Bayesian Classifiers

Prof. Dr. Ronny Hartanto

HOCHSCHULE RHEIN-WAAL
Rhine-Waal University of Applied Sciences

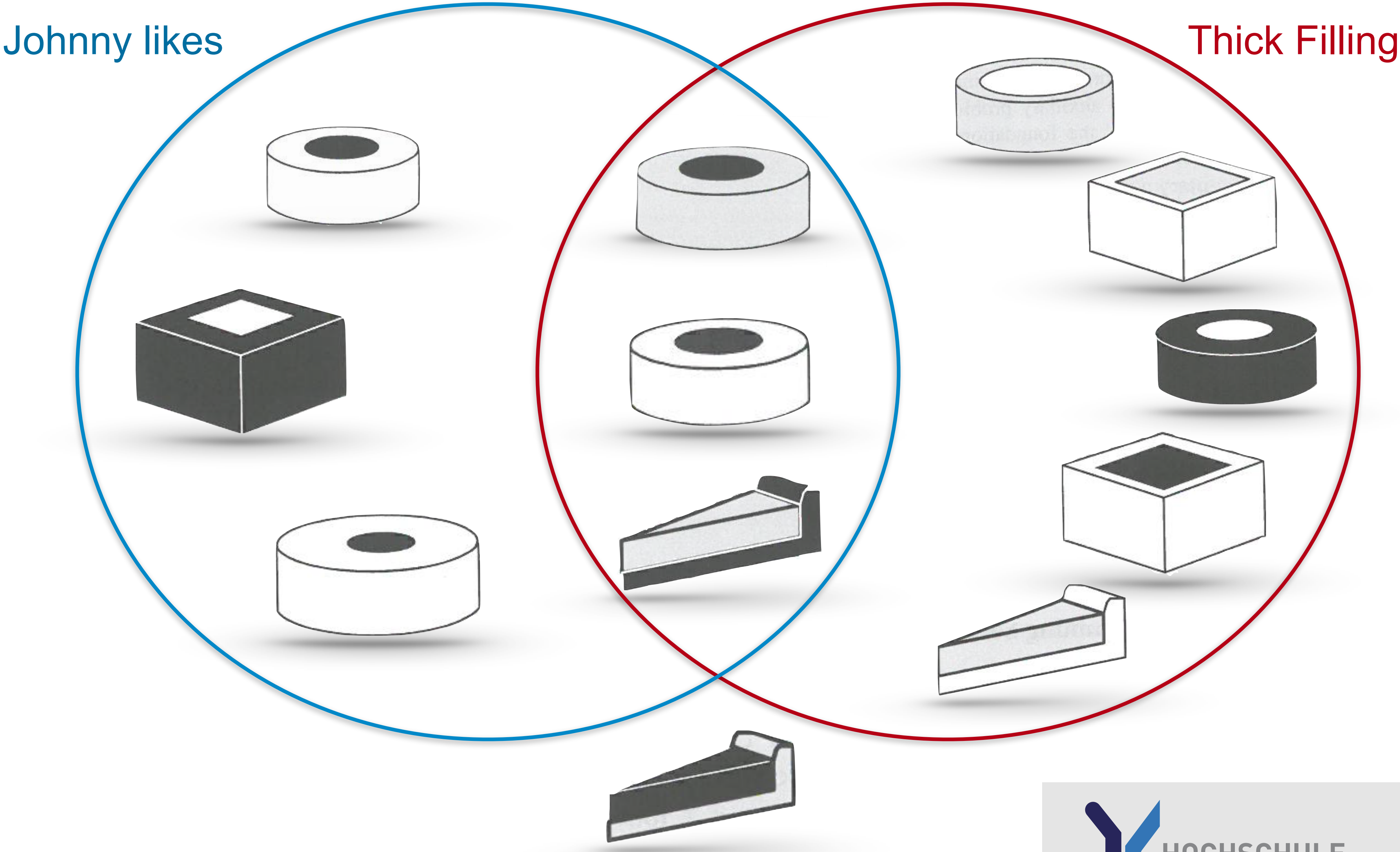
Single-Attribute Case

Probabilities: (Johnny's pies example)

- Training set consists of twelve pies (N_{all} = 12)
- Six positive examples (N_{pos} = 6)
- Six negative examples (N_{neg} = 6)

1. what is the probability of Johnny's liking a randomly picked pie?
2. what is the relative frequency of positive examples among those with thick filling?

