

## Comparing Classifiers - Exercise

**Problem 1:** Given the training dataset presented on Table 1, classify the following record using k-Nearest Neighbors (K=1; Euclidian Distance) and Naïve Bayes classification.

outlook	temperature	humidity	windy	play
overcast 3	cool 3	normal 2	true 2	?

*Table 1: Training Data from "weather.nominal" data set*

	outlook	temperature	humidity	windy	play
1	sunny 1	hot 1	high 1	false 1	no 1
2	sunny 1	hot 1	high 1	true 2	no 1
3	rainy 2	mild 2	high 1	false 1	yes 2
4	rainy 2	cool 3	normal 2	true 2	no 1
5	sunny 1	cool 3	normal 2	false 1	yes 2
6	sunny 1	mild 2	normal 2	true 2	yes 2
7	overcast 3	mild 2	high 1	true 2	yes 2
8	overcast 3	hot 1	normal 2	false 1	yes 2
9	rainy 2	mild 2	high 1	true 2	no 1
10	rainy 2	cool 3	normal 2	false 1	yes 2
11	sunny 1	mild 2	high 1	false 1	no 1
12	overcast 3	hot 1	high 1	false 1	yes 2
13	rainy 2	mild 2	normal 2	false 1	yes 2

**Problem 2:** The "J48 pruned tree" for the training data set presented on Table 1 was obtained using the WEKA software. The decision tree is given below.

```

Outlook = sunny
|  humidity = high: no (3.0)
|  humidity = normal: yes (2.0)
outlook = overcast: yes (3.0)
outlook = rainy
|  windy = true: no (2.0)
|  windy = false: yes (3.0)
    
```

- How the test record from Problem 1 is classified using this decision tree?
- ✓ What are the classification rules that can be obtained from this tree?



### Exercise 7

#### ⊙ K-Nearest Neighbors

No.	E.D	Yes=2   No=1
1	$\sqrt{2^2+2^2+1^2+1^2} = \sqrt{10}$	1
2	$\sqrt{2^2+2^2+1^2+0} = 3$	1
3	$\sqrt{1^2+1^2+1^2+1^2} = 2$	2
4	$\sqrt{1^2+0+0+0} = 1$	1
5	$\sqrt{2^2+0+0+1^2} = \sqrt{5}$	2
6	$\sqrt{2^2+1^2+0+0} = \sqrt{5}$	2
7	$\sqrt{0+1^2+1^2+0} = \sqrt{2}$	2
8	$\sqrt{0+2^2+0+1^2} = \sqrt{5}$	2
9	$\sqrt{1^2+1^2+1^2+0} = \sqrt{3}$	1
10	$\sqrt{1^2+0+0+1^2} = \sqrt{2}$	2
11	$\sqrt{2^2+1^2+1^2+1^2} = \sqrt{7}$	1
12	$\sqrt{0+2^2+1^2+1^2} = \sqrt{6}$	2
13	$\sqrt{1^2+1^2+0+1^2} = \sqrt{3}$	2

Since  $k=1$ , choose minimum E.D.  $\Rightarrow$  therefore, "No"

\* 가장 가까운 이웃을 선택할 때, 가장 가까운 이웃만 선택하는 것이 아니라!!

TEST

d : 0.707106781

x



### ⊙ Naive Bayes

$$P(\text{yes}) = 8/13 = 0.615$$

$$P(\text{no}) = 5/13 = 0.385$$

$$1. P(\text{overcast} | \text{yes}) = 3/8 = 0.375$$

$$2. P(\text{overcast} | \text{no}) = 0$$

$$3. P(\text{cool} | \text{yes}) = 2/8 = 0.25$$

$$4. P(\text{cool} | \text{no}) = 1/5 = 0.2$$

$$5. P(\text{normal} | \text{yes}) = 5/8 = 0.625$$

$$6. P(\text{normal} | \text{no}) = 1/5 = 0.2$$

$$7. P(\text{true} | \text{yes}) = 2/8 = 0.25$$

$$8. P(\text{true} | \text{no}) = 3/5 = 0.6$$

$$\begin{aligned} \text{Therefore, } P(\text{yes} | \text{Given}) &= (0.375 \times 0.25 \times 0.625 \times 0.25) \times 0.615 \\ &= 0.0087 \end{aligned}$$

$$P(\text{no} | \text{Given}) = 0$$

⇒ "YES" (Play)

\* 큰 P를 갖는 것보다 그건이 아닌 것의 P를 고려!!