Lorenzo Gangi
CS557 Univeristy of Montana
Homework2 Naïve Bayes Classifier
2.27.08

Theory

1) Prove the following using axioms of probability and the definition of conditional probability given three binary variables X,Y,Z

a) $P(x \cap y) = P(x \cap y \cap \neg z) + P(x \cap y \cap z)$

a.1First we need to show $P(x) = P(x \cap y) + P(x \cap \neg y)$
By Axiom 1 $\Omega = Universe = 1$
$P(x) = P(x \cap \Omega) = p(x \cap (y \cap \neg y)) = P((x \cap y)) \cup (x \cap \neg y))$
By Axiom 4 (addition rule)
$P(x \cap y) + P(x \cap \neg y) - P(x \cap y \cap x \cap \neg y)$
$P(x) = P(x \cap y) + P(x \cap \neg y)$

a.2 Substitute w for $x \cap y$
$P(w) = P(x \cap z) + P(x \cap \neg z)$
$P(w) = P(w) = P(x \cap y)$

b) $P(x) = P(x \cap y \cap z) + P(x \cap y \cap \neg z) + P(x \cap \neg y \cap z) + P(x \cap \neg y \cap \neg z)$

b.1
a.1 proved
$P(x) = P(x \cap y) + (x \cap \neg y)$

b.2 Substiture $x = x \cap y$ and $v = x \cap \neg y$
$P(x \cap y \cap z) \Rightarrow P(w \cap z)$
$P(x \cap y \cap \neg z) \Rightarrow P(w \cap \neg z)$
$P(x \cap \neg y \cap z) \Rightarrow P(w \cap v)$
$P(x \cap \neg y \cap \neg z) \Rightarrow P(w \cap \neg v)$
$P(x) = P(w \cap z) + P(w \cap \neg z) + P(w \cap v) + P(w \cap \neg v)$
$P(x) = P(w) + P(z)$
$P(x) = P(x \cap y) + P(x \cap \neg y)$
$P(x) = P(x)$

c) $P(x \cup y \cup z) = P(x) + P(y) + (z) - P(x \cap y) - P(x \cap z) + P(x \cap y \cap z)$

c.1 Subtitute $w = x \cup y$

by Axiom 4 we now have
$P(w \cup z) = P(w) + P(z) - P(w \cap z)$

c.2 Replace w and apply axiom 4 a couple times
$P(x \cup y \cup z) = P(x \cup y) + P(z) - P((x \cup y) \cap z))$
$P(x \cup y \cup z) = P(x) + P(y) + P(z) - P(x \cap y) - P((x \cap z) \cup (y \cap z))$
$P(x \cup y \cup z) = P(x) + P(y) + P(z) - P(x \cap y) - P(x \cap z) + P(y \cap z) - P((x \cap z) \cap (y \cap z))$
$P(x \cup y \cup z) = P(x) + P(y) + P(z) - P(x \cap y) - P(x \cap z) + P(y \cap z) - P(x \cap y \cap z)$

d) $P(x|y) = 1 - P(\neg x|y)$
d.1 by the definition of conditional probability
$$P(x|y) = \frac{P(x \cap y)}{P(y)}$$
$$1 - P(\neg x|y) = 1 - \frac{P(\neg x \cap y)}{P(y)}$$
A little rearranging gives us
$$1 = \frac{P(x \cap y)}{P(y)} + \frac{P(\neg x \cap y)}{P(y)} = \frac{P(x \cap y) + P(\neg x \cap y)}{P(y)} = \frac{P(y)}{P(y)} = 1$$

2) Bayes Rule

Know
$P(p|vCJD) = .95$
$P(p|\neg vCJD) = .30$
$$P(vCJD) = \frac{1}{1{,}000{,}000}$$
$$P(\neg vCJD) = \frac{999{,}999}{1{,}000{,}000}$$