The Prior Probabilities



Probability

- Conditional Probability
  - $P(pos|thick) = 3/8$
- Joint Probability
  - $P(pos,thick) = P(pos \cap thick) = 3/12$

- $P(pos|thick).P(thick) = P(thick|pos).P(pos)$

Exercise

Calculate the following probabilities with the filling-size (thick or thin) to recognise the positive class given the following training examples:

	ex1	*ex2*	*ex3*	*ex4*	*ex5*	*ex6*	*ex7*	*ex8*
size	*thick*	*thick*	*thin*	*thin*	*thin*	*thick*	*thick*	*thick*
class	*pos*	*pos*	*pos*	*pos*	*neg*	*neg*	*neg*	*neg*

- P(thin), P(thick), P(pos), P(neg)
- P(thin|pos), P(thick|pos), P(thin|neg), P(thick|neg)
- P(pos|thin), P(pos|thick), P(neg|thin), P(neg|thick)

Bayes Formula

$$P(A|B) = \frac{P(B|A).P(A)}{P(B)}$$

$$P(c_i|\mathbf{x}) = \frac{P(\mathbf{x}|c_i).P(c_i)}{P(\mathbf{x})}$$

$$\mathbf{x} = (x_1, x_2, \ldots, x_n)$$

Naive Bayes

under assumption of mutually independent attributes, a random representative of c_j is described by $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is calculated as follows:

$$P(\mathbf{x}|c_j) = \prod_{i=1}^{n} P(x_i|c_j)$$

an object will be labelled c_j is this class maximises the Bayes formula's numerator:

$$P(c_j).\prod_{i=1}^{n} P(x_i|c_j)$$

Training Examples "Johnny's Pie"

example	shape	crust		filling		class
		size	shade	size	shade	
ex1	circle	thick	gray	thick	dark	pos
ex2	circle	thick	white	thick	dark	pos
ex3	triangle	thick	dark	thick	gray	pos
ex4	circle	thin	white	thin	dark	pos
ex5	square	thick	dark	thin	white	pos
ex6	circle	thick	white	thin	dark	pos
ex7	circle	thick	gray	thick	white	neg
ex8	square	thick	white	thick	gray	neg
ex9	triangle	thin	gray	thin	dark	neg
ex10	circle	thick	dark	thick	white	neg
ex11	square	thick	white	thick	dark	neg
ex12	triangle	thick	white	thick	gray	neg

Exercise

Given the "Johnny's pie domain" determine the class of the following object:

```
x = [shape=square, crust-size=thick, crust-shade=gray, filling-
size=thin, filling-shade=white]
```

(Hint: use the two classes P(pos) and P(neg) for calculating the probabilities)

P(x|pos) ?

P(x|neg) ?

Expert Intuition (m-estimate)

Example in coins P(heads) given frequency of the experiments N_{all} and N_{heads}

