Bayes Classification

- Uncertainty & Probability
- Baye's rule
- Choosing Hypotheses- Maximum a posteriori
- Maximum Likelihood - Baye's concept learning
- Maximum Likelihood of real valued function
- Bayes optimal Classifier
- Joint distributions
- Naive Bayes Classifier

Uncertainty

- Our main tool is the probability theory, which assigns to each sentence numerical degree of belief between *0* and *1*
- It provides a way of summarizing the uncertainty

Variables

- Boolean random variables: cavity might be true or false
- Discrete random variables: weather might be sunny, rainy, cloudy, snow
  - *P(Weather=sunny)*
  - *P(Weather=rainy)*
  - *P(Weather=cloudy)*
  - *P(Weather=snow)*
- Continuous random variables: the temperature has continuous values

Where do probabilities come from?

- Frequents:
  - From experiments: form any finite sample, we can estimate the true fraction and also calculate how accurate our estimation is likely to be
- Subjective:
  - Agent’s believe
- Objectivist:
  - True nature of the universe, that the probability up heads with probability 0.5 is a probability of the coin

- Before the evidence is obtained; prior probability
  - $P(a)$ the prior probability that the proposition is true
  - $P(cavity)=0.1$
- After the evidence is obtained; posterior probability
  - $P(a|b)$
  - The probability of a given that all we know is b
  - $P(cavity|toothache)=0.8$

Axioms of Probability

(Kolmogorov's axioms, first published in German 1933)

- All probabilities are between 0 and 1. For any proposition a $\quad 0 \leq P(a) \leq 1$
- $P(true)=1, \quad P(false)=0$
- The probability of disjunction is given by

$$P(a \vee b) = P(a) + P(b) - P(a \wedge b)$$

- Product rule

$$P(a \wedge b) = P(a \mid b)P(b)$$

$$P(a \wedge b) = P(b \mid a)P(a)$$

Theorem of total probability

If events A_1, ... , A_n are mutually

exclusive with $\sum_{i=1}^{n} P(A_i) = 1$
then

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i)$$

$$P(B) = \sum_{i=1}^{n} P(B, A_i)$$

Bayes’s rule

- (Reverent Thomas Bayes 1702-1761)
  - He set down his findings on probability in "Essay Towards Solving a Problem in the Doctrine of Chances" (1763), published posthumously in the *Philosophical Transactions of the Royal Society of London*

$$P(b \mid a) = \frac{P(a \mid b)P(b)}{P(a)}$$

Diagnosis

- What is the probability of meningitis in the patient with stiff neck?

  - A doctor knows that the disease meningitis causes the patient to have a stiff neck in 50% of the time -> *P(s|m)*

  - Prior Probabilities:
    - That the patient has meningitis is 1/50.000 -> *P(m)*
    - That the patient has a stiff neck is 1/20 -> *P(s)*

$$P(m \mid s) = \frac{P(s \mid m)P(m)}{P(s)}$$

$$P(m \mid s) = \frac{0.5 * 0.00002}{0.05} = 0.0002$$

Normalization

$$P(y \mid x) = \frac{P(x \mid y)P(y)}{P(x)}$$

$$P(\neg y \mid x) = \frac{P(x \mid \neg y)P(\neg y)}{P(x)}$$

$$1 = P(y \mid x) + P(\neg y \mid x)$$

$$P(Y \mid X) = \alpha \times P(X \mid Y) P(Y)$$

$$\alpha \left\langle P(y \mid x), P(\neg y \mid x) \right\rangle$$

$$\alpha \left\langle 0.12, 0.08 \right\rangle = \left\langle 0.6, 0.4 \right\rangle$$

Bayes Theorem

$$P(h|D) = \frac{P(D|h)P(h)}{P(D)}$$

- $P(h)$ = prior probability of hypothesis h
- $P(D)$ = prior probability of training data D
- $P(h|D)$ = probability of h given D
- $P(D|h)$ = probability of D given h

Example

- Does patient have cancer or not?

