Abelardo Riojas
Data Mining HW 4
Prof. Meyer-Baese

1. Attributes:

AC: Working, Broken
Engine: Good, Bad
Milage: High, Medium, Low
Rust: Yes, No

Class: Value $\rightarrow$ (high, low)

Rules (R)

r_1: (mileage = high) $\rightarrow$ Value = low
r_2: (milage = low) $\rightarrow$ Value = high
r_3: (AC = ~~good~~ W) $\wedge$ (Engine = Good) $\rightarrow$ Value = high
r_4: (AC = W) $\wedge$ (Engine = Bad) $\rightarrow$ Value = low
r_5: (AC = B) $\rightarrow$ Value = low

a) Mutually exclusive?

~~The rules are mutually exclusive as there isn't a combination of the attribute set that would trigger more than one rule.~~

The rules are not mutually exclusive as a car with low milage and a broken AC would be classified under both classes by r_2 and r_5.

b) Exhaustive?

The rules are exhaustive because all records will be triggered by either rule 1 or rule 2.

c) ordering needed?

Yes because the set of rules is not mutually exclusive, there could be conflict issues between rules

d) Default class needed?

No because the rules are exhaustive over all possible combinations of the training set.

5. Figure 5.1

~~a) Likelihood ratio statistic~~

29 Positive Examples 21 Negative Examples 50 total

r_1 = covers 12 positive and 3 negative
r_2 = covers ~~8~~ 7 positive and 3 negative
r_3 = covers 8 positive and 4 negative

r_1 accuracy = $\frac{12}{15}$ = 80% Coverage = $\frac{15}{50}$ = ~~33~~ 30%
r_2 accuracy = $\frac{7}{10}$ = 70% Coverage = $\frac{10}{50}$ = 20%
r_3 accuracy = $\frac{8}{12}$ = 66% Coverage = $\frac{12}{50}$ = 24%

a) Likelihood Ratio Statistic

$$R(r_n) = 2 \sum_{i=1}^{K} f_i \log_2(f_i/e_i)$$

K = # of classes
f_i = Frequency of class i in rule.

$$f_i = \frac{\text{\# of class } i}{\text{total \# of instances in rule}}$$

$$e_i = \text{\# of instances covered by rule} \times \frac{\text{total \# of class } i}{\text{total instances}}$$

~~$R(r_1) = 2 \times (12 \times \log_2 (50 / (15 \times \frac{29}{50}))) \times (3 \times \log_2 \frac{50}{22} (15 \times \frac{21}{50})))$~~

$R(r_1) = 2 \times (12 \times \log_2 (12 / (15 \times \frac{29}{50})) \times (3 \times \log_2 (3 / (15 \times \frac{21}{50}))),$

$e_+ = 8.7 \quad e_- = 6.3$

$R(r_1) = 2 \cdot (12 \cdot \log_2 (12/8.7)) \times (3 \cdot \log_2 (3/6.3)) =$

$\boxed{R(r_1) = 13.09}$ Best

R(r2)= This was tedious so I made a function

```
function R = like(cP, cN, tP, tN)      % cP = covered Pos
eP= (cP+cN) . (tP/(tP+tN));            % tP = total Pos
eN = (cP+cN).(tN/(tP+tN));             % cN = covered Neg
R = 2 x ((cP . log2(cP/eP)) + (cN . log2(   % tN = total Neg
  cN/eN)));
end
```

$R(r_2) = \text{like}(7, 3, 29, 21) = 4.6116$

$R(r_3) = \text{like}(8, 4, 29, 21) = \boxed{4.2725}$ Worst

b) Laplace Measure

$L = \frac{f_+ + 1}{n + k}$

n = # of covers

f_+ = # of positive covers

k = classes

$L(R_1) = \frac{13}{17} = .764$

Best

$L(R_2) = \frac{8}{12}$

.666

$L(R_3) = \frac{9}{14}$

.64

Worst

c) m- estimate

$$m = \frac{F_+ + kP_+}{n+k}$$

$\boxed{m(r_1) = \frac{12 + 2(.58)}{17}}$ $m(r_2) = \frac{7 + 2(.58)}{12}$ $\boxed{m(r_3) = \frac{8 + 2(.58)}{14}}$

Best Worst.

d) Accuracy (R_1 examples not deleted)

$\boxed{r_1 = \frac{12}{15} = 80\%}$ Best $r_2 = \frac{7}{10} = 70\%$ $\boxed{r_3 = \frac{8}{12} = 66\%}$ Worst

e) Accuracy (R_1 + examples deleted).

