Quiz 1 - sample

Note Title 9/3/2007

1. For the following "widget" dataset, compute

   a) decision rule
   b) error rate on training set

   using the 1R algorithm.

2. For the same dataset, but using the naïve Bayes algorithm, compute

   a) decision rule only for "normal, loud" widgets
   b) what is the probability that a "normal, loud" widget is "bad"?
   c) clearly identify the <u>prior probabilities</u>, <u>likelihoods</u>, and <u>posterior probabilities</u> you used for (a) & (b).

   d) what specific independence assumption did you use to answer (a) & (b)?

Data set for widgets :

appearance	sound	class
normal	quiet	good
normal	quiet	good
deformed	quiet	bad
deformed	loud	bad
normal	quiet	bad
deformed	quiet	good
normal	quiet	good
normal	quiet	good
normal	loud	bad
normal	loud	bad
normal	loud	bad
deformed	quiet	bad
normal	loud	good
normal	quiet	good
normal	quiet	good
normal	quiet	good

Solution:

1. 1R for appearance:

appearance	good	bad	decision	errors
normal	8	4	good	4/12
deformed	1	3	bad	1/4

total errors: 5/16

1R for sound:

sound	good	bad	decision	errors
quiet	8	3	good	3/11
loud	1	4	bad	1/5

total errors: 4/16

a) decision rule: attribute 'sound' has lowest error rate, so choose that.

Rule is: quiet → good
loud → bad

b) error rate on training set: 4/16 (i.e. 1/4).

2. Need to compute $P(G|N,L)$ and $P(B|N,L)$ using Bayes' Rule, assuming independence of features:

$$P(G|N,L) = P(N,L|G) \times P(G) \;/\; P(N,L)$$
$$= P(N|G) \times P(L|G) \times P(G) \;/\; P(N,L)$$

$$P(B|N,L) = P(N,L|B) \times P(B) \;/\; P(N,L)$$
$$= P(N|B) \times P(L|B) \times P(B) \;/\; P(N,L)$$

summary statistics we will need :

instances :	16
good :	9
bad :	7
normal & good :	8
normal & bad :	4
loud & good :	1
loud & bad :	4

probability estimates :

$$P(G) = 9/16$$
$$P(B) = 7/16$$
$$P(N|G) = 8/9$$
$$P(N|B) = 4/7$$
$$P(L|G) = 1/9$$
$$P(L|B) = 4/7$$

plugging in:

$$P(G|N,L) = \frac{8}{9} \times \frac{1}{9} \times \frac{9}{16} \Big/ P(N,L)$$

$$= \frac{1}{18} \Big/ P(N,L) = 0.055 \Big/ P(N,L)$$

$$P(B|N,L) = \frac{4}{7} \times \frac{4}{7} \times \frac{7}{16} \Big/ P(N,L)$$

$$= \frac{1}{28} \Big/ P(N,L) = 0.036 \Big/ P(N,L)$$

These 2 probabilities must sum to 1, so we know immediately that $P(N,L) = \frac{1}{18} + \frac{1}{28} = 0.091$

So,

$$P(G|N,L) = \frac{0.055}{0.091} = 0.61$$

$$P(B|N,L) = \frac{0.036}{0.091} = 0.39$$

So 'Good' has higher posterior probability for normal, loud widgets

i.e. decision rule is

noisy, loud $\rightarrow$ good.

b) prob that normal, loud widget is bad $= P(B|N,L) \approx 0.39$

c) Prior probs $-$ $P(G)$, $P(B)$

Likelihoods $-$ $P(N|G)$, $P(N|B)$, $P(L|G)$, $P(L|B)$

Posteriors $-$ $P(G|N,L)$, $P(B|N,L)$

d) 2 independence assumptions:

$$P(N,L|G) = P(N|G)\,P(L|G)$$

$$P(N,L|B) = P(N|B)\,P(L|B)$$