Microsoft Equation 3.0

a)What is the probability of having paralysis and not vCJD
By the product rule we can say:
$P(p \cap \neg vCJD) = P(p|\neg vCJD) * P(\neg vCJD)$
$$P(p \cap \neg vCJD) = .30 * \frac{999{,}999}{1{,}000{,}000} = .2999$$

b)What is the probability some one who has paralysis in the us has vCJD

$$P(vCJD|p) = \frac{P(vCJD \cap p)}{P(p)} = \frac{P(p|vCJD)P(vCJD)}{P(p \cap vCJD) + P(p \cap \neg vCJD)}$$

$$P(p \cap \neg vCJD) = .2999$$

$$P(p \cap vCJD) = P(p|vCJD) * P(vCJD) = .95 * \frac{1}{1{,}000{,}000} = .00000095$$

$$\Rightarrow P(vCJD|p) = \frac{95 * \frac{1}{1{,}000{,}000}}{.2999 + .00000095} = 3.33e-8$$

Not very likely

c)what is the probility some one in from Great Britain who has paralysis has vCJD; P(vCJD)=1/100,000
again

$$P(vCJD|p) = \frac{P(vCJD \cap p)}{P(p)} = \frac{P(p|vCJD)P(vCJD)}{P(p \cap vCJD) + P(p \cap \neg vCJD)}$$

But

$$P(p \cap \neg vCJD) = .30 * \frac{99{,}999}{100{,}000} = .2999997$$

$$P(p \cap vCJD) = .95 * \frac{1}{100{,}000} = .0000095$$

$$\Rightarrow P(vCJD|p) = \frac{95 * \frac{1}{100{,}000}}{.2999997 + .0000095} = 3.16 - 5$$

Still not very likely

2) Density Estimator and Bayes Classifier

a)Estimate all terms of all X,Y,Z that JD will produce

$$P(x, y, z) = \frac{\#ofcases(X = x, Y = y, Z = z) + .125}{totalcases + 1}$$

Example case A

$$P(x = t, y = t, z = t) = \frac{\#ofcases(X = x, Y = y, Z = z) = 1 + .125}{totalcases + 1 = 9 + 1} = \frac{1.125}{10} = .1125$$

Cases	X	Y	Z	Occurences	Calculated P(case)
A	t	t	t	1	.1125

B	t	f	t	3	.3125
C	f	t	f	2	.2125
D	f	f	f	1	.1125
E	t	t	f	1	.1125
F	f	t	t	1	.1125
G	f	f	t	0	.0125
H	t	f	f	0	.0125

b)

b.1 Use JD to calculate P(x=t|y=t)

$$P(x=t|y=t)=\frac{P(x\cap y)}{P(y)}=\frac{P(x\cap y\cap z)+P(x\cap y\cap\neg z)}{P(x\cap y\cap z)+P(x\cap y\cap\neg z)+P(\neg x\cap y\cap z)+P(\neg x\cap y\cap\neg z)}$$

$$P(x=t|y=t)=\frac{.1125+.1125}{.1125+.1125+.1125+.2125}=.40909$$

b.2 Use JD to calculate P(z=t|x=t)

$$P(z=t|x=t)=\frac{P(z\cap x)}{P(x)}=\frac{P(x\cap y\cap z)+P(x\cap\neg y\cap z)}{P(x\cap y\cap z)+P(x\cap y\cap\neg z)+P(x\cap\neg y\cap z)+P(x\cap\neg y\cap\neg z)}$$

$$P(z=t|x=t)=\frac{.1125+.3125}{.1125+.1125+.3125+.0125}=.77272$$

c) Estimate all terms of all X,Y,Z that Naïve Bayes Classifier will produce

$$P(x,y)=\frac{\#ofcases(X=x,Y=y)+.25}{totalcases+1}$$

$$P(x,z)=\frac{\#ofcases(X=x,Z=z)+.25}{totalcases+1}$$

c.1 calculate $P(x\cap y)$ and $P(x\cap z)$ from set D

x	y|z	Occurrences $x\cap y$	Occurrences $x\cap z$	$P(x\cap y)$	$P(x\cap z)$
T	T	2	4	.225	.425
T	F	3	1	.325	.125
F	T	3	1	.325	.125
F	F	1	3	.125	.325

c.2 calculate the Bayes terms conditional probability terms

Calculation example term a

$$P(y|x) = \frac{P(y \cap x)}{P(x)} = \frac{P(y \cap x)}{P(x \cap y) + P(x \cap \neg y)} = \frac{.225}{.225 + .325} = .40909$$

