电子科技大学
University of Electronic Science and Technology of China

Discussion about Data Stream

Data Mining Lab, Big Data Research Center, UESTC
Email：junmshao@uestc.edu.cn
http://staff.uestc.edu.cn/shaojunming

Outline

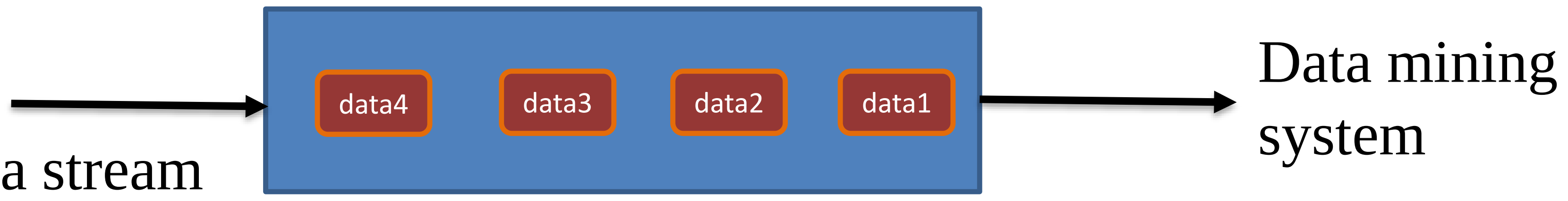
1. Background
2. Discuss, discuss and discuss

妈！他懒得好有个性

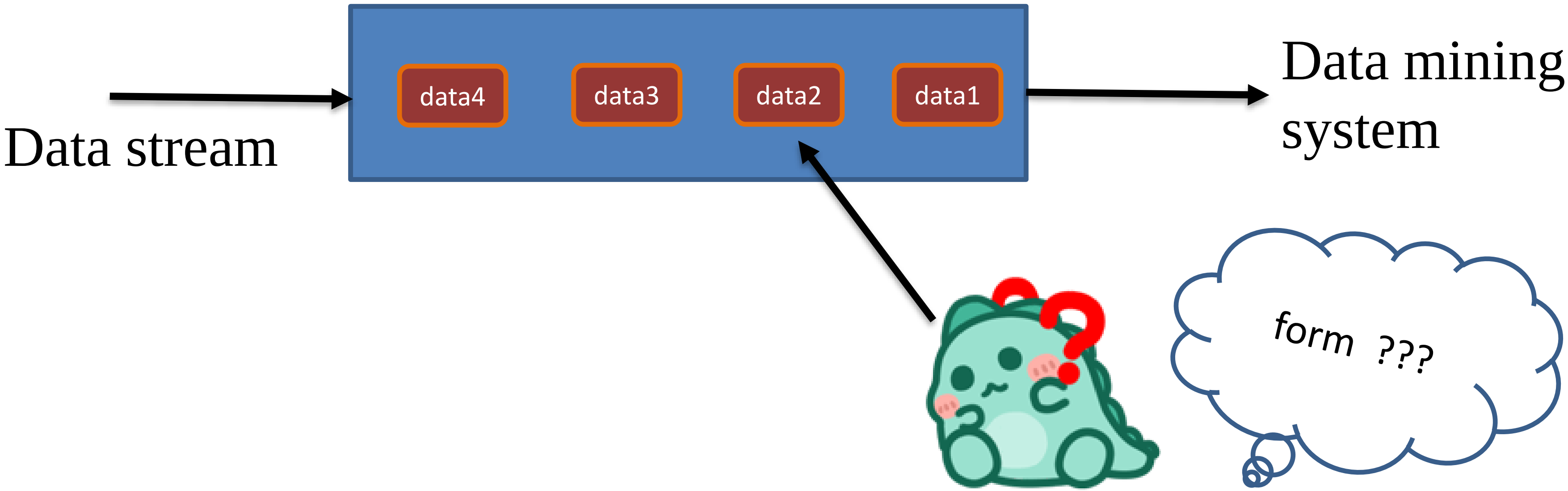
What is data stream

A data stream is a massive sequence of data objects which have some unique features:

