

13. Rule Based Classification

20170718

박지혜

어서와~ 머신러닝은 처음이지?

6	국어	주말	점심	B	YES
7	과학	평일	저녁	C	NO
8	국어	주말	점심	A	YES
9	국어	주말	저녁	B	YES
10	수학	평일	저녁	B	NO
11	과학	주말	점심	C	NO
12	국어	주말	아침	A	YES
13	수학	주말	낮	C	NO
14	과학	평일	저녁	B	YES
15	수학	주말	낮	A	NO
16	수학	주말	아침	A	YES
17	영어	주말	아침	A	YES
18	수학	주말	저녁	A	YES
19	영어	주말	낮	C	NO
20	영어	평일	저녁	A	YES
21	영어	평일	저녁	C	NO

```
> test <- read.csv("rule.csv" , header = T)
```

```
> test
```

	과목	수업일	수업시간대	class	효과
1	수학	평일	저녁	C	NO
2	과학	주말	아침	A	YES
3	국어	평일	저녁	A	YES
4	수학	평일	저녁	A	YES
5	국어	주말	점심	B	NO
6	과학	평일	저녁	C	NO
7	국어	주말	점심	A	YES
8	국어	주말	저녁	B	YES
9	수학	평일	저녁	B	NO
10	과학	주말	점심	C	NO

Rule Based Classification

- 규칙기반분류
- 어떤 규칙이 가장 효율적인지
- 수치로 나타낼 필요가 있다!!

분류규칙 표시

• $R : (\text{condition}) \rightarrow (\text{result})$

or

$C \rightarrow R$

1. RULE1 : (과목=수학) \wedge (날짜=주말) \rightarrow 효과없음(NO)
2. RULE2 : (과목=과학) \rightarrow 효과없음(NO)
3. RULE3 : (과목=과학 or 수학) \wedge (시간=저녁) \wedge (class=A or B) \rightarrow 효과없음(NO)
4. RULE4 : (과목=영어 or 국어) \wedge (class=A or B) \rightarrow 효과있음(YES)

적용도 (coverage)

$C \rightarrow R$ 이라는 분류규칙이 있다면
전체데이터에서
 C 에 만족하는 데이터의 비율

$$\text{적용도(coverage)} = \frac{|C|}{n}$$

정확도(accuracy)

- 적용도를 만족하는 집합 중에서 R에 해당하는 집합

$$\text{정확도(accuracy)} = \frac{|\cap(C, R)|}{|C|}$$

가능도 비율(likelihood ratio)

- 규칙의 좋고 나쁨을 비교할 수 있는 지수
- 기대도수와 관찰도수의 비를 이용??
- 관찰도수(observed frequency) : 조건 C를 만족하는 실제 각 집단(R_i)을 만족하는 데이터 개수

