CSIT5210

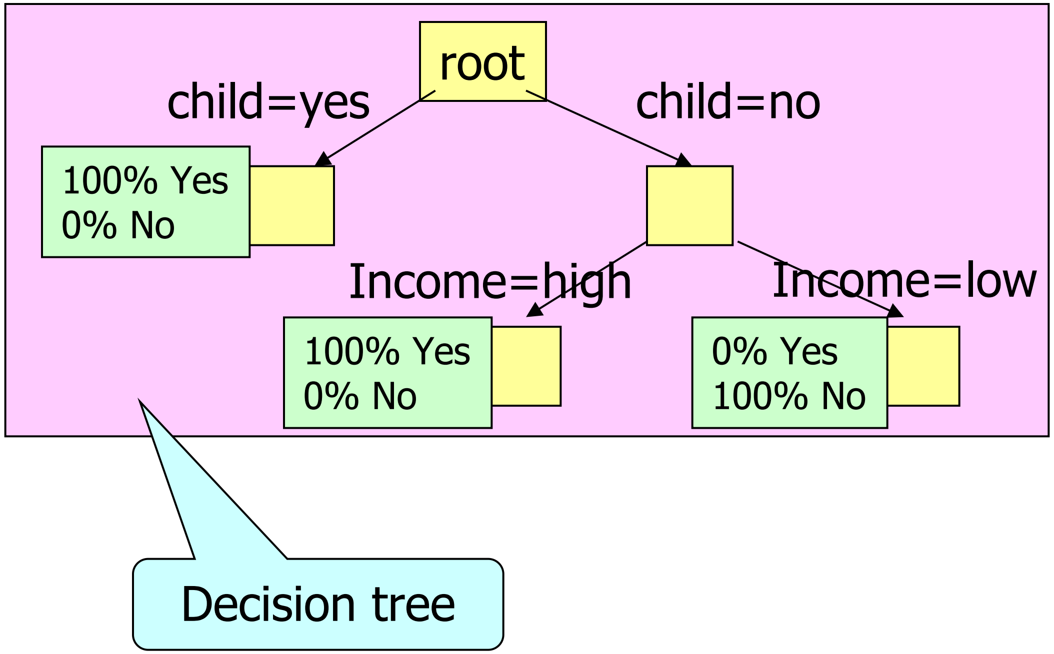
Classification

Prepared by Raymond Wong
The examples used in Decision Tree are borrowed from LW Chan's notes
Presented by Raymond Wong
raywong@cse

Classification

Suppose there is a person.

Race	Income	Child	Insurance
white	high	no	?



Classification

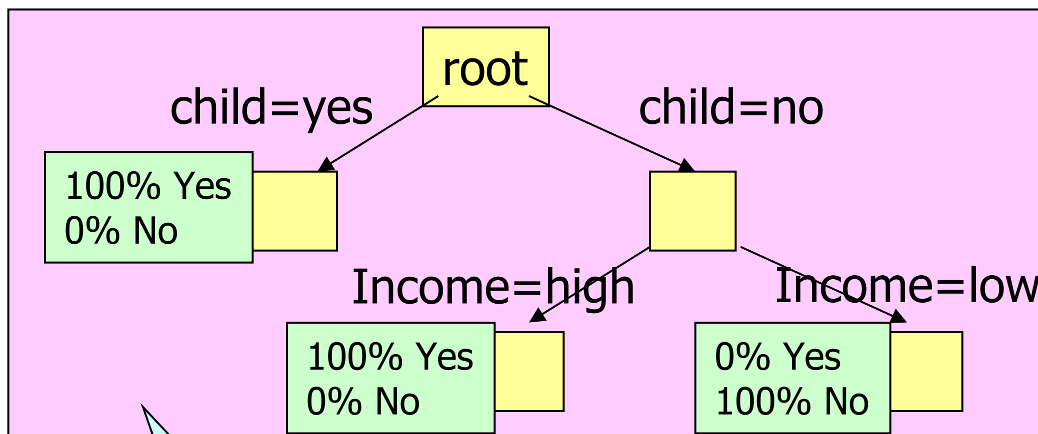
Suppose there is a person.