A patient takes a lab test and the result comes back positive. The test returns a correct positive result (+) in only 98% of the cases in which the disease is actually present, and a correct negative result (-) in only 97% of the cases in which the disease is not present

Furthermore, 0.008 of the entire population have this cancer

Suppose a positive result (+) is returned...

$$P(cancer) = 0.008 \qquad P(\neg cancer) = 0.992$$

$$P(+|cancer) = 0.98 \qquad P(-|cancer) = 0.02$$

$$P(+|\neg cancer) = 0.03 \qquad P(-|\neg cancer) = 0.97$$

$$P(+|cancer) \cdot P(cancer) = 0.98 \cdot 0.008 = 0.0078$$

$$P(+|\neg cancer) \cdot P(\neg cancer) = 0.03 \cdot 0.992 = 0.0298$$

Normalization

$$\frac{0.0078}{0.0078+0.0298} = 0.20745 \qquad \frac{0.0298}{0.0078+0.0298} = 0.79255$$

- The result of Bayesian inference depends strongly on the prior probabilities, which must be available in order to apply the method

Naive Bayes Classifier

- Along with decision trees, neural networks, nearest nbr, one of the most practical learning methods
- When to use:
  - Moderate or large training set available
  - Attributes that describe instances are conditionally independent given classification
- Successful applications:
  - Diagnosis
  - Classifying text documents

Training dataset

Class:
C1:buys_computer='yes'
C2:buys_computer='no'

Data sample:

X = (age<=30, Income=medium, Student=yes

Credit_rating=Fair)

age	income	student	credit_rating	buys_computer
<=30	high	no	fair	no
<=30	high	no	excellent	no
30…40	high	no	fair	yes
>40	medium	no	fair	yes
>40	low	yes	fair	yes
>40	low	yes	excellent	no
31…40	low	yes	excellent	yes
<=30	medium	no	fair	no
<=30	low	yes	fair	yes
>40	medium	yes	fair	yes
<=30	medium	yes	excellent	yes
31…40	medium	no	excellent	yes
31…40	high	yes	fair	yes
>40	medium	no	excellent	no

Naïve Bayesian Classifier: Example

- Compute $P(X|C_i)$ for each class

P(buys_computer=„yes“)=9/14

P(buys_computer=„no“)=5/14

P(age="<30" | buys_computer="yes") = 2/9=0.222
P(age="<30" | buys_computer="no") = 3/5 =0.6
P(income="medium" | buys_computer="yes")= 4/9 =0.444
P(income="medium" | buys_computer="no") = 2/5 = 0.4
P(student="yes" | buys_computer="yes)= 6/9 =0.667
P(student="yes" | buys_computer="no")= 1/5=0.2
P(credit_rating="fair" | buys_computer="yes")=6/9=0.667
P(credit_rating="fair" | buys_computer="no")=2/5=0.4

- X=(age<=30 ,income =medium, student=yes,credit_rating=fair)

$P(X|C_i)$: P(X|buys_computer="yes")= 0.222 x 0.444 x 0.667 x 0.0.667 =0.044
P(X|buys_computer="no")= 0.6 x 0.4 x 0.2 x 0.4 =0.019

$P(X|C_i)*P(C_i)$: P(X|buys_computer="yes") * P(buys_computer="yes")=0.028
P(X|buys_computer="no") * P(buys_computer="no")=0.007