$$P(z|x) = \frac{P(z \cap x)}{P(x)} = \frac{P(z \cap x)}{P(x \cap z) + P(x \cap \neg z)} = \frac{.425}{.425 + .125} = .77272$$

$$P(y|\neg x) = \frac{P(y \cap \neg x)}{P(\neg x)} = \frac{P(y \cap \neg x)}{P(\neg x \cap y) + P(\neg x \cap \neg y)} = \frac{.325}{.325 + .125} = .72222$$

$$P(z|\neg x) = \frac{P(z \cap \neg x)}{P(\neg x)} = \frac{P(z \cap \neg x)}{P(\neg x \cap z) + P(\neg x \cap \neg z)} = \frac{.125}{.125 + .325} = .27777$$

$P(y\|x)$	.40909	$\Rightarrow P(\neg y\|x) = 1 - .40909 = .59091$
$P(z\|x)$	.77272	$\Rightarrow P(\neg z\|x) = 1 - .77272 = .22728$
$P(x)$	.525	$\Rightarrow P(\neg x) = 1 - .525 = .475$
$P(y\|\neg x)$	.72222	$\Rightarrow P(\neg y\|\neg x) = 1 - .72222 = .27778$
$P(z\|\neg x)$	.27777	$\Rightarrow P(\neg z\|\neg x) = 1 - .27777 = .72223$

d)

d.1 Use NB to calculate $P(x|y)$

$$P(x|y) = \frac{P(x \cap y)}{P(y)} = \frac{P(x \cap y)}{P(y \cap x) + P(y \cap \neg x)} = \frac{.225}{.225 + .325} = .40909$$

d.2 Use NB to calculate $P(z|x)$

$$P(z|x) = \frac{P(z \cap x)}{P(x)} = \frac{P(z \cap x)}{P(x \cap z) + P(x \cap \neg z)} = \frac{.425}{.425 + .125} = .77272$$

e) Use JD to predict the following cases

e.1 calculation examples for y=t, z=t and y=t, z=f subsequent probabilities are calculated in the same way but the y and z boolean values are substituted according to the table below

$$x = t \Rightarrow P(y \cap z|x) = \frac{P(y \cap z \cap x)}{P(x)} = \frac{P(x \cap y \cap z)}{P(x \cap y \cap z) + P(x \cap y \cap \neg z) + P(x \cap \neg y \cap z) + P(x \cap \neg y \cap \neg z)} = \frac{.1125}{.1125 + .1125 + .3125 + .0125} = \frac{.1125}{.55} = .20454$$

$$x = f \Rightarrow P(y \cap z|\neg x) = \frac{P(y \cap z \cap \neg x)}{P(\neg x)} = \frac{P(\neg x \cap y \cap z)}{P(\neg x \cap y \cap z) + P(\neg x \cap y \cap \neg z) + P(\neg x \cap \neg y \cap z) + P(\neg x \cap \neg y \cap \neg z)} = \frac{.1125}{.1125 + .2125 + .0125 + .1125} = \frac{.1125}{.45} = .25$$

$$x = t \Rightarrow P(y \cap \neg z|x) = \frac{P(y \cap \neg z \cap x)}{P(x)} = \frac{P(x \cap y \cap \neg z)}{P(x \cap y \cap z) + P(x \cap y \cap \neg z) + P(x \cap \neg y \cap z) + P(x \cap \neg y \cap \neg z)} = \frac{.1125}{.1125 + .1125 + .3125 + .0125} = \frac{.1125}{.55} = .20454$$

$$x = f \Rightarrow P(y \cap \neg z|\neg x) = \frac{P(y \cap z \cap \neg x)}{P(\neg x)} = \frac{P(\neg x \cap y \cap \neg z)}{P(\neg x \cap y \cap z) + P(\neg x \cap y \cap \neg z) + P(\neg x \cap \neg y \cap z) + P(\neg x \cap \neg y \cap \neg z)} = \frac{.2125}{.1125 + .2125 + .0125 + .1125} = \frac{.1125}{.45} = .47222$$

Y	Z	$P(y \cap z\|x)$	$P(y \cap z\|\neg x)$	Prediction of X
T	T	.20454	.25	F
T	F	.20454	.47222	F
F	T	.56818	.02777	T