- **One by one**
- **Potentially Unbounded**
- **Concept Drifting**



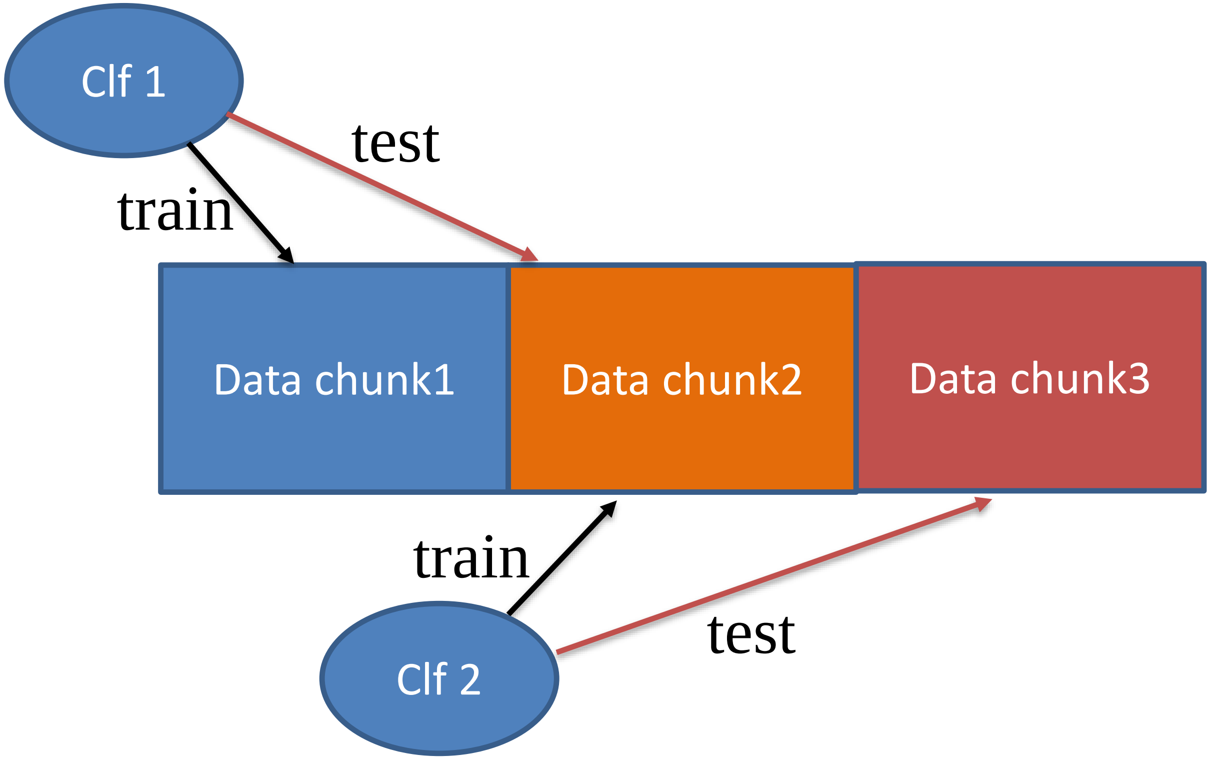
What is data stream



