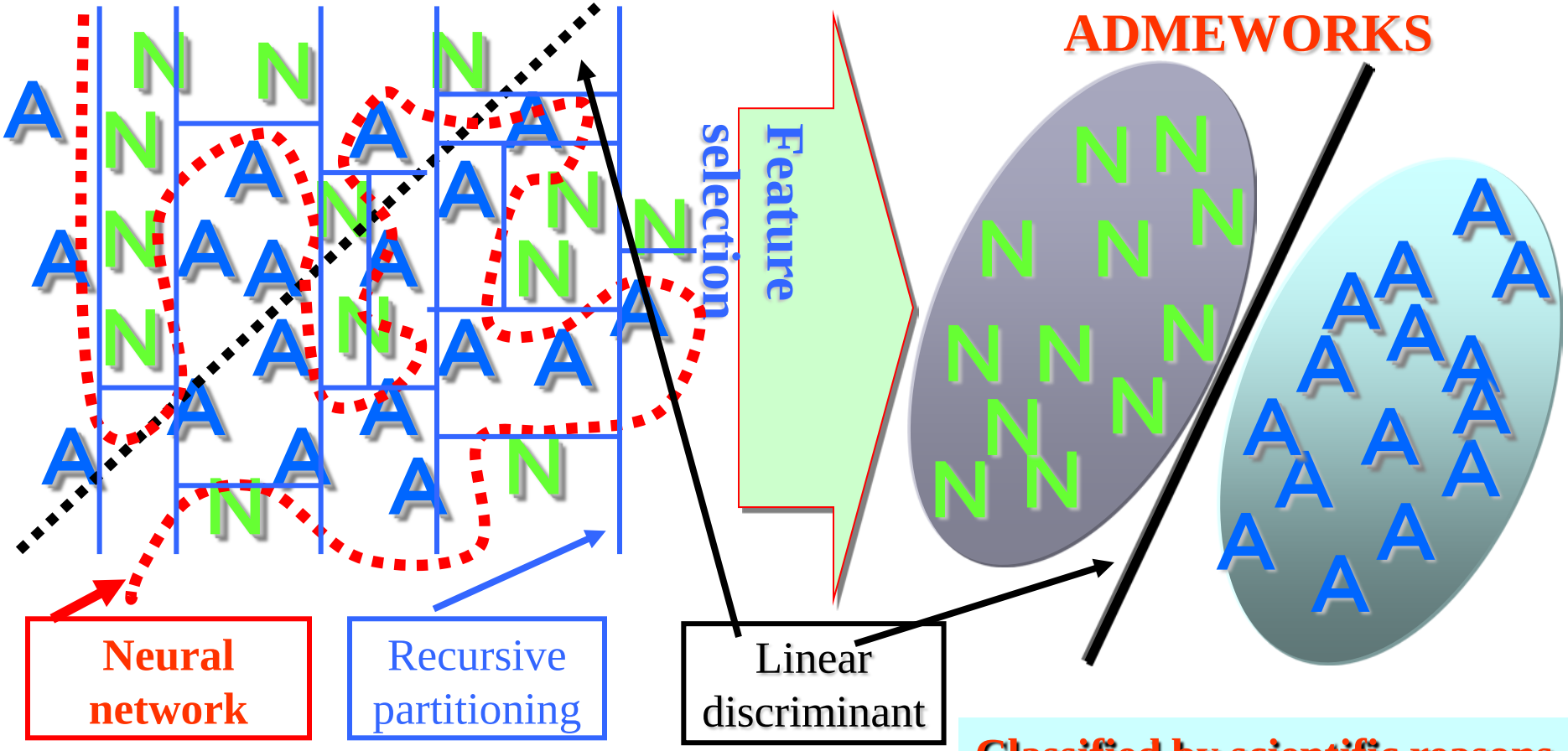
Discriminant function ： Linear and non-linear



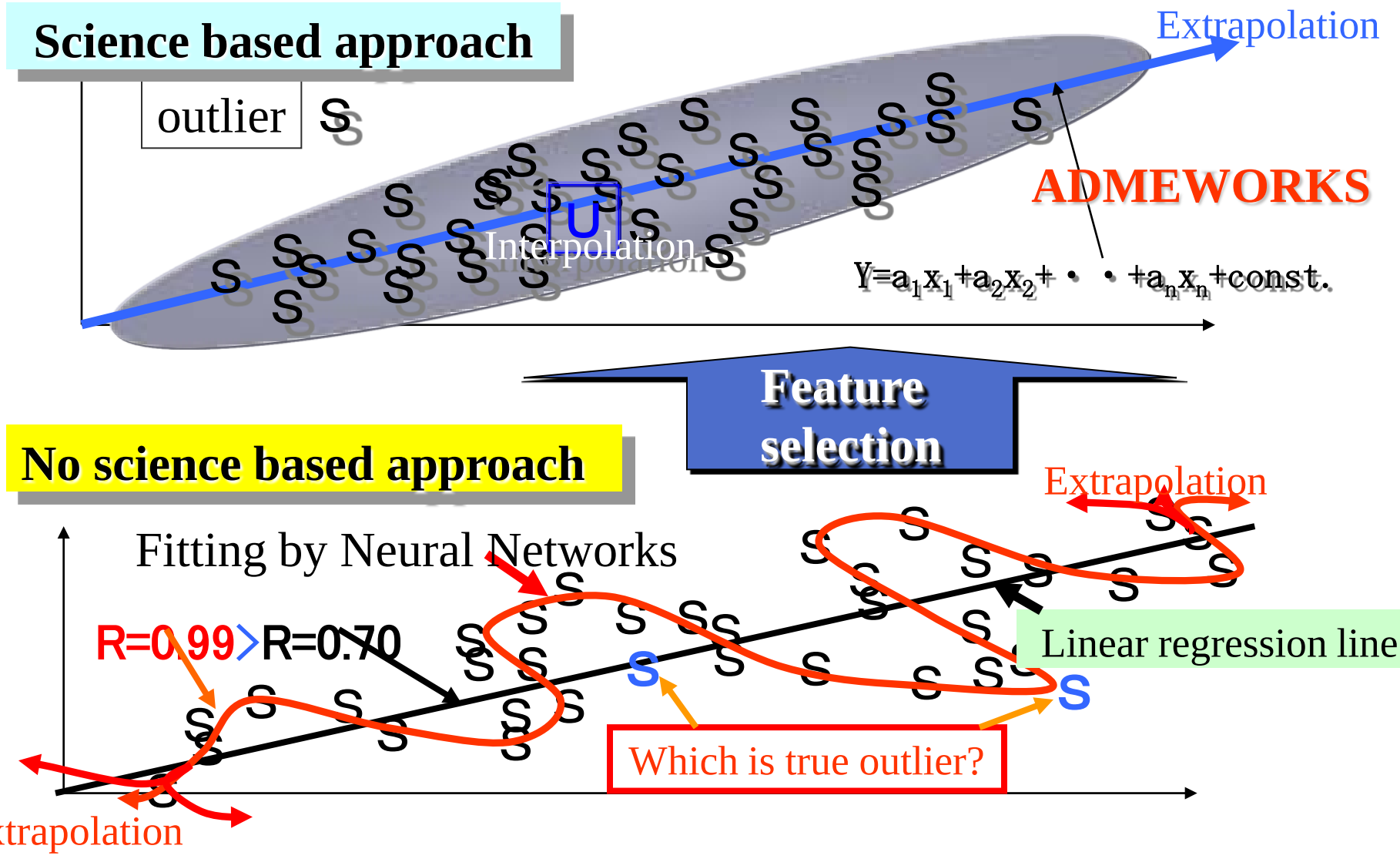
Simple classification and scientific classification