F	F	.02272	.25	F

f)Use NB to predict the following cases

f.1 calculation examples for y=t, z=t and y=t, z=f subsequent probabilities are calculated in the same way but the y and z boolean values are substituted according to the table below

$$x=t \Rightarrow P(y \cap z|x)=\frac{P(y \cap z \cap x)}{P(x)}=\frac{P(y \cap z|x)P(x)}{P(x)}=P(y|x)P(z|x)=.40909*.77272=.31611$$

$$x=f \Rightarrow P(y \cap z|\neg x)=\frac{P(y \cap z \cap \neg x)}{P(\neg x)}=\frac{P(y \cap z|\neg x)P(\neg x)}{P(\neg x)}=P(y|\neg x)P(z|\neg x)=.72222*.27777=.20061$$

$$x=t \Rightarrow P(y \cap \neg z|x)=\frac{P(y \cap \neg z \cap x)}{P(x)}=\frac{P(y \cap \neg z|x)P(x)}{P(x)}=P(y|x)P(\neg z|x)=.40909*.22728=.09297$$

$$x=f \Rightarrow P(y \cap \neg z|\neg x)=\frac{P(y \cap \neg z \cap \neg x)}{P(\neg x)}=\frac{P(y \cap \neg z|\neg x)P(\neg x)}{P(\neg x)}=P(y|\neg x)P(\neg z|\neg x)=.72222*.72223=.52160$$

Y	Z	$P(y \cap z\|x)$	$P(y \cap z\|\neg x)$	Prediction of X
T	T	.31611	.20061	T
T	F	.09297	.52160	F
F	T	.45660	.07715	T
F	F	.13430	.20062	F

g)Use JD to predict cases for MAPC

calculations

$$P(x|y \cap z)=\frac{P(x \cap y \cap z)}{P(y \cap z)}=\frac{P(x \cap y \cap z)}{P(x \cap y \cap z)+P(\neg x \cap y \cap z)}=\frac{.1125}{.1125+.1125}=.5$$

$$P(\neg x|y \cap z)=\frac{P(\neg x \cap y \cap z)}{P(y \cap z)}=\frac{P(\neg x \cap y \cap z)}{P(x \cap y \cap z)+P(\neg x \cap y \cap z)}=\frac{.1125}{.1125+.1125}=.5$$

$$P(x|y \cap \neg z)=\frac{P(x \cap y \cap \neg z)}{P(y \cap \neg z)}=\frac{P(x \cap y \cap \neg z)}{P(x \cap y \cap \neg z)+P(\neg x \cap y \cap \neg z)}=\frac{.1125}{.1125+.2125}=.34615$$

$$P(\neg x|y\cap\neg z)=\frac{P(\neg x\cap y\cap\neg z)}{P(y\cap\neg z)}=\frac{P(\neg x\cap y\cap\neg z)}{P(x\cap y\cap\neg z)+P(\neg x\cap y\cap\neg z)}=\frac{.2125}{.1125+.2125}=.68$$

$$P(x|\neg y\cap z)=\frac{P(x\cap\neg y\cap z)}{P(\neg y\cap z)}=\frac{P(x\cap\neg y\cap z)}{P(x\cap\neg y\cap z)+P(\neg x\cap\neg y\cap z)}=\frac{.3125}{.3125+.0125}=.9615$$

$$P(\neg x|\neg y\cap z)=\frac{P(\neg x\cap\neg y\cap z)}{P(\neg y\cap z)}=\frac{P(\neg x\cap\neg y\cap z)}{P(x\cap\neg y\cap z)+P(\neg x\cap\neg y\cap z)}=\frac{.0125}{.3125+.0125}=.03846$$

$$P(x|\neg y\cap\neg z)=\frac{P(x\cap\neg y\cap\neg z)}{P(\neg y\cap\neg z)}=\frac{P(x\cap\neg y\cap\neg z)}{P(x\cap\neg y\cap\neg z)+P(\neg x\cap\neg y\cap\neg z)}=\frac{.0125}{.0125+.1125}=.1$$

$$P(\neg x|\neg y\cap\neg z)=\frac{P(\neg x\cap\neg y\cap\neg z)}{P(\neg y\cap\neg z)}=\frac{P(\neg x\cap\neg y\cap\neg z)}{P(x\cap\neg y\cap\neg z)+P(\neg x\cap\neg y\cap\neg z)}=\frac{.1125}{.0125+.1125}=.9$$