Test set

Race	Income	Child	Insurance
white	high	no	?

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

Training set



Decision tree

Applications

- Insurance
  - According to the attributes of customers,
    - Determine which customers will buy an insurance policy
- Marketing
  - According to the attributes of customers,
    - Determine which customers will buy a product such as computers
- Bank Loan
  - According to the attributes of customers,
    - Determine which customers are “risky” customers or “safe” customers

Applications

- Network
  - According to the traffic patterns,
    - Determine whether the patterns are related to some "security attacks"
- Software
  - According to the experience of programmers,
    - Determine which programmers can fix some certain bugs

Same/Difference

- Classification
- Clustering

Classification Methods

- Decision Tree
- Bayesian Classifier
- Nearest Neighbor Classifier

Decision Trees

- ID3
- C4.5
- CART

Iterative Dichotomiser

Classification

Classification And Regression Trees

Entropy

- Example 1
  - Consider a random variable which has a uniform distribution over 32 outcomes
  - To identify an outcome, we need a label that takes 32 different values.
  - Thus, 5 bit strings suffice as labels

Entropy

- **Entropy** is used to measure how informative is a node.
- If we are given a probability distribution $P = (p_1, p_2, \ldots, p_n)$ then the **Information** conveyed by this distribution, also called the **Entropy** of P, is:
  $I(P) = - (p_1 \times \log p_1 + p_2 \times \log p_2 + \ldots + p_n \times \log p_n)$
- All logarithms here are in base 2.

Entropy

- For example,
  - If P is (0.5, 0.5), then I(P) is 1.
  - If P is (0.67, 0.33), then I(P) is 0.92,
  - If P is (1, 0), then I(P) is 0.
- The **entropy** is a way to measure the amount of information.
- The smaller the entropy, the more informative we have.

Entropy

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

Info(T) = - ½ log ½ - ½ log ½
= 1

For attribute Race,

$\text{Info}(T_{black}) = -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} = 0.8113$

$\text{Info}(T_{white}) = -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} = 0.8113$

$\text{Info}(\text{Race}, T) = \frac{1}{2} \times \text{Info}(T_{black}) + \frac{1}{2} \times \text{Info}(T_{white}) = 0.8113$

Gain(Race, T) = Info(T) – Info(Race, T) = 1 – 0.8113 = 0.1887

For attribute Race, Gain(Race, T) = 0.1887

Entropy

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

$\text{Info}(T) = -\frac{1}{2} \log \frac{1}{2} - \frac{1}{2} \log \frac{1}{2}$
$= 1$

For attribute Income,

$\text{Info}(T_{high}) = -1 \log 1 - 0 \log 0 = 0$

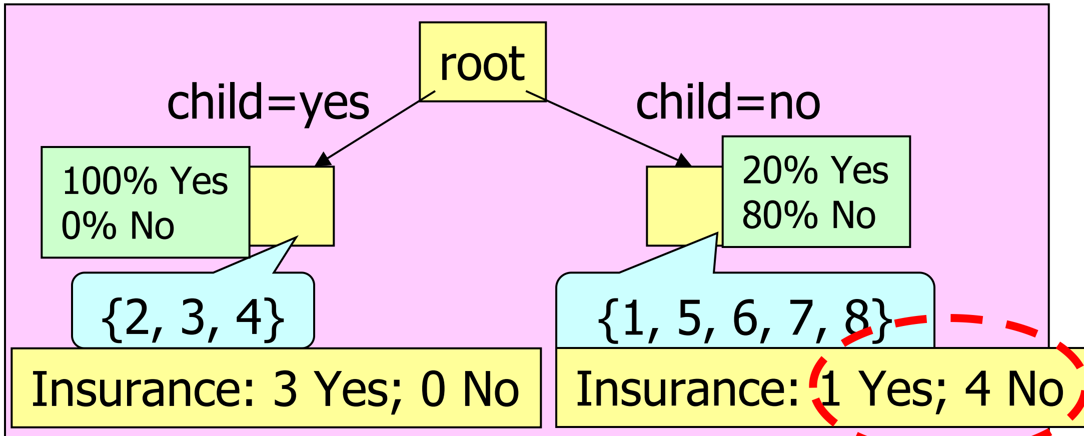
$\text{Info}(T_{low}) = -1/3 \log 1/3 - 2/3 \log 2/3 = 0.9183$