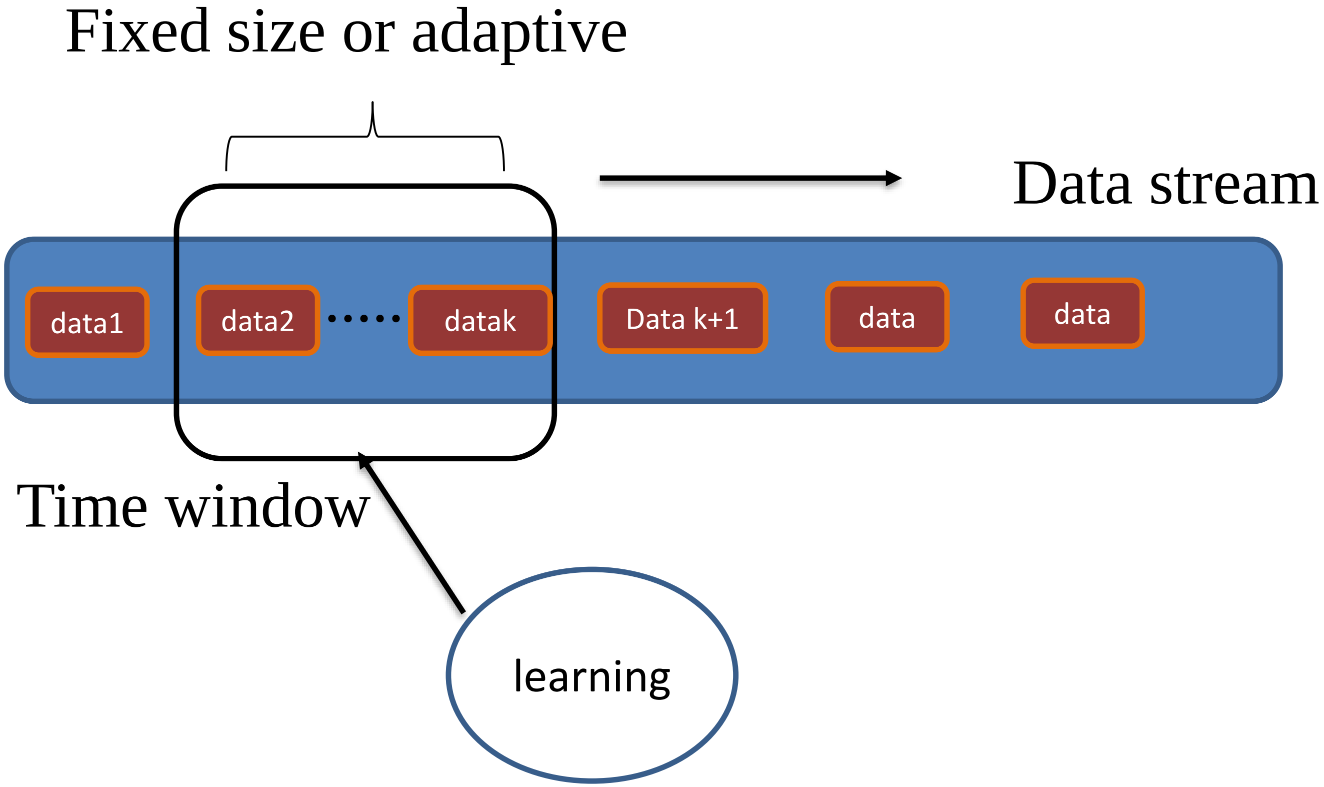
- Classification/regression (X, y)
- Clustering X
- Frequent rule (a,c,j)
- …….