Y	Z	$P(x\|y\cap z)$	$P(\neg x\|y\cap z)$	Prediction of X
T	T	.5	.5	F
T	F	.34615	.68	F
F	T	.96153	.03846	T
F	F	.1	.9	F

h)Use NB to predict cases for MAPC

calculations

$$P(x|y\cap z)=\frac{P(x\cap y\cap z)}{P(y\cap z)}=\frac{P(y\cap z|x)P(x)}{P(x\cap y\cap z)+P(\neg x\cap y\cap z)}=\frac{P(y|x)P(z|x)P(x)}{P(y\cap z|x)P(x)+P(y\cap z|\neg x)}=\frac{P(y|x)P(z|x)P(x)}{P(y|x)P(z|x)P(x)+P(y|\neg x)P(z|\neg x)P(x)}$$

$$P(x|y\cap z)=\frac{P(y|x)P(z|x)P(x)}{P(y|x)P(z|x)P(x)+P(y|\neg x)P(z|\neg x)P(\neg x)}=\frac{.40*.77*.52}{.40*.77*.52+.73*.28*.48}=\frac{.16}{.27}=.62$$

$$P(\neg x|y\cap z)=\frac{P(y|\neg x)P(z|\neg x)P(\neg x)}{P(y|\neg x)P(z|\neg x)P(\neg x)+P(y|x)P(z|x)P(x)}=\frac{.73*.28*.48}{.40*.77*.52+.73*.28*.48}=\frac{.10}{.27}=.38$$

$$P(x|y\cap\neg z)=\frac{P(y|x)P(\neg z|x)P(x)}{P(y|x)P(\neg z|x)P(x)+P(y|\neg x)P(\neg z|\neg x)P(\neg x)}=\frac{.40*.23*.52}{.40*.23*.52+.73*.72*.48}=\frac{.05}{.30}=.17$$

$$P(\neg x|y\cap\neg z)=\frac{P(y|\neg x)P(\neg z|\neg x)P(\neg x)}{P(y|x)P(\neg z|x)P(x)+P(y|\neg x)P(\neg z|\neg x)P(\neg x)}=\frac{.73*.72*.48}{.40*.23*.52+.73*.72*.48}=\frac{.25}{.30}=.83$$

$$P(x|\neg y \cap z) = \frac{P(\neg y|x)P(z|x)P(x)}{P(\neg y|x)P(z|x)P(x) + P(\neg y|\neg x)P(z|\neg x)P(\neg x)} = \frac{.59*.77*.52}{.59*.77*.52 + .28*.28*.48} = \frac{.24}{.28} = .86$$

$$P(\neg x|\neg y \cap z) = \frac{P(\neg y|\neg x)P(z|\neg x)P(\neg x)}{P(\neg y|x)P(z|x)P(x) + P(\neg y|\neg x)P(z|\neg x)P(\neg x)} = \frac{.72*.28*.48}{.59*.77*.52 + .28*.28*.48} = \frac{.10}{.49} = .14$$

$$P(x|\neg y \cap \neg z) == \frac{P(\neg y|x)P(\neg z|x)P(x)}{P(\neg y|x)P(\neg z|x)P(x) + P(\neg y|\neg x)P(\neg z|\neg x)P(\neg x)} = \frac{.59*.23*.52}{.59*.23*.52 + .28*.72*.48} = \frac{.07}{.17} = .41$$

$$P(\neg x|\neg y \cap \neg z) = \frac{P(\neg y|\neg x)P(\neg z|\neg x)P(\neg x)}{P(\neg y|x)P(\neg z|x)P(x) + P(\neg y|\neg x)P(\neg z|\neg x)P(\neg x)} = \frac{.28*.72*.48}{.59*.23*.52 + .28*.72*.48} = \frac{.10}{.17} = .59$$

Y	Z	$P(x\|y \cap z)$	$P(\neg x\|y \cap z)$	Prediction of X
T	T	.62	.38	T
T	F	.17	.83	F
F	T	.86	.14	T
F	F	.41	.59	F