$\text{Info}(\text{Income}, T) = \frac{1}{4} \times \text{Info}(T_{high}) + \frac{3}{4} \times \text{Info}(T_{low}) = 0.6887$

$\text{Gain}(\text{Income}, T) = \text{Info}(T) - \text{Info}(\text{Income}, T) = 1 - 0.6887 = 0.3113$

For attribute Race, Gain(Race, T) = 0.1887

For attribute Income, Gain(Income, T) = 0.3113



	Race	Income	Child	Insurance
1	black	high	no	yes
2	white	high	yes	yes
3	white	low	yes	yes
4	white	low	yes	yes
5	black	low	no	no
6	black	low	no	no
7	black	low	no	no
8	white	low	no	no

$$\text{Info}(T) = -\tfrac{1}{2}\log\tfrac{1}{2} - \tfrac{1}{2}\log\tfrac{1}{2} = 1$$

For attribute Child,

$$\text{Info}(T_{yes}) = -1\log 1 - 0\log 0 = 0$$

$$\text{Info}(T_{no}) = -1/5\log 1/5 - 4/5\log 4/5 = 0.7219$$

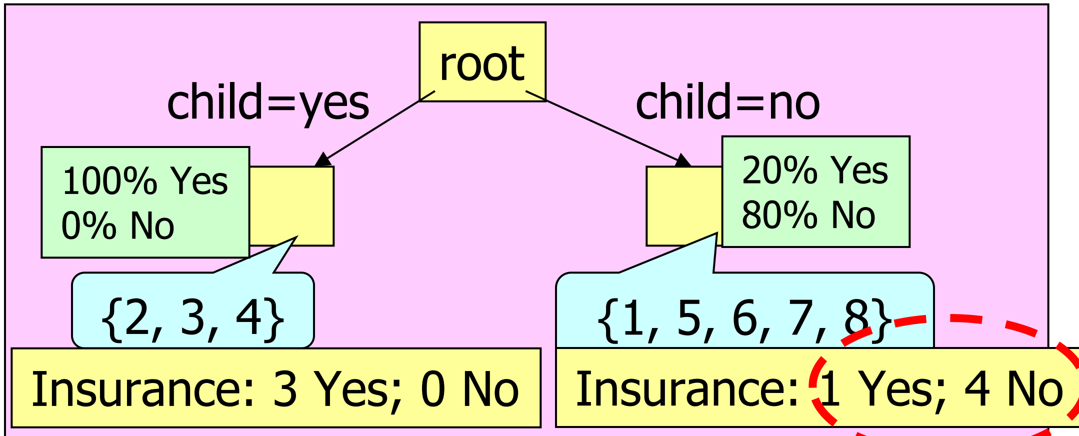
$$\text{Info}(\text{Child}, T) = 3/8 \times \text{Info}(T_{yes}) + 5/8 \times \text{Info}(T_{no}) = 0.4512$$

$$\text{Gain}(\text{Child}, T) = \text{Info}(T) - \text{Info}(\text{Child}, T) = 1 - 0.4512 = 0.5488$$

