

COSE362 Machine Learning Assignment 2

181113

김병주

Data Mining & Information Systems Lab.
Department of Computer Science and Engineering,
College of Informatics, Korea University

Problem Definition



- Classification task
 - ✓ # of feature : 6
 - ✓ # of samples : train 2000, test 1500
 - ✓ # of class : 18
- 4가지 model(logistic regression, decision tree, random forest, support vector machine)을 train, validation 하여 model selection
 - ✓ (Optional) 다른 classification model 비교 가능(only in *scikit-learn*)
- Validation을 통해 model selection을 진행하고 적절한 model을 선택하여, test data에 대한 prediction result를 제출

Dataset Description



Training set

	A	B	C	D	E	F	G
1	feature1	feature2	feature3	feature4	feature5	feature6	target
2	a	1	c	5	d	2	3
3	b	2	a	8	e	2	15
4	b	1	c	3	h	7	6
5	c	1	a	5	h	7	13
6	d	3	h	2	g	4	15
7	b	1	g	3	g	8	6
8	d	1	b	2	d	8	6
9	b	1	d	6	g	6	11
10	b	2	h	8	d	1	12
11	a	1	c	1	e	3	3
12	c	2	a	8	g	1	1
13	d	3	f	2	c	5	15
14	b	1	d	7	c	4	15
15	c	2	g	3	f	6	13
16	c	3	c	1	h	3	2
17	c	1	a	1	a	5	13
18	b	1	b	6	e	6	3

Test set

	A	B	C	D	E	F
1	feature1	feature2	feature3	feature4	feature5	feature6
2	d	1	h	5	h	6
3	d	1	c	1	e	5
4	c	1	d	3	e	5
5	d	2	a	3	g	8
6	d	1	e	1	b	3
7	b	1	c	3	h	8
8	a	1	c	3	e	1
9	b	1	a	6	g	8
10	d	3	d	6	f	3
11	c	1	d	6	h	3
12	b	1	c	5	c	8
13	a	1	c	4	f	5
14	c	3	b	8	d	1
15	d	1	c	8	e	6
16	c	2	g	5	h	7
17	b	1	d	4	d	7
18	d	3	h	3	a	1