Pattern space impossible to be classified by linear discriminant

Pattern space classified by linear discriminant



Simple fitting and scientific fitting



☐Non-linear approach

Fit lines on existed sample space

No-remake sample space

Scientific approach

Remake sample space

Strong feature selection is required

Fit samples for individual end point

☐Linear approach

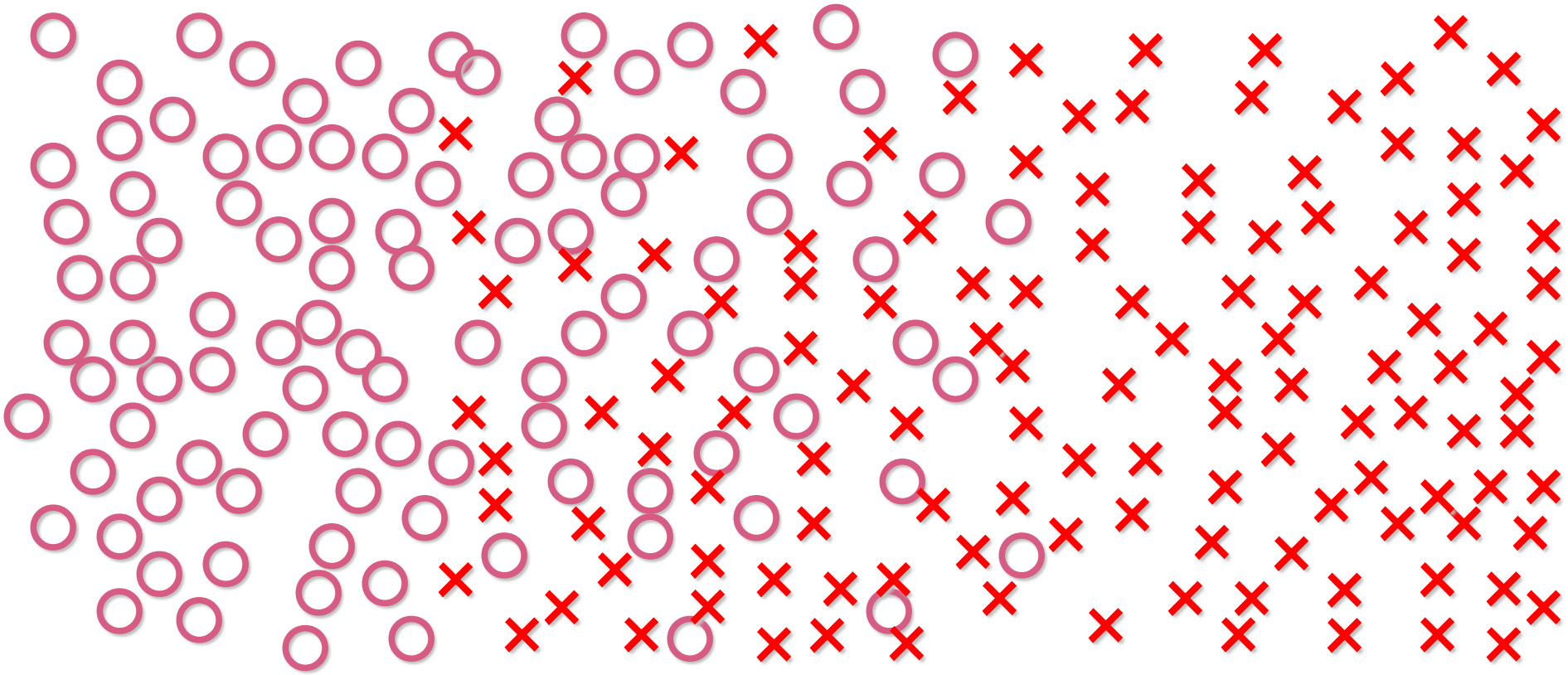
Building process to the features of "K-step Yard sampling method"