For attribute Race, Gain(Race, T) = 0.1887

For attribute Income, Gain(Income, T) = 0.3113

For attribute Child, Gain(Child, T) = 0.5488



	Race	Income	Child	Insurance
1	black	high	no	yes
2	white	high	yes	yes
3	white	low	yes	yes
4	white	low	yes	yes
5	black	low	no	no
6	black	low	no	no
7	black	low	no	no
8	white	low	no	no

$$\text{Info}(T) = - 1/5 \log 1/5 - 4/5 \log 4/5$$
$$= 0.7219$$

For attribute Race,

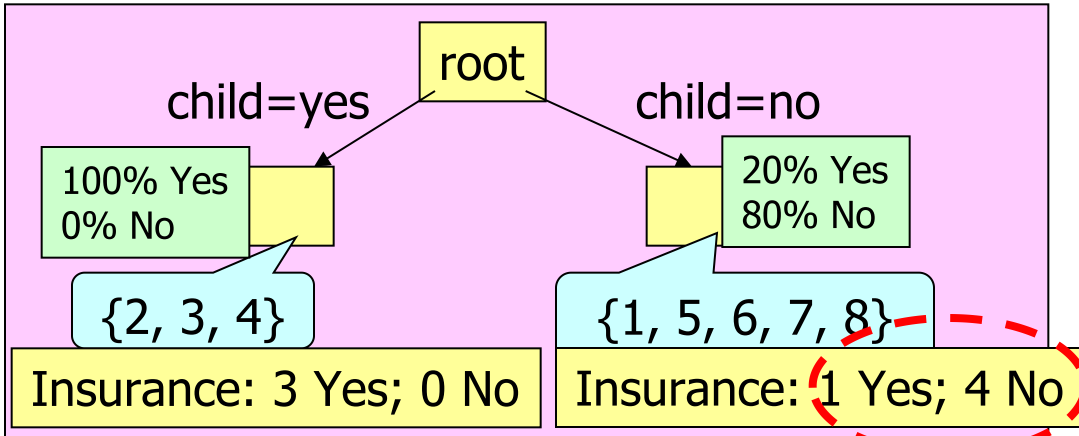
$$\text{Info}(T_{black}) = - ¼ \log ¼ - ¾ \log ¾ \quad = 0.8113$$

$$\text{Info}(T_{white}) = - 0 \log 0 - 1 \log 1 \ = 0$$

$$\text{Info}(\text{Race}, T) = 4/5 \times \text{Info}(T_{black}) + 1/5 \times \text{Info}(T_{white}) \quad = 0.6490$$

$$\text{Gain}(\text{Race}, T) = \text{Info}(T) - \text{Info}(\text{Race}, T) = 0.7219 - 0.6490 \quad = 0.0729$$

For attribute Race, Gain(Race, T) = 0.0729



	Race	Income	Child	Insurance
1	black	high	no	yes
2	white	high	yes	yes
3	white	low	yes	yes
4	white	low	yes	yes
5	black	low	no	no
6	black	low	no	no
7	black	low	no	no
8	white	low	no	no

$$\text{Info}(T) = -\ 1/5 \log 1/5 - 4/5 \log 4/5$$
$$= 0.7219$$

For attribute Income,

$$\text{Info}(T_{high}) = -\ 1 \log 1 - 0 \log 0 \quad = 0$$

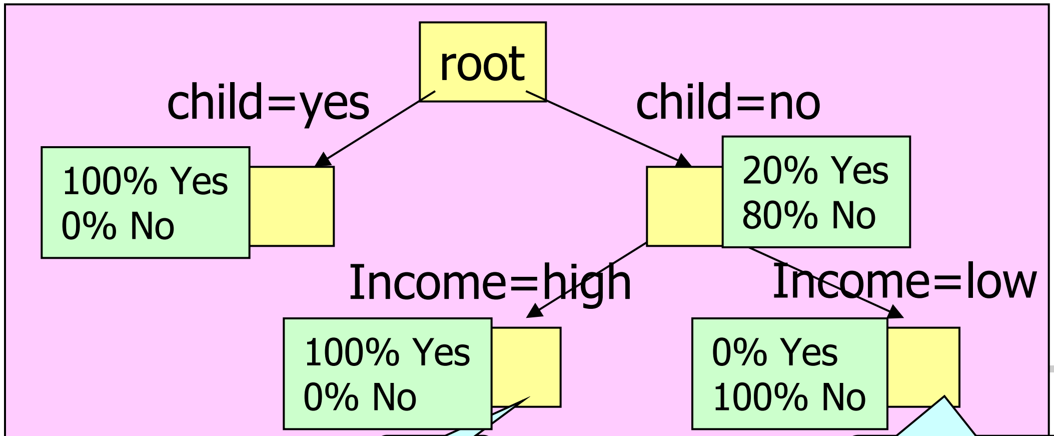
$$\text{Info}(T_{low}) = -\ 0 \log 0 - 1 \log 1 \quad = 0$$