$$P_{heads} = \frac{N_{heads} + m\pi_{heads}}{N_{all} + m}$$

Exercise

Calculate the relative frequency and m-estimate of the given successive trial

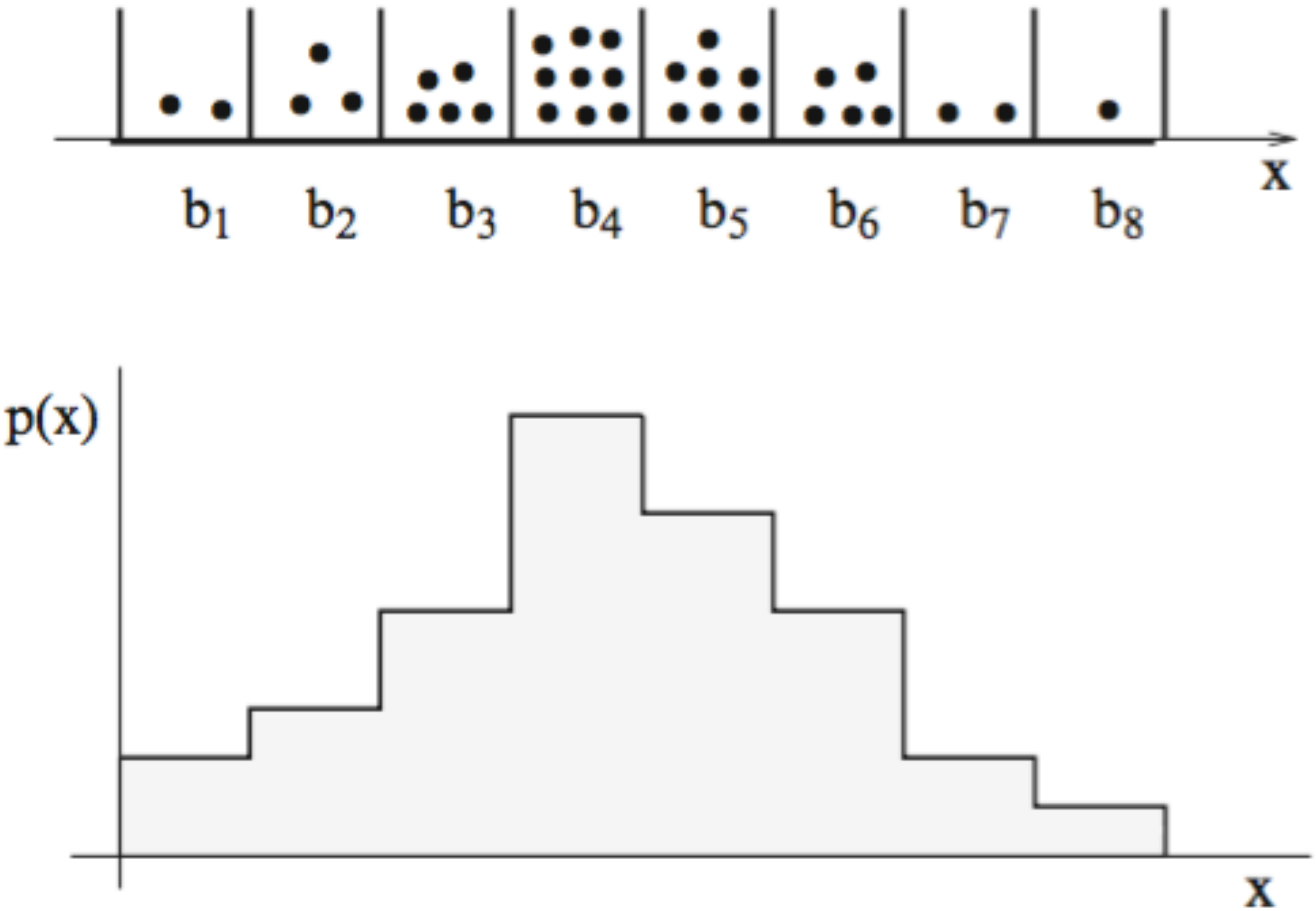
Toss number	*1*	*2*	*3*	*4*	*5*	*6*
Outcome	*Heads*	*Heads*	*Tails*	*Heads*	*Tails*	*Tails*
Relative Frequency						
m-estimate						

How about Continuous Attributes?