$\cancel{r_1 = \frac{0}{3}}$ $\cancel{r_2 = \frac{6}{\cancel{8}10}}$ $\cancel{r_3 = \frac{7}{10}}$

~~Worst~~ ~~Best~~ $r_1 = \frac{0}{3}$ Worst $r_2 = \frac{7}{10}$ Best $r_3 = \frac{6}{10}$

(~~Not fair though,~~ ~~r_3 is the worst (acc).~~

f) Accuracy (R_1 examples deleted)

$r_3 = \frac{\cancel{6}\,7}{\cancel{8}\,10}$ Worst $r_3 = \frac{6}{8}$ Best

r_1 not a rule anymore does not cover anything.

6. Baye's Theorem

$X =$ ~~Grad~~ UG (0) or G (1)

$Y =$ ~~Smo~~ NS (0) or S (1)

$P(Y=1 \mid X=0) = .15 \quad P(Y=1 \mid X=1) = .23$

$P(X=0) = .8 \quad P(X=1) = .2$

$$P(X=0 \mid Y=1) = \frac{P(Y=1 \mid X=0)\,P(X=0)}{P(Y=1 \mid X=0)\,P(X=0) + P(Y=1 \mid X=1)\,P(X=1)}$$

$$\frac{.15 \times .8}{(.15 \times .8) + (.23 \times .2)} = .722$$

$P(X=1 \mid Y=1) = 1 - P(X=0 \mid Y=1) = \boxed{.277}$

b) Undergraduate because they make up 80% of the class.

c) Assuming they smoke, it's still more likely they're undergrad as demonstrated in part A.

d) Assuming independence.

$P(\text{Smoking Grad} \cap \text{Dorm}) = P(X=1 \mid Y=1) \times P(Y=1 \mid D=1)$

$.277 \times .3 = .0831$

$P(\text{Smoking Ugrad} \cap \text{Dorm}) = .722 \times .1 = .0722$

$\boxed{\text{Smoking Graduate student more likely}}$

$P(B = \ldots) = \frac{2}{\ldots}$ $P(B \ldots)$

$P(C = 1 \mid +) = \frac{4}{5}$ $P(C \ldots)$

$P(C = 1 \mid \ldots) = \frac{\ldots}{5}$ $P(C = \ldots)$

$X = \{\overset{A}{0}, \overset{B}{1}, \overset{C}{\ldots}\}$

B) $P(+ \mid X) \propto P(+) \cdot P(A \ldots)$

$P(+ \mid X) \propto \frac{1}{2} \cdot \frac{2}{5} \cdot \ldots$

$P(- \mid X) \propto P(-) \cdot P(A = 0 \mid -) \ldots$

$P(- \mid X) \propto \frac{1}{2} \cdot \frac{3}{5} \cdot \frac{\ldots}{5} \ldots$

$\propto \frac{1}{250}$ ... $+$!

C) M estimate $p = 1/2$ $m = 4$

$P(A = 0 \mid +) = \frac{4}{9}$ $P(A = \ldots \mid -) = \frac{5}{\ldots}$

$P(B = 1 \mid +) = \frac{3}{9}$ $P(A = 1 \mid -) = 4/\ldots$

$P(B = 0 \mid +) = \frac{5}{9}$ $P(C = 0 \mid -) = \frac{2}{9}$

$$P(+|X) \propto P(+) \times P(A=0|+) \times P(B=1|+) \times P(C=0|+)$$

$$\propto .5\ (4/9)\ (3/9)\ (5/9) = .041$$

$$P(-|X) \propto .5\ (5/9)\ (4/9)\ (2/9) = .027$$

Class label is still +

~~Naive Bayes~~

e) Naive Bayes allows probabilities to be zero, which cancels out the whole expression. That means no matter how good A and B are, C can determine the result. Using the m estimate, it does not "zero out."

9. Figure 5.2

~~Both XF~~

A) NB will not do well as the probabilities for ~~both classes~~ all attributes given a class are the same for A and B.