$$\text{Info}(\text{Income}, T) = 1/5 \times \text{Info}(T_{high}) + 4/5 \times \text{Info}(T_{low}) = 0$$

$$\text{Gain}(\text{Income}, T) = \text{Info}(T) - \text{Info}(\text{Income}, T) = 0.7219 - 0 = 0.7219$$

For attribute Race, Gain(Race, T) = 0.0729

For attribute Income, Gain(Income, T) = 0.7219



	Race	Income	Child	Insurance
1	black	high	no	yes
2	white	high	yes	yes
3	white	low	yes	yes
4	white	low	yes	yes
5	black	low	no	no
6	[illegible]	low	no	no
7	black	low	no	no
8	white	low	no	no

Info(T) = - 1/5 ... /5 – 4/5 log 4/5

{1}

{5, 6, 7, 8}

Insurance: 1 Yes; 0 No

Insurance: 0 Yes; 4 No

For attribute Income,

Info(T_{high}) = - 1 log 1 – 0 log 0 = 0

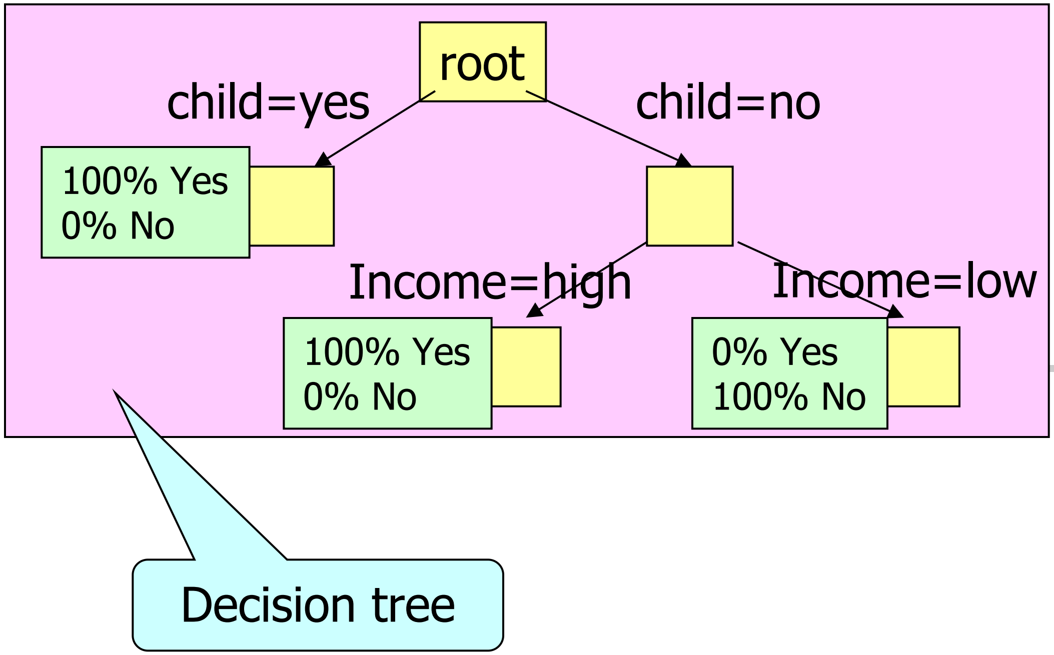
Info(T_{low}) = - 0 log 0 – 1 log 1 = 0

Info(Income, T) = 1/5 x Info(T_{high}) + 4/5 x Info(T_{low}) = 0

Gain(Income, T) = Info(T) – Info(Income, T) = 0.7219 – 0 = 0.7219

For attribute Race, Gain(Race, T) = 0.0729

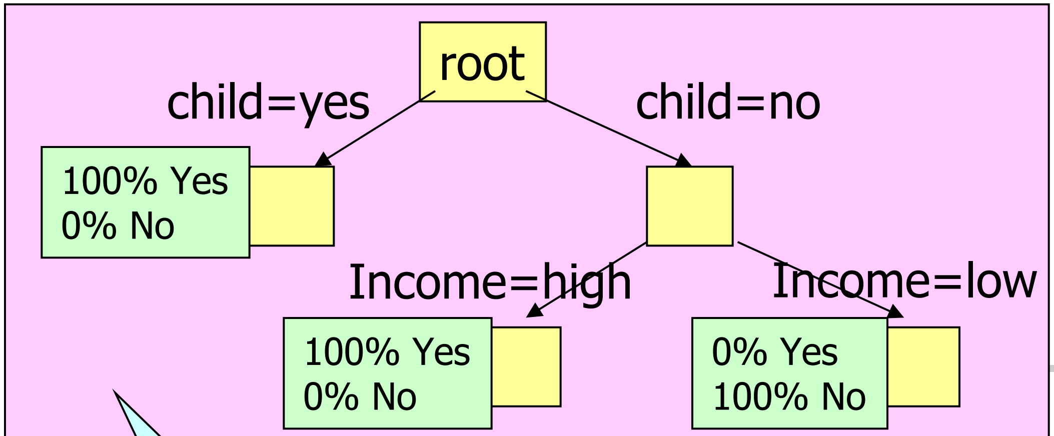
For attribute Income, Gain(Income, T) = 0.7219



	Race	Income	Child	Insurance
1	black	high	no	yes
2	white	high	yes	yes
3	white	low	yes	yes
4	white	low	yes	yes
5	black	low	no	no
6	black	low	no	no
7	black	low	no	no
8	white	low	no	no

Suppose there is a new person.

Race	Income	Child	Insurance
white	high	no	?



Decision tree

	Race	Income	Child	Insurance
1	black	high	no	yes
2	white	high	yes	yes
3	white	low	yes	yes
4	white	low	yes	yes
5	black	low	no	no
6	black	low	no	no
7	black	low	no	no
8	white	low	no	no

Termination Criteria?

e.g., height of the tree
e.g., accuracy of each node

