Q1 HW1.

ANIL AKYILDIRIM

①

prospecting dataset				
age-group	networth	status	credit_rating	class:prospect
youth	high	employed	fair	no
youth	high	employed	excellent	no
middle	high	employed	fair	yes
senior	medium	employed	fair	yes
senior	low	unemployed	fair	yes
senior	low	unemployed	excellent	no
middle	low	unemployed	excellent	yes
youth	medium	employed	fair	no
youth	low	unemployed	fair	yes
senior	medium	unemployed	fair	yes
youth	medium	unemployed	excellent	yes
middle	medium	employed	excellent	yes
middle	high	unemployed	fair	yes
senior	medium	employed	excellent	no

You have been hired by a local electronics retailer and the above dataset has been given to you.

Manager Bayes Jr.9th wants to create a spreadsheet to predict is a customer is likely prospect. To that end

1)Compute prior probabilities for the Prospect Yes/No

2) Compute the conditional probabilities

P(age-group=youth|prospect=yes) and

P(age-group=youth|prospect=no)

where age-group is a predictor variable.

Compute the conditional probabilities for each predictor variable namely (age_group networth status credit_rating)

3) Assuming the assumptic compute the posterior probability P(prospect|X) where X is one of the predictor variable.

Recommendation: Do this by hand and submit name=value format where name

Prior Priorities P(prospect=yes)

Prior Priorities P(prospect=no)

Conditional Probabilities:

P(age-group=youth|prospect=yes)

P(age-group=middle|prospect=yes)

P(age-group=senior|prospect=yes)

P(age-group=youth|prospect=no)

P(age-group=middle|prospect=no)

P(age-group=senior|prospect=no)

P(networth=high|prospect=yes)

P(networth=low|prospect=yes)

P(networth=medium|prospect=yes)

①

class = prospect

$P(\text{prospect} = \text{yes}) = \frac{9}{14}$ $\quad p(\text{prospect} = \text{no}) = \frac{5}{14}$

INDIVIDUAL PROBABILITIES OF EACH VARIABLE

$p(\text{age-group} = \text{youth} \mid \text{yes}) = \frac{2}{9}$

$p(\text{age-group} = \text{youth} \mid \text{prospect} = \text{no}) = \frac{3}{5}$

$p(\text{age-group} = \text{middle} \mid \text{prospect} = \text{yes}) = \frac{4}{9}$

$p(\text{age-group} = \text{middle} \mid \text{prospect} = \text{no}) = \frac{0}{5}$

$p(\text{age-group} = \text{senior} \mid \text{prospect} = \text{yes}) = \frac{3}{9}$

$p(\text{age-group} = \text{senior} \mid \text{prospect} = \text{no}) = \frac{2}{9}$

$p(\text{net-worth} = \text{high} \mid \text{prospect} = \text{yes}) = \frac{2}{9}$

$p(\text{net-worth} = \text{high} \mid \text{prospect} = \text{no}) = \frac{2}{5}$

$p(\text{net-worth} = \text{medium} \mid \text{prospect} = \text{yes}) = \frac{4}{9}$

$p(\text{net-worth} = \text{medium} \mid \text{prospect} = \text{no}) = \frac{2}{5}$

$p(\text{net-worth} = \text{low} \mid \text{prospect} = \text{yes}) = \frac{3}{9}$

$p(\text{net-worth} = \text{low} \mid \text{prospect} = \text{no}) = \frac{1}{5}$

$P(\text{status} = \text{employed} \mid \text{prospect} = \text{yes}) = \frac{3}{9}$

$P(\text{status} = \text{employed} \mid \text{prospect} = \text{no}) = \frac{4}{5}$

$P(\text{status} = \text{unemployed} \mid \text{prospect} = \text{yes}) = \frac{6}{9}$

$P(\text{status} = \text{unemployed} \mid \text{prospect} = \text{no}) = \frac{1}{5}$

$P(\text{credit\_rating} = \text{fair} \mid \text{prospect} = \text{yes}) = \frac{6}{9}$

$P(\text{credit\_rating} = \text{fair} \mid \text{prospect} = \text{no}) = \frac{2}{5}$

$P(\text{credit\_rating} = \text{excellent} \mid \text{prospect} = \text{yes}) = \frac{3}{9}$

$P(\text{credit\_rating} = \text{excellent} \mid \text{prospect} = \text{no}) = \frac{3}{5}$

Prior probability $\Rightarrow P(\text{prospect} = \text{yes}) = \frac{9}{14}$

Prior probability $\Rightarrow P(\text{prospect} = \text{no}) = \frac{5}{14}$

Probability that the customer is likely a prospect

prospect = yes

$P(\text{age-group} = \text{youth} \mid \text{prospect} = \text{yes}) = \frac{2}{9}$

$P(\text{net\_worth} = \text{low} \mid \text{prospect} = \text{yes}) = \frac{3}{9}$

$P(\text{status} = \text{employed} \mid \text{prospect} = \text{yes}) = \frac{3}{9}$

$P(\text{credit\_rating} = \text{excellent} \mid \text{prospect} = \text{Yes}) = \frac{3}{9}$

Probability that the customer is likely ~~not~~ a prospect

prospect = no

$P(\text{age-group} = \text{youth} \mid \text{prospect} = \text{No}) = \frac{3}{5}$

$P(\text{net\_worth} = \text{low} \mid \text{prospect} = \text{No}) = \frac{1}{5}$

$P(\text{status} = \text{employed} \mid \text{prospect} = \text{No}) = \frac{4}{5}$

$P(\text{credit\_rating} = \text{excellent} \mid \text{prospect} = \text{No}) = \frac{3}{5}$

$X = \{$ age-group, net_worth, status, credit $\}$

$C = \{$ prospect $)$

NAIVE BAYES

POSTERIOR PROBABILITY $\Rightarrow P(A|C)$

LIKELIHOOD $\Rightarrow P(C|A)$

PRIOR PROBABILITY $\Rightarrow P(A)$

NORMALIZING CONSTANT $\Rightarrow P(C)$

$P(X \mid prospect=yes) \cdot P(prospect=yes) =$

$= \frac{2}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{3}{9} \cdot \frac{9}{14}$

$= \frac{486}{91854} = 0.00529$

$P(X \mid prospect=no) \cdot P(prospect=no) =$

$= \frac{3}{5} \cdot \frac{1}{5} \cdot \frac{4}{5} \cdot \frac{3}{5} \cdot \frac{5}{14}$

$= \frac{180}{8750} = 0.0205$

Normalizing Constant $= 0.00529 + 0.0205$

$= 0.02579$

Posterior Probability for X={age-group=youth, net_worth=low, status=employed, credit_rating=excellent}

$$P(\text{prospect=yes} \mid X) = \frac{0.00529}{0.0259} = 0.204 \simeq \%20$$

$$P(\text{prospect=no} \mid X) = \frac{0.0205}{0.0259} = 0.791 \simeq \%80$$

Total probability $\Rightarrow$ $\%20 + \%80 = 1$