How to do data stream mining

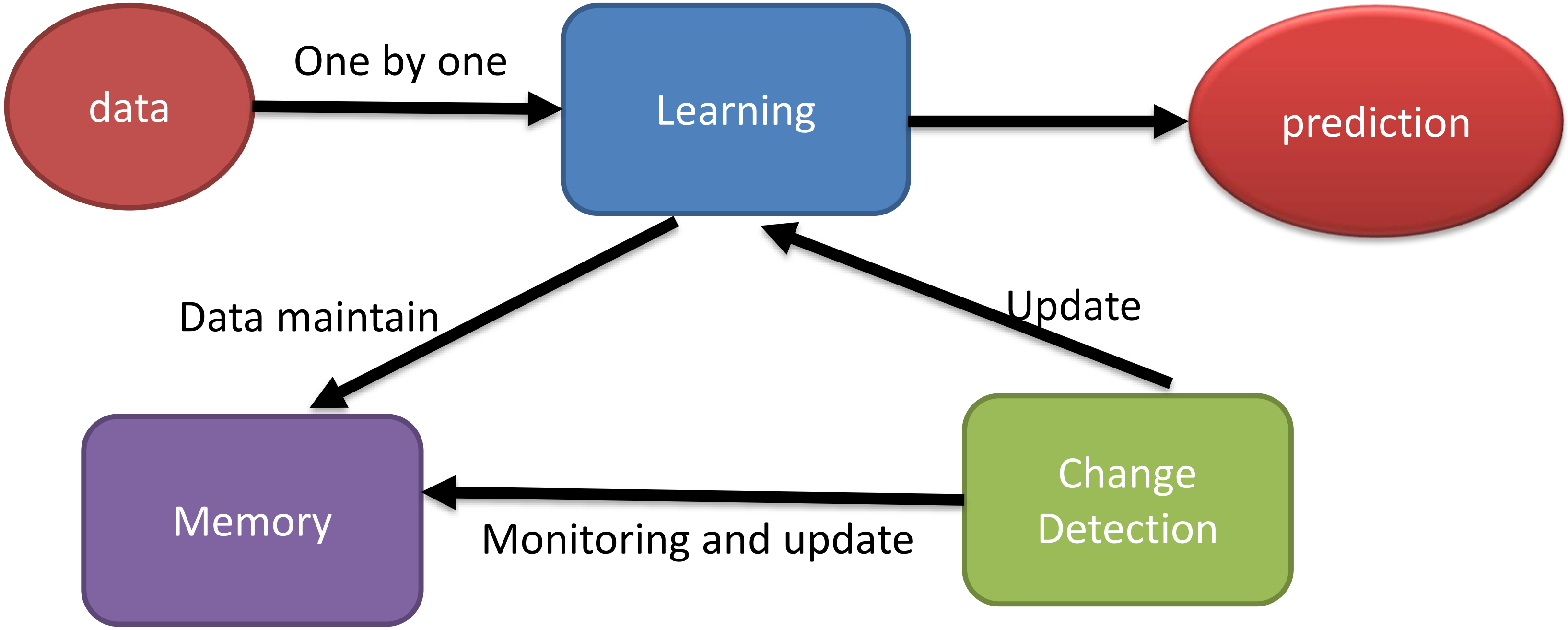
If we just consider the first two features of data stream， do you have some ideas?