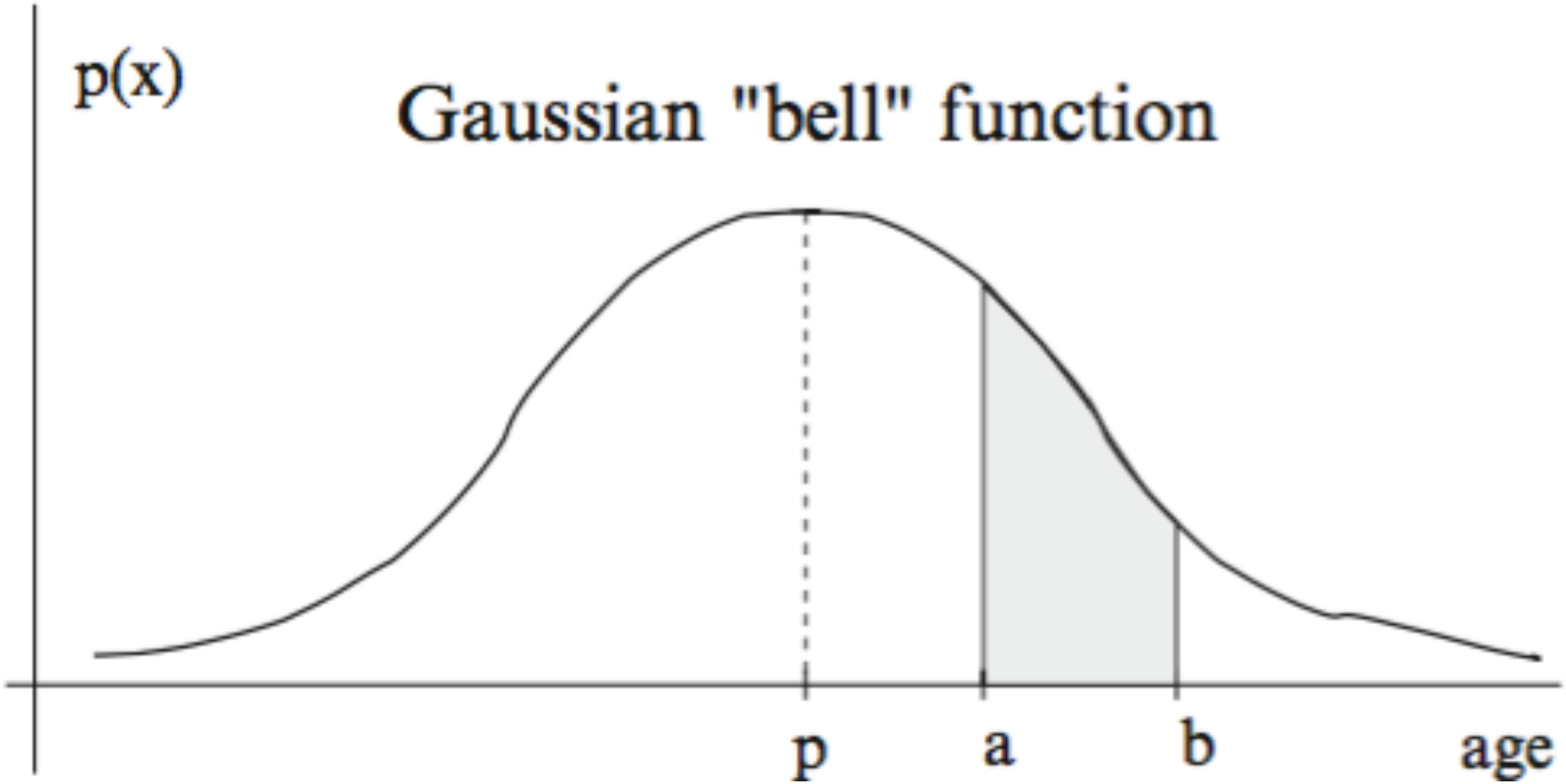
- Many attributes are continuous, e.g. age
- How to deal with such attributes?
  - discretization with certain intervals
    - (0, 10], … (90,100]
    - count the value and place the results on each interval bin
    - N_i/N - relative frequency of the value in the i-th bin
    - total all relative frequency is one

$$\frac{\Sigma N_i}{N} = 1$$

Discretization Method



Gaussian Function



p(x) is a probability density function

Bayes Formula for Continuous Domain

$$P(c_i|x) = \frac{P_{ci}(x).P(c_i)}{P(x)}$$

Naive Bayes:

$$P_{cj}(\mathbf{x}) = \prod_{i=1}^{n} P_{cj}(x_i)$$

HOCHSCHULE RHEIN-WAAL
Rhine-Waal University of Applied Sciences

Gaussian "Bell" Function

$$p(x) = k.e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

where:

$$k = \frac{1}{\sqrt{2\pi\sigma^2}}$$

$$\mu = \frac{1}{m}\sum_{i=1}^{m} x_i \qquad \sigma^2 = \frac{1}{m-1}\sum_{i=1}^{m}(x_i - \mu)^2$$

Combining Gaussian Functions

$$p(x) = k.\sum_{i=1}^{m} e^{-\frac{(x-\mu_i)^2}{2\sigma^2}}$$

$$k = \frac{1}{(2\pi)^{m/2}\sigma^m}$$

Exercise

Combine bell functions of a training set consisting of m = 3 examples, where $x_1 = 0.4$, $x_2 = 0.5$ and $x_3 = 0.7$. The variance of all examples is one. The mean $\mu_1 = 0.4$, $\mu_2 = 0.5$, and $\mu_3 = 0.7$.

Homework

The following is a training set with three continuous attributes

Example	*at1*	*at2*	*at3*	*class*
ex1	*3.2*	*2.1*	*2.1*	*pos*
ex2	*5.2*	*6.1*	*7.5*	*pos*
ex3	*8.5*	*1.3*	*0.5*	*pos*
ex4	*2.3*	*5.4*	*2.45*	*neg*
ex5	*6.2*	*3.1*	*4.4*	*neg*
ex6	*1.3*	*6.0*	*3.35*	*neg*

a. find the most probable class of x = (9, 2.6, 3.3)

b. find the most probable class of x = (1.4, 3.3, 3.0)