Step1: Yard sampling methods

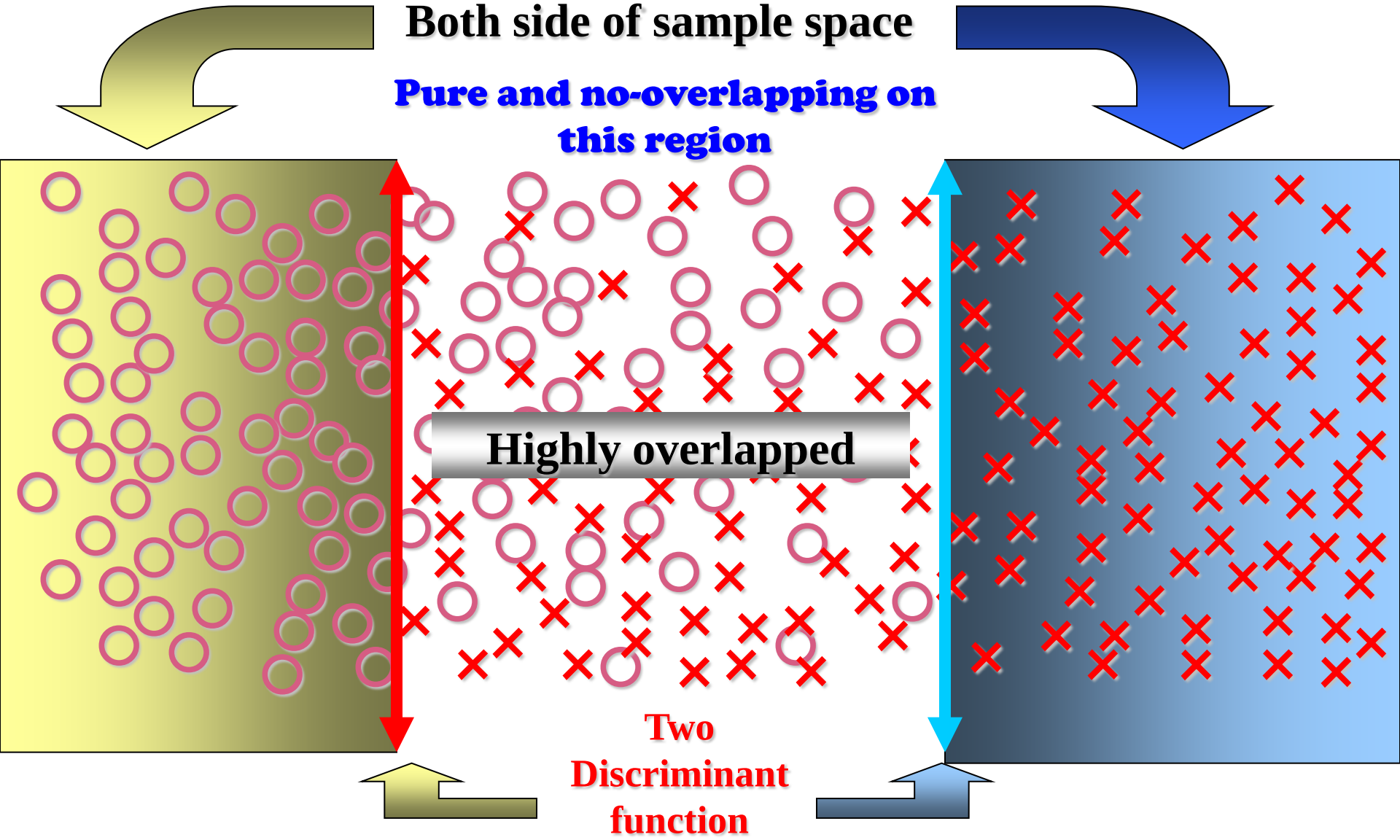
Spatial region on sample space

Both side of sample space

Pure and no-overlapping on this region

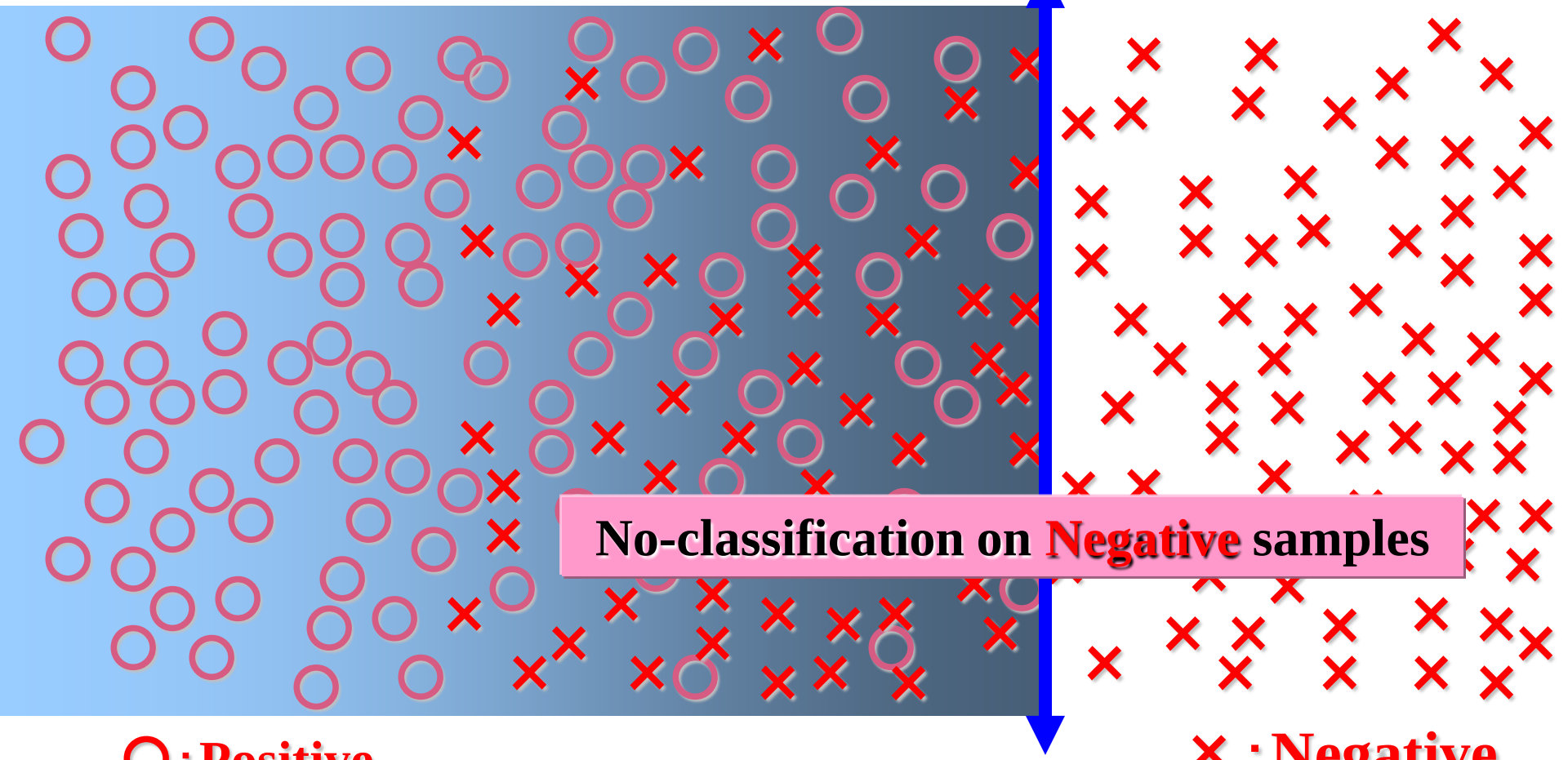


