Classifiers: toy example of Naive Bayes

Data mining lab 3

Input data

	skin	color	size	flesh	class
1	hairy	brown	large	hard	safe
2	hairy	green	large	hard	safe
3	smooth	red	large	soft	dangerous
4	hairy	green	large	soft	safe
5	hairy	red	small	hard	safe
6	smooth	red	small	hard	safe
7	smooth	brown	small	hard	safe
8	hairy	green	small	soft	dangerous
9	smooth	green	small	hard	dangerous
10	hairy	red	large	hard	safe
11	smooth	brown	large	soft	safe
12	smooth	green	small	soft	dangerous
13	hairy	red	small	soft	safe
14	smooth	red	large	hard	dangerous
15	smooth	red	small	hard	safe
16	hairy	green	small	hard	dangerous

Query: hairy skin, red color, large with soft flesh.
The question is: is it safe to eat this unknown animal?

Naive Bayes I

1. To answer the query we need to calculate the relative values for probabilities: P(safe | E) and P(dangerous|E), where E=(hairy,red,large,soft).

Naive Bayes II

2. According to the formula, P(safe|E)=P(hairy|safe)*P(red|safe)*P(large|safe)*P(soft|safe)*P(safe)*alpha.

P(hairy|safe) is estimated as a fraction of hairy animals from all animals of class safe, similarly for the rest of the probabilities.

P(safe|E)=6/10*5/10*5/10*3/10*10/16*alpha=alpha*0.028

Naive Bayes III

3. Compute P(dangerous|E)=P(hairy|dangerous)*P(red|dangerous)*P(large|dangerous)*P(soft|dangerous)*P(dangerous)*alpha.

P(dangerous|E)=2/6*2/6*2/6*3/6*6/16*alpha=alpha*0.007

Naive Bayes IV

4. Compare P(safe|E) with P(dangerous|E). What is the prediction? Is it safe or dangerous to eat this animal?

P(safe|E)=alpha*0.028

P(dangerous|E)=alpha*0.007