Time sliding window



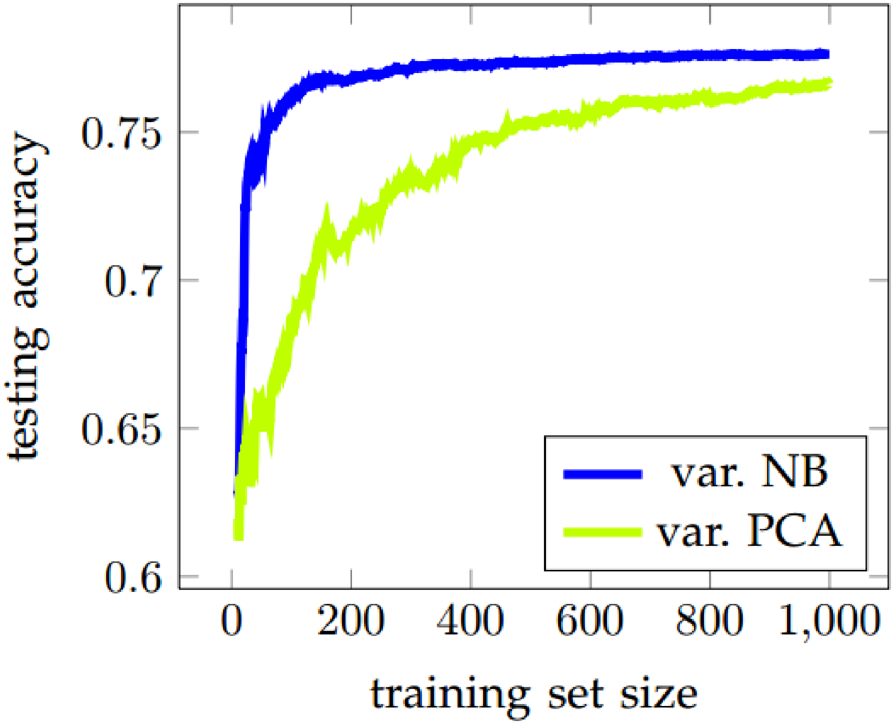
Data stream Framework



Streamed Preprocessing

Types of pre-processing

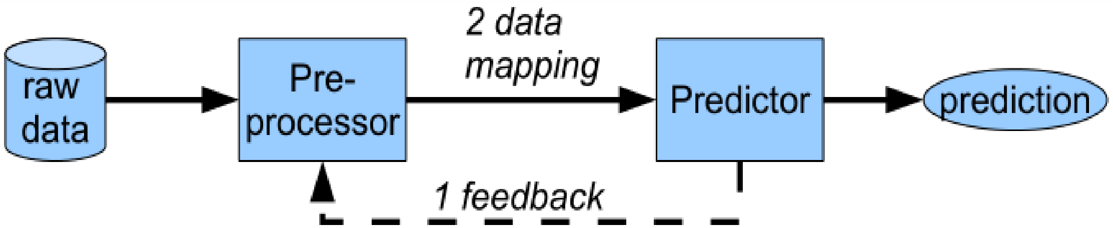
- Normalization…
- Feature selection
- Feature transformation



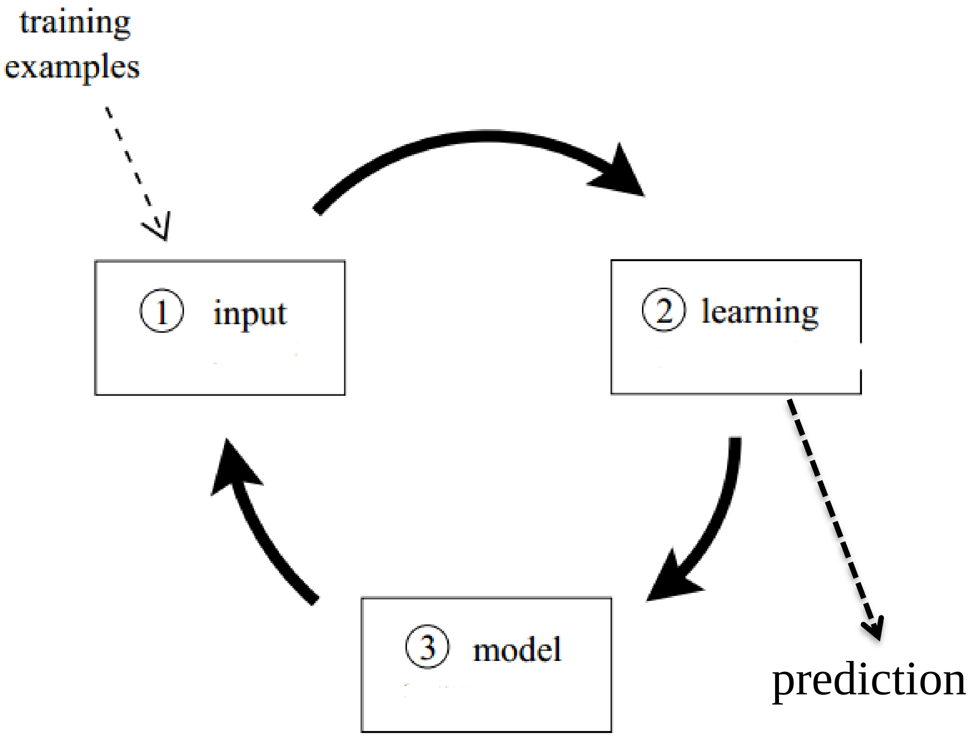
Problem 1

Streamed Preprocessing

How to do combing predictor?



Data stream classification



1. One by one

2. The algorithm predict the example, and then update its model based on X and Label Y.

3. The algorithm is ready to accept the next example

Problem 2

Handling Delayed Information

- How to learn from unlabeled data?
- How to preserve those unlabeled data?
- When delayed label information is available, how should you judge whether it is outdated?

Problem3

Evaluation on data stream mining

Data stream classification?
Data stream clustering?
Data stream outlier detection?

Conclusion

懒出啦新高度…

Thanks