Spatial region on sample space

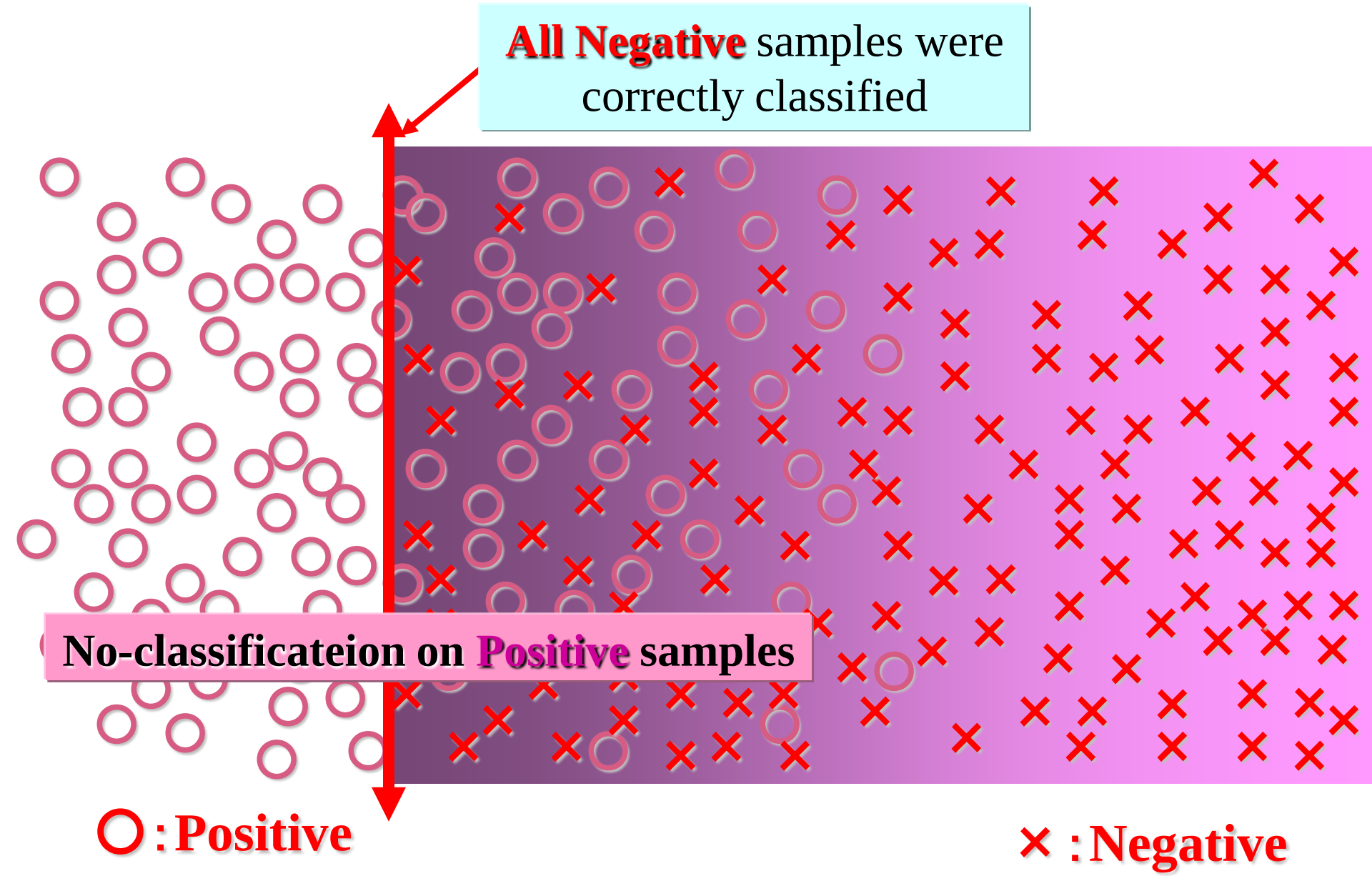


Property of AP（All Positive）model

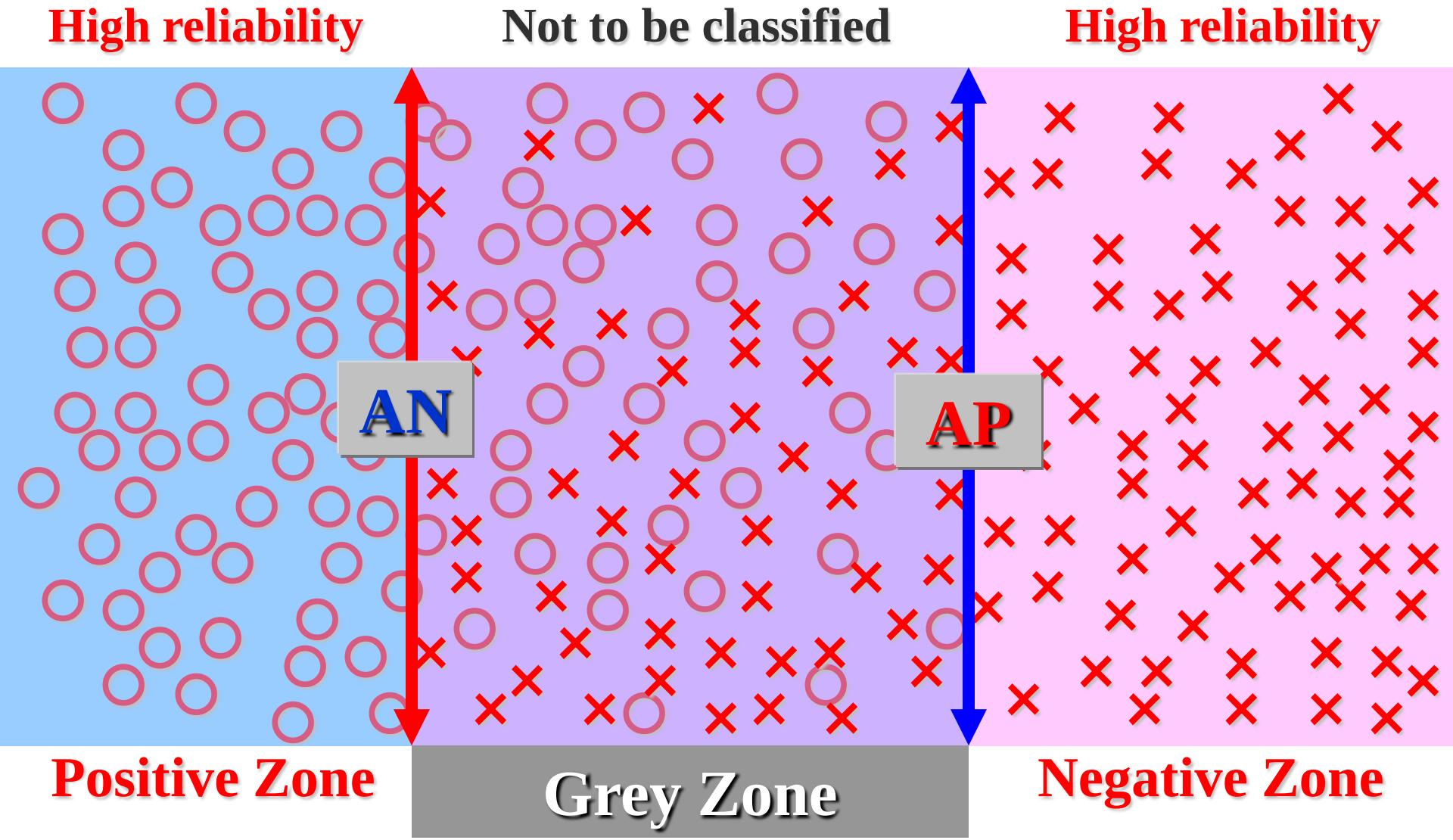
All Positive samples were correctly classified



