

1. 개요

WEKA의 K-Nearest Neighbor method를 이용한 숫자 인식 성능 비교

A. 목적

- i. K 값의 변화에 따른 성능 변화 비교
 - 가) 테스트 데이터의 소속 클래스를 결정할 기본 클러스터의 크기 변화에 따른 성능 변화 비교
- ii. Feature의 sub-sampling 후 성능 차이 비교
 - 가) Feature를 구성하는 $8 \times 8 = 64$ 개의 비트 맵 중 일부 비트만 샘플링 한 후 성능 변화 비교
- iii. 보고서 제출: 하드카피(분량제한 최대 10페이지)
- iv. 제출 마감: 12월 14일 목요일 (301동 419호 조교에게 제출)

2. WEKA

A. <http://www.cs.waikato.ac.nz/ml/weka/>

B. Download

- i. http://www.cs.waikato.ac.nz/ml/weka/index_downloading.html

3. 제공 데이터

A. $8 \times 8 = 64$ 개의 비트맵으로 표현된 숫자 데이터

1	2	3	4	5	6	7	8
9	10	11	12	13	14	15	16
17	18	19					
							64

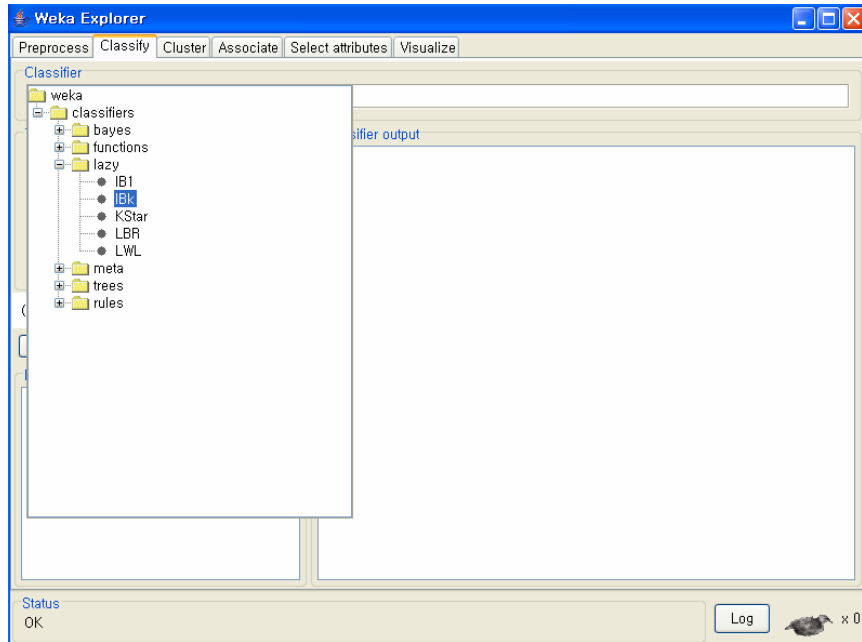
위와 같은 식으로 이미지 비트맵이 표현된 데이터입니다.

B. WEKA가 요구하는 arff 포맷으로 정리된 데이터 제공

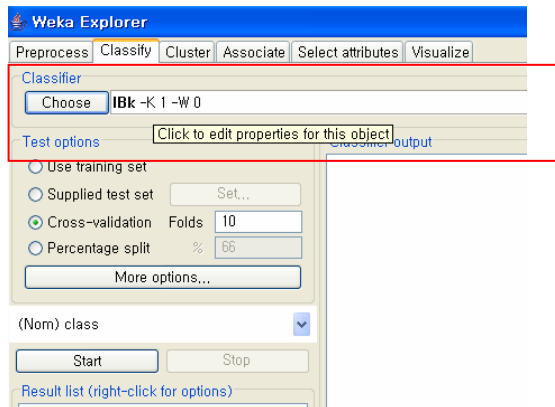
C. data1.arff와 data2.arff 두 개의 데이터 집합 중 선택하여 사용 가능

4. WEKA에서 KNN 사용법

A. "Classify" 탭에서 "lazy" 밑의 "IBk" 선택



B. K 값 변화시키기: 적색 사각형 내부의 “IBk -k 1 -w 0”이라는 텍스트를 마우스로 클릭



C. “KNN” 필드의 값을 변화시키면서 성능 변화 관찰