Decision Trees

- ID3
- C4.5
- CART

C4.5

- ID3
  - Impurity Measurement
    - Gain(A, T)
      = Info(T) – Info(A, T)
- C4.5
  - Impurity Measurement
    - Gain(A, T)
      = (Info(T) – Info(A, T))/SplitInfo(A)
    - where SplitInfo(A) = $-\sum_{v \in A} p(v) \log p(v)$

Entropy

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

$$Info(T) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$$

For attribute Race,

$$Info(T_{black}) = -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} = 0.8113$$

$$Info(T_{white}) = -\frac{3}{4}\log\frac{3}{4} - \frac{1}{4}\log\frac{1}{4} = 0.8113$$

$$Info(Race, T) = \frac{1}{2} \times Info(T_{black}) + \frac{1}{2} \times Info(T_{white}) = 0.8113$$

$$SplitInfo(Race) = -\frac{1}{2}\log\frac{1}{2} - \frac{1}{2}\log\frac{1}{2} = 1$$

$$Gain(Race, T) = (Info(T) - Info(Race, T))/SplitInfo(Race) = (1 - 0.8113)/1 = 0.1887$$

For attribute Race,

Gain(Race, T) = 0.1887

Entropy

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

Info(T) = - ½ log ½ - ½ log ½
= 1

For attribute Income,

$\text{Info}(T_{high}) = - 1 \log 1 – 0 \log 0 = 0$

$\text{Info}(T_{low}) = - 1/3 \log 1/3 – 2/3 \log 2/3 = 0.9183$

$\text{Info}(\text{Income}, T) = ¼ \times \text{Info}(T_{high}) + ¾ \times \text{Info}(T_{low}) = 0.6887$

SplitInfo(Income) = - 2/8 log 2/8 – 6/8 log 6/8 = 0.8113

Gain(Income, T)= (Info(T)–Info(Income, T))/SplitInfo(Income) = (1–0.6887)/0.8113
= 0.3837

For attribute Race, Gain(Race, T) = 0.1887

For attribute Income, Gain(Income, T) = 0.3837

For attribute Child, Gain(Child, T) = ?