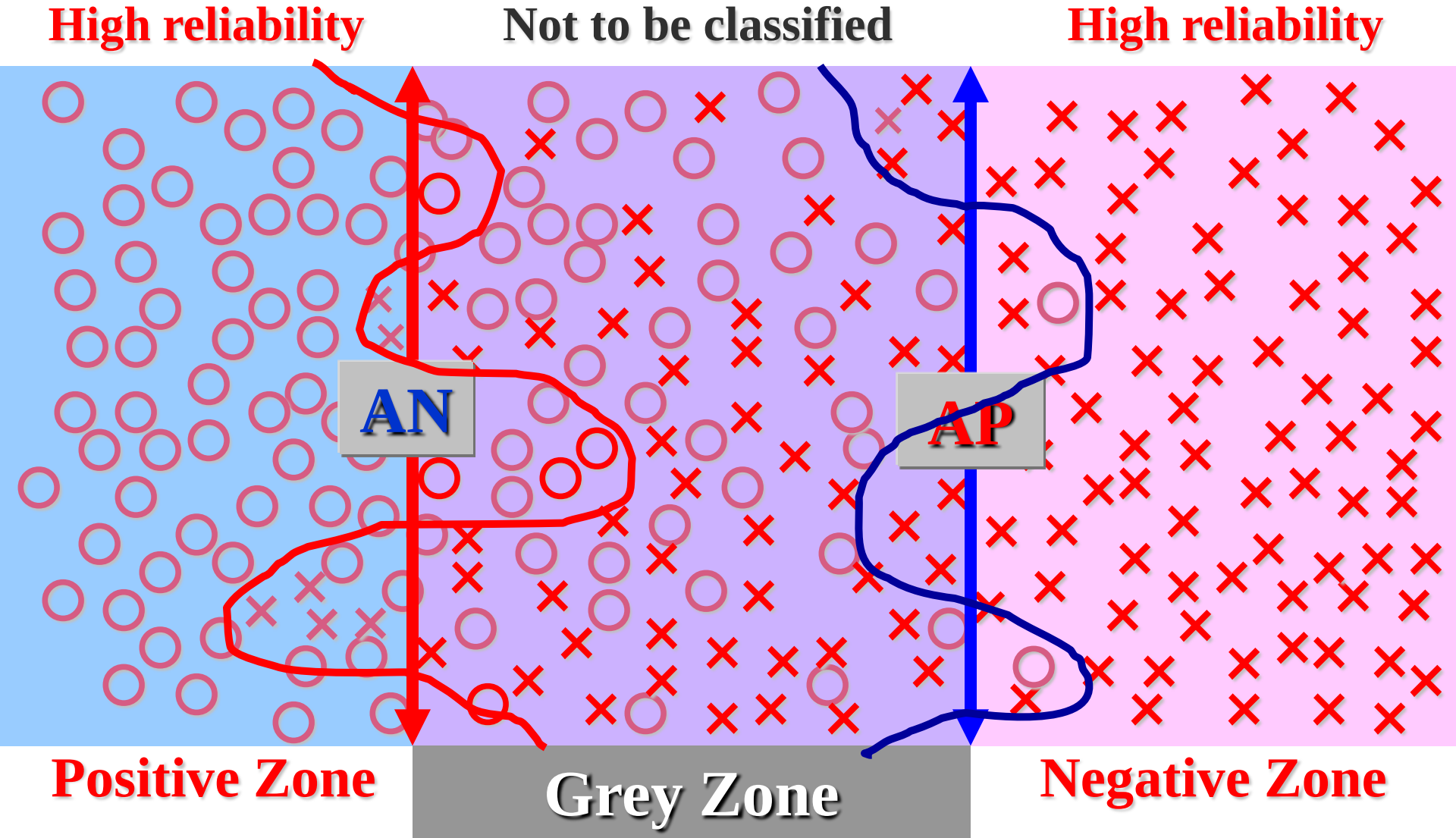
Property of AN（All Negative）model



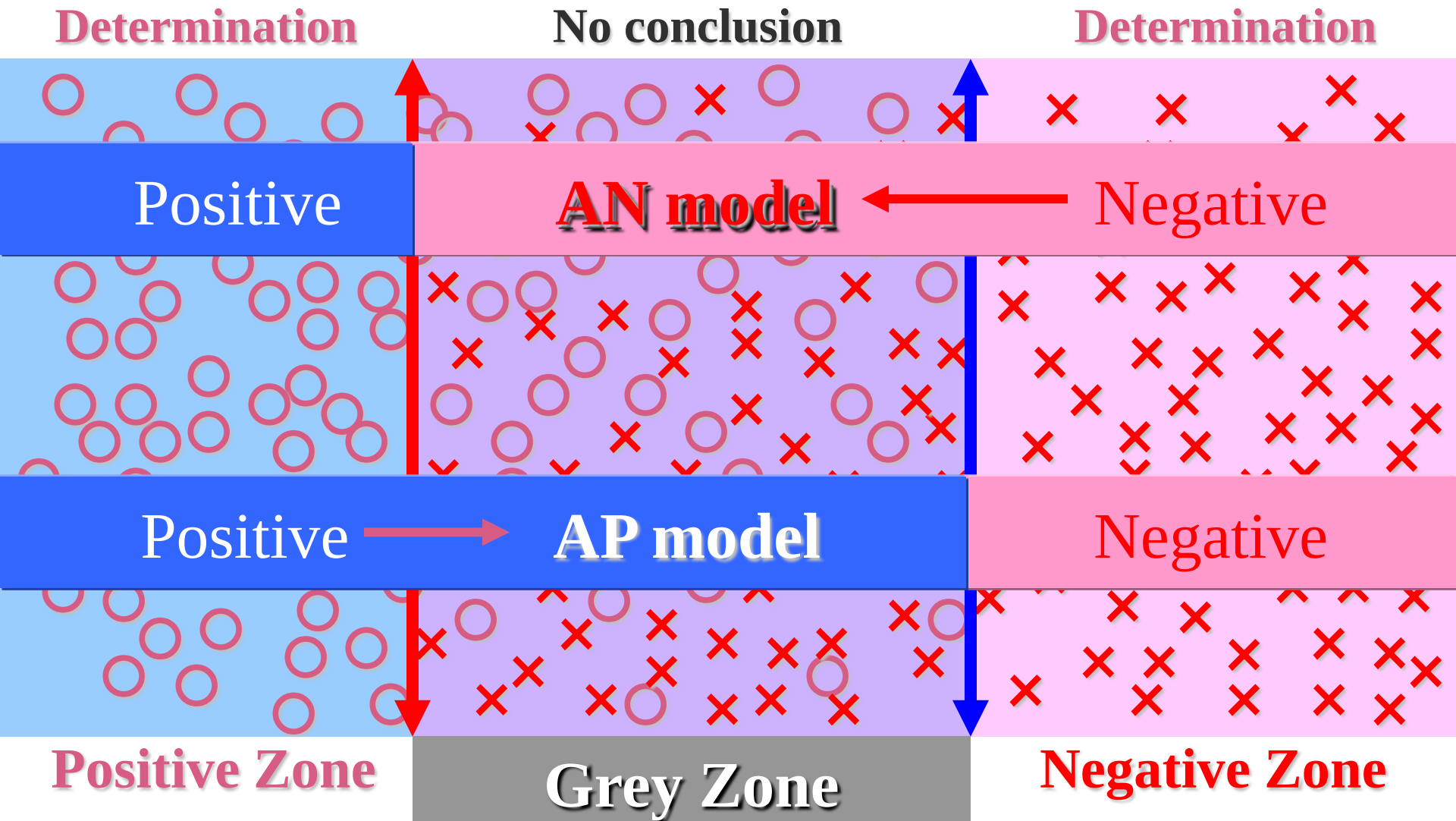
Combination of AN and AP models



Linear and non-linear discriminant on AP and AN models



Relations between Sample space & AN and AP models