B) NB will perform better now since the distinguishing attributes now have different probabilities for each ~~class~~ subclass (products are different).

C) Entropy will not improve in 2 class problem, but will improve in 4 class.

13. point $x = 5.0$

A) 1 NN

$$|5.0 - \overset{+}{4.7}| < |5.0 - \overset{-}{5.2}| \Rightarrow$$

3 NN $\quad \underset{+}{4.9}, \underset{-}{5.2}, \underset{-}{5.3} \longrightarrow -$

5 NN $\quad \underset{+}{4.5}, \underset{+}{4.6}, \underset{+}{4.9}, \underset{-}{5.2}, \underset{-}{5.3}, \underset{+}{5.5}$ (4.5 or 5.5 Doesn't matter in this case)

$\longrightarrow +$

9 NN $\quad \underset{-}{.05}, \underset{+}{3.0}, \underset{+}{4.5}, \underset{-}{5.2}, \underset{-}{5.3}, \underset{+}{5.5}, \underset{-}{7.0}, \underset{-}{9.5}$ (.05 or 9.5 doesn't matter)

$\longrightarrow -$

B) 1 NN

$$w_1 = \frac{1}{.1} = 10+$$

3 NN

$$w_1^+ = \frac{1}{.1} = 10+$$

$$w_2^- = \frac{1}{.2} = \cancel{3.33}\ 5-$$

$$w_3 = \frac{1}{.3} = 3.33-$$

$10+ > 8.33-$

$\downarrow$

$\boxed{+}$

5 NN

$$w_1^+ = \frac{1}{.1} = 10+ \quad w_2^- = 5- \quad w_3^- = \frac{1}{.3} = 3.33^-$$

$$w_4^+ = \frac{1}{.4} = 2.5+ \quad w_5^+ = \frac{1}{.5} = 2^+$$

$14.5+ > 8.33^-$

$\boxed{+}$

13. (cont).

$W_1 = \frac{1}{.1}$ $W_2 = \frac{-1}{.2}$ $W_3 = \frac{-1}{.3}$ $W_4 = \frac{1}{.4}$

$W_5 = \frac{1}{.5}$ $W_6 = \frac{1}{.3}$ $= \frac{1}{2}$ $W_7 = \frac{-1}{2}$ $W_8 = \frac{-1}{2}$

$W_9 = -\frac{1}{4.5}$ $\rightarrow$ $\boxed{+}$

16.
A)

$y = \varphi(x_1 + x_2 - 1.5)$ And

$y = \varphi(x_1 + x_2 - .5)$ OR

$x_1 = 0$ $x_2 = 0$ (And). (or)

$y = \varphi(0 + 0 - 1.5) = 0$ $y = \varphi(0 + 0 - .5) = 0$

$x_1 = 1$ $x_2 = 0$

$y = \varphi(1 + 0 - 1.5) = 0$ $y = \varphi(1 + 0 - .5) = 1$

$x_1 = 0$ $x_2 = 1$

$y = \varphi(0 + 1 - 1.5) = 0$ $y = \varphi(0 + 1 + .5) = 1$

$x_1 = 1$ $x_2 = 1$

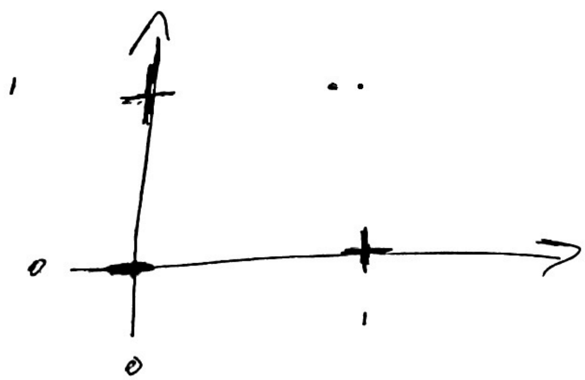
$y = \varphi(1 + 1 - 1.5) = 1$ $y = \varphi(1 + 1 - .5) = 1$

b)

Using Linear functions instead of the activation functions just makes your network architecture into a "fancy" perceptron.

22.

$$\phi = (1, \sqrt{2}x_1, \sqrt{2}x_2, \sqrt{2}x_1x_2, x_1^2, x_2^2)$$



? I have no idea what to do here :(.