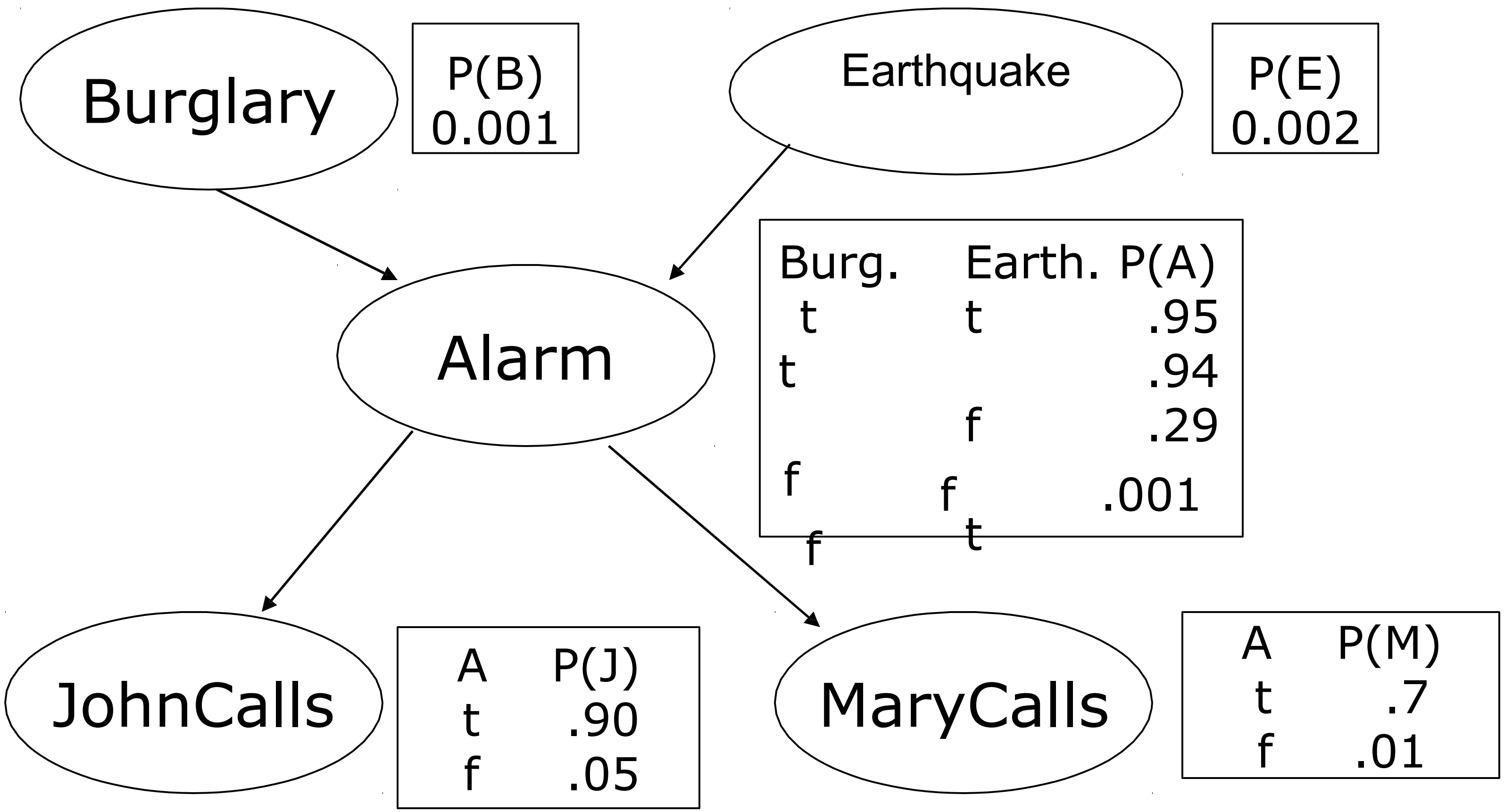
- X belongs to class "buys_computer=yes"

Naïve Bayesian Classifier: Comments

- Advantages :
  - Easy to implement
  - Good results obtained in most of the cases
- Disadvantages
  - Assumption: class conditional independence , therefore loss of accuracy
  - Practically, dependencies exist among variables
  - E.g., hospitals: patients: Profile: age, family history etc
    Symptoms: fever, cough etc., Disease: lung cancer, diabetes etc
  - Dependencies among these cannot be modeled by Naïve Bayesian Classifier
- How to deal with these dependencies?
  - Bayesian Belief Networks

- Uncertainty & Probability
- Baye's rule
- Choosing Hypotheses- Maximum a posteriori
- Maximum Likelihood - Baye's concept learning
- Maximum Likelihood of real valued function
- Bayes optimal Classifier
- Joint distributions
- Naive Bayes Classifier

Bayesian Belief Networks



Thank you !!!!

Any Questions ????

Utkarsh Kulshrestha

(kuls.utkarsh1205@gmail.com)