Class determination by AN and AP models

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

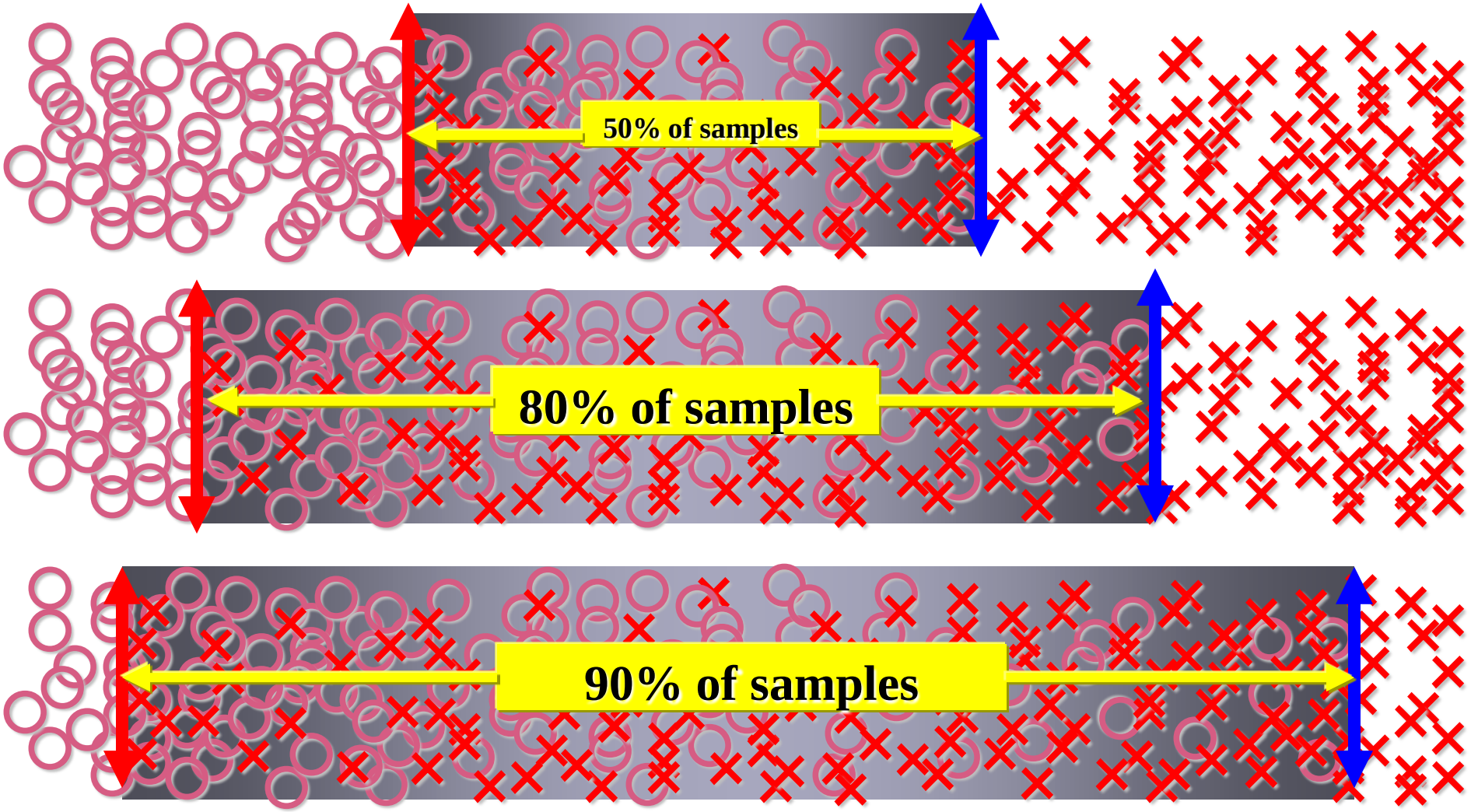
	AP model	AN model	Results
①	AP ; POSI,	AN ; POSI	→ POSI
②	AP ; POSI,	AN ; NEGA	→ GREY
③	AP ; NEGA,	AN ; POSI	→ GREY
④	AP ; NEGA,	AN ; NEGA	→ NEGA

