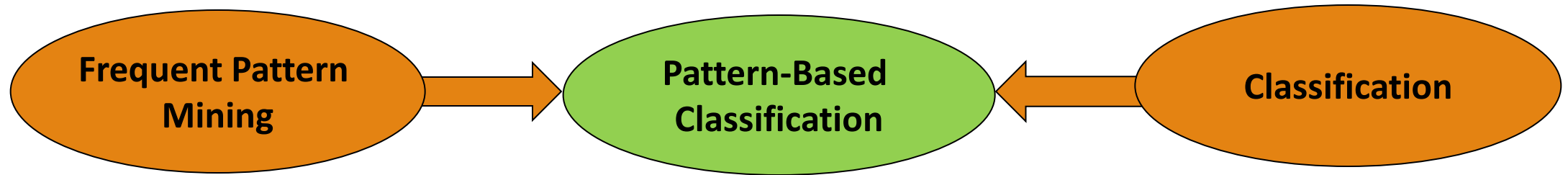


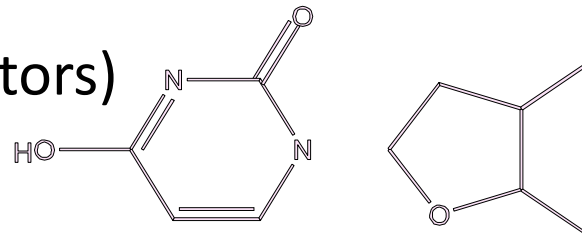
The background of the slide is a complex, abstract composition. It features a dark, muted purple or brownish background. Overlaid on this are several geometric and data-like elements: a network of thin, light-colored lines forming a mesh or web-like structure; numerous small, green and blue dots scattered across the field; and a prominent, lighter-colored, semi-transparent geometric shape (a large triangle or polygon) that serves as a backdrop for the title. In the bottom-left corner, there is a small, square inset image showing a cluster of orange and red dots, possibly representing a specific data set or a visualization of a pattern.

Session 2. Pattern-Based Classification

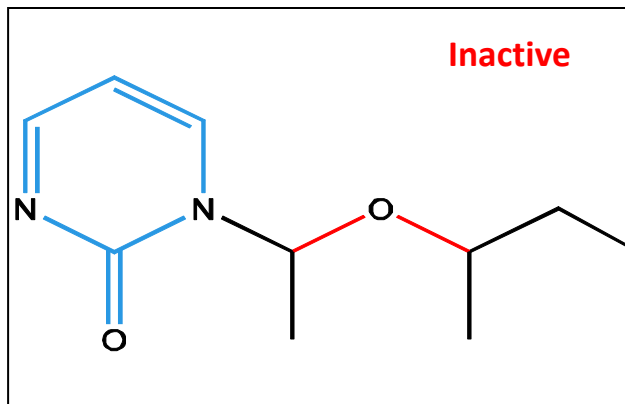
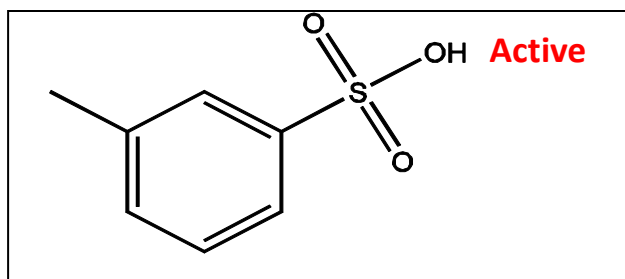
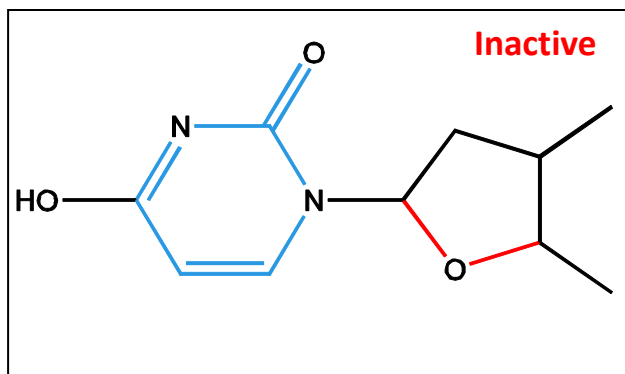
Pattern-Based Classification, Why?



- ❑ **Pattern-based classification:** An integration of both themes
- ❑ **Why pattern-based classification?**
 - ❑ **Feature construction**
 - ❑ Higher order; compact; discriminative
 - ❑ E.g., single word → phrase (Apple pie, Apple i-pad)
 - ❑ **Complex data modeling**
 - ❑ Graphs (no predefined feature vectors)
 - ❑ Sequences
 - ❑ Semi-structured/unstructured Data

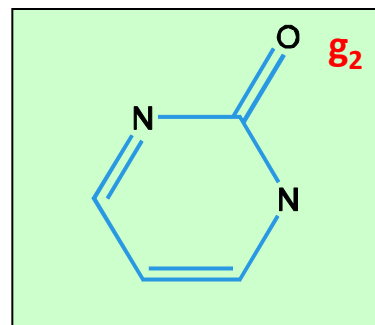
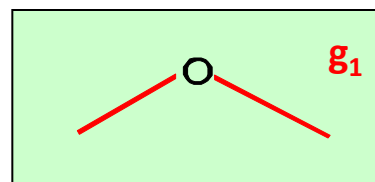


Pattern-Based Classification on Graphs



Mining
min_sup=2

Frequent subgraphs



Transform

Use frequent patterns as features for classification

g_1	g_2	Class
1	1	0
0	0	1
1	1	0

Associative or Pattern-Based Classification

- ❑ **Data:** Transactions, microarray data, ... → **Patterns or association rules**
- ❑ **Classification Methods** (Some interesting work):
 - ❑ CBA [Liu, Hsu & Ma, KDD'98]: Use high-conf., high-support *class association rules* to build classifiers To be discussed here
 - ❑ Emerging patterns [Dong & Li, KDD'99]: Patterns whose support changes significantly between the two classes
 - ❑ CMAR [Li, Han & Pei, ICDM'01]: Multiple rules in prediction To be discussed here
 - ❑ CPAR [Yin & Han, SDM'03]: Beam search on multiple prediction rules
 - ❑ RCBT [Cong et al., SIGMOD'05]: Build classifier based on mining top-k covering rule groups with row enumeration (for high-dimensional data)
 - ❑ Lazy classifier [Veloso, Meira & Zaki, ICDM'06]: For a test t , project training data D on t , mine rules from D_t , predict on the best rule
 - ❑ Discriminative pattern-based classification [Cheng et al., ICDE'07] To be discussed here