Decision Trees

- ID3
- C4.5
- CART

CART

- Impurity Measurement
  - Gini
    $I(P) = 1 - \sum_j p_j^2$

Gini

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

$\text{Info}(T) = 1 - (\tfrac{1}{2})^2 - (\tfrac{1}{2})^2$
$= \tfrac{1}{2}$

For attribute Race,

$\text{Info}(T_{black}) = 1 - (\tfrac{3}{4})^2 - (\tfrac{1}{4})^2 \quad = 0.375$

$\text{Info}(T_{white}) = 1 - (\tfrac{3}{4})^2 - (\tfrac{1}{4})^2 \quad = 0.375$

$\text{Info}(\text{Race}, T) = \tfrac{1}{2} \times \text{Info}(T_{black}) + \tfrac{1}{2} \times \text{Info}(T_{white}) \quad = 0.375$

$\text{Gain}(\text{Race}, T) = \text{Info}(T) - \text{Info}(\text{Race}, T) = \tfrac{1}{2} - 0.375 \quad = 0.125$

For attribute Race, Gain(Race, T) = 0.125

Gini

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

$$Info(T) = 1 – (½)^2 – (½)^2$$
$$= ½$$

For attribute Income,

$$Info(T_{high}) = 1 – 1^2 – 0^2 \;\; = 0$$

$$Info(T_{low}) = 1 – (1/3)^2 – (2/3)^2 \;\; = 0.444$$

$$Info(Income, T) = 1/4 \text{ x } Info(T_{high}) + 3/4 \text{ x } Info(T_{low}) \;\; = 0.333$$

$$Gain(Income, T) = Info(T) – Info(Race, T) \;\; = ½ \; – 0.333 \;\; = 0.167$$

For attribute Race, Gain(Race, T) = 0.125

For attribute Income, Gain(Race, T) = 0.167

For attribute Child, Gain(Child, T) = ?

Classification Methods

- Decision Tree
- Bayesian Classifier
- Nearest Neighbor Classifier

Bayesian Classifier

- Naïve Bayes Classifier
- Bayesian Belief Networks

Naïve Bayes Classifier

- Statistical Classifiers
- Probabilities
- Conditional probabilities

Naïve Bayes Classifier

- Conditional Probability
  - A: a random variable
  - B: a random variable
  - $$P(A \mid B) = \frac{P(AB)}{P(B)}$$

Naïve Bayes Classifier

- Bayes Rule
  - A : a random variable
  - B: a random variable
  - $$P(A \mid B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Naïve Bayes Class

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

- Independent Assumption
  - Each attribute are independent
  - e.g.,
    $P(X, Y, Z \mid A) = P(X \mid A) \times P(Y \mid A) \times P(Z \mid A)$

Suppose there is a new person.

Race	Income	Child	Insurance
white	high	no	?

Naïve Bayes Class

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

Insurance = Yes

For attribute Race,

P(Race = black | Yes) = ¼
P(Race = white | Yes) = ¾
P(Race = black | No) = ¾
P(Race = white | No) = ¼

P(Yes) = ½

P(No) = ½

For attribute Income,

P(Income = high | Yes) = ½
P(Income = low | Yes) = ½
P(Income = high | No) = 0
P(Income = low | No) = 1

For attribute Child,

P(Child = yes | Yes) = ¾
P(Child = no | Yes) = ¼
P(Child = yes | No) = 0
P(Child = no | No) = 1

Naïve Bayes Classifier

P(Race = white, Income = high, Child = no| Yes)
= P(Race = white | Yes) x P(Income = high | Yes)
x P(Child = no | Yes)
= ¾ x ½ x ¼
= 0.09375

P(Race = white, Income = high, Child = no| No)
= P(Race = white | No) x P(Income = high | No)
x P(Child = no | No)
= ¼ x 0 x 1
= 0

Suppose there is a new person.