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

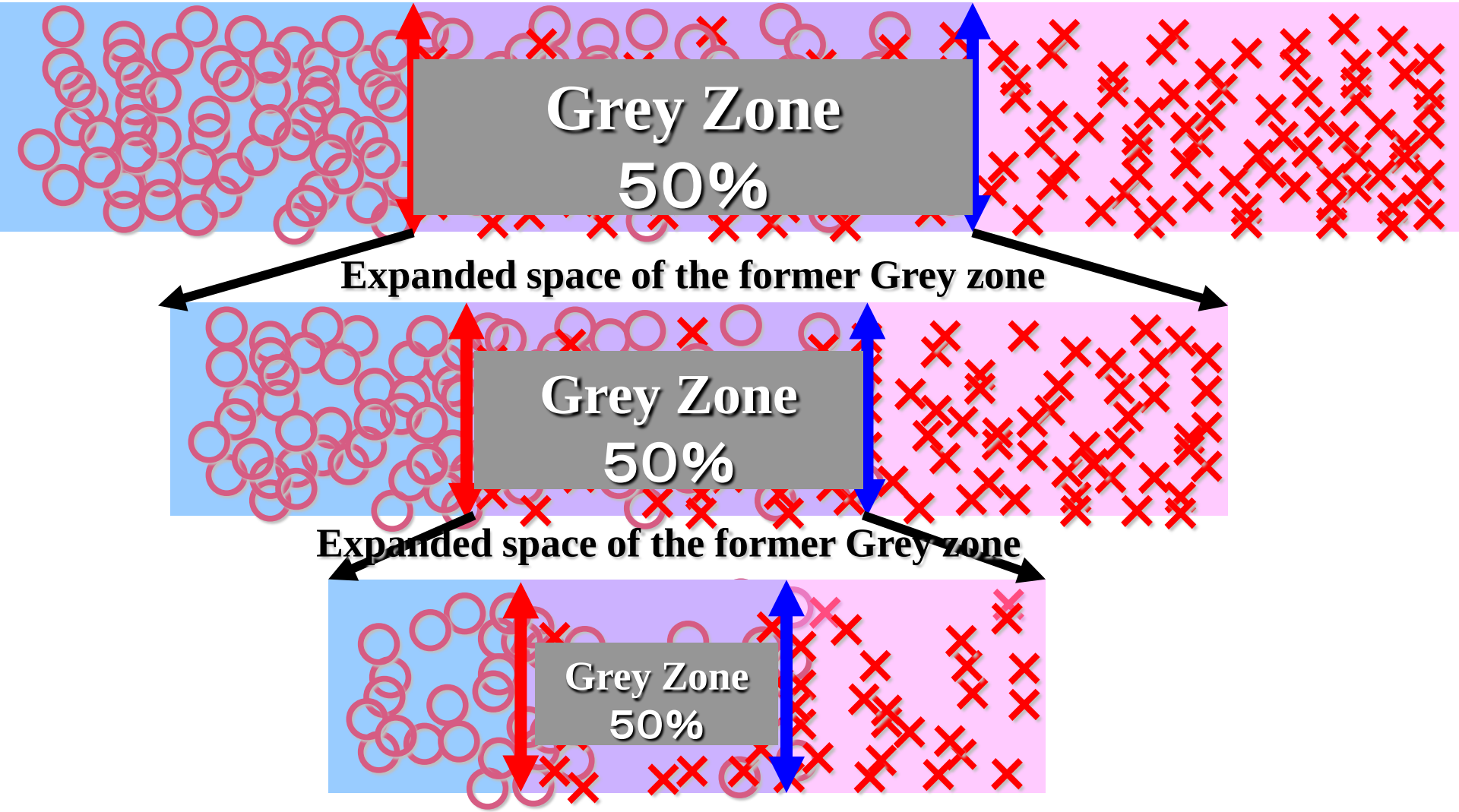
Step2: K-step approach

Problems of Yard sampling methods

The ratio of Grey zone:Highly overlapped sample space



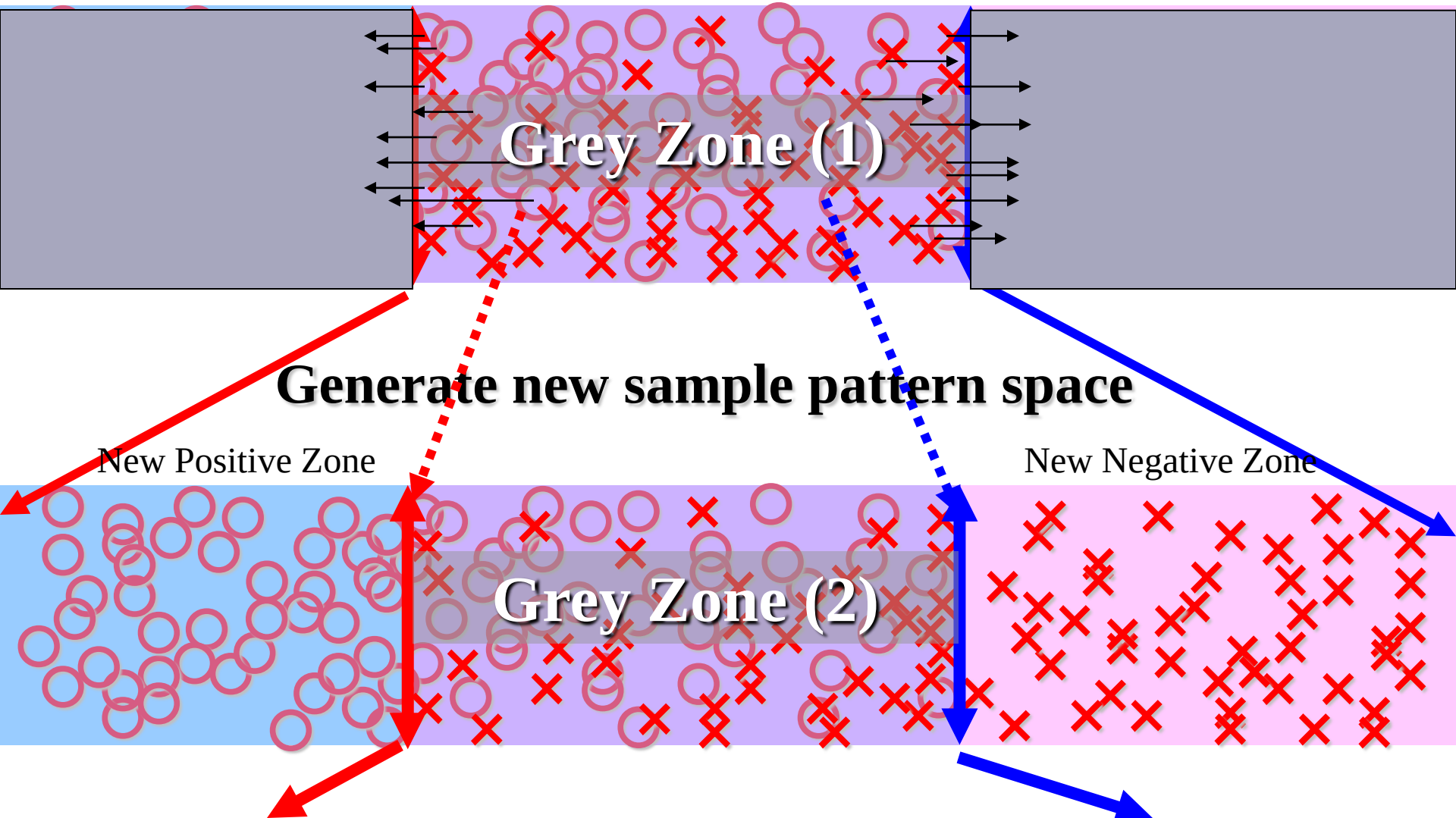
Steps to the K-step methods



In Silico Data
Miracles by the KY-methods

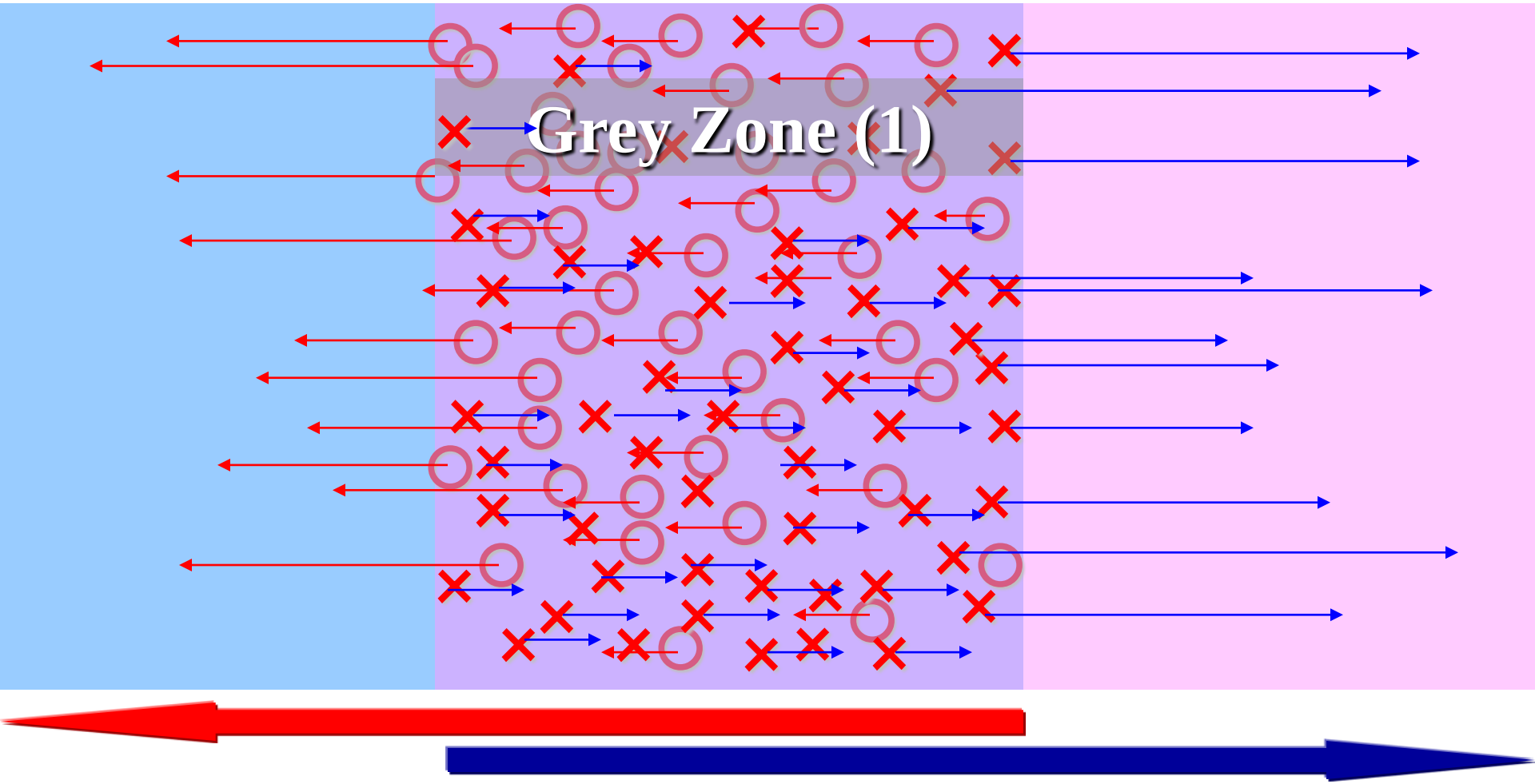
"K-step Yard sampling（KY）Method"

Improvement by repeated classification of Grey Zone samples



“K-step Yard sampling（KY）Method”

☐ **Relocation of Grey Zone samples on new sample space**

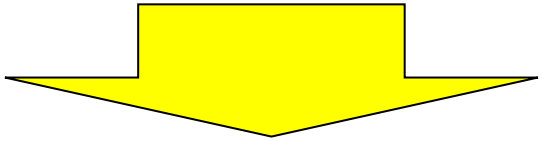


Building steps to the features of "K-step Yard sampling method"

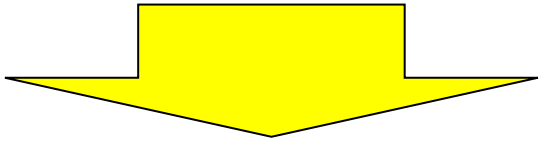
Step3: Merge two approaches: Yard sampling and K-step handling

The way to perfect classification

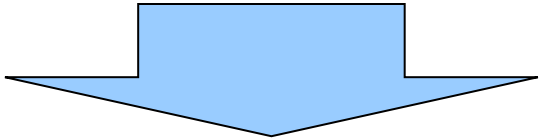
Partially realized by Yard sampling process



Not determined class on Grey zone compounds



Fixed up by K-step approach



Perfect classification for all samples:
any case, any time, any condition, others

"K-step Yard sampling" method

Yard sampling
process

For perfect classification

K-step
repeated processes

For no Grey zone