Race	Income	Child	Insurance
white	high	no	?

Naïve Bayes Classifier

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

Insurance = Yes

For attribute Race,

P(Race = black | Yes) = ¼
P(Race = white | Yes) = ¾
P(Race = black | No) = ¾
P(Race = white | No) = ¼

P(Yes) = ½

P(No) = ½

For attribute Income,

P(Income = high | Yes) = ½
P(Income = low | Yes) = ½
P(Income = high | No) = 0
P(Income = low | No) = 1

For attribute Child,

P(Child = yes | Yes) = ¾
P(Child = no | Yes) = ¼
P(Child = yes | No) = 0
P(Child = no | No) = 1

Naïve Bayes Classifier

P(Race = white, Income = high, Child = no| Yes)

= 0.09375

P(Race = white, Income = high, Child = no| No)
= P(Race = white | No) x P(Income = high | No)
x P(Child = no | No)
= ¼ x 0 x 1
= 0

Naïve Bayes Class

Suppose there is a new person.

Race	Income	Child	Insurance
white	high	no	?

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

Insurance = Yes

For attribute Race,

P(Race = black | Yes) = ¼
P(Race = white | Yes) = ¾
P(Race = black | No) = ¾
P(Race = white | No) = ¼

For attribute Income,

P(Income = high | Yes) = ½
P(Income = low | Yes) = ½
P(Income = high | No) = 0
P(Income = low | No) = 1

For attribute Child,

P(Child = yes | Yes) = ¾
P(Child = no | Yes) = ¼
P(Child = yes | No) = 0
P(Child = no | No) = 1

P(Yes) = ½

P(No) = ½

Naïve Bayes Classifier

P(Race = white, Income = high, Child = no| Yes)
= 0.09375

P(Race = white, Income = high, Child = no| No)
= P(Race = white | No) x P(Income = high | No)
x P(Child = no | No)
= ¼ x 0 x 1
= 0

Suppose there is a new person.

Race	Income	Child	Insurance
white	high	no	?

Naïve Bayes Classifier

Insurance = Yes

For attribute Race,

P(Race = black | Yes) = ¼
P(Race = white | Yes) = ¾
P(Race = black | No) = ¾
P(Race = white | No) = ¼

P(Yes) = ½

P(No) = ½

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

For attribute Income,

P(Income = high | Yes) = ½
P(Income = low | Yes) = ½
P(Income = high | No) = 0
P(Income = low | No) = 1

Naïve Bayes Classifier

P(Race = white, Income = high, Child = no| Yes)
= 0.09375

For attribute Child,

P(Child = yes | Yes) = ¾
P(Child = no | Yes) = ¼
P(Child = yes | No) = 0
P(Child = no | No) = 1

P(Race = white, Income = high, Child = no| No)

= 0

Naïve Bayes Classifier

Suppose there is a new person.

Race	Income	Child	Insurance
white	high	no	?

Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

Insurance = Yes

For attribute Race,

P(Race = black | Yes) = ¼
P(Race = white | Yes) = ¾
P(Race = black | No) = ¾
P(Race = white | No) = ¼

P(Yes) = ½

P(No) = ½

For attribute Income,

P(Income = high | Yes) = ½
P(Income = low | Yes) = ½
P(Income = high | No) = 0
P(Income = low | No) = 1

For attribute Child,

P(Child = yes | Yes) = ¾
P(Child = no | Yes) = ¼
P(Child = yes | No) = 0
P(Child = no | No) = 1

Naïve Bayes Classifier

P(Race = white, Income = high, Child = no| Yes)
= 0.09375

P(Race = white, Income = high, Child = no| No)
= 0

Suppose there is a new person.

Race	Income	Child	Insu…
white	high	no	?

$$P(\text{Yes} \mid \text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no}) = \frac{0.046875}{P(\text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no})}$$

Naïve Bayes Class…

white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

Insurance = Yes

For attribute Race,

$P(\text{Race} = \text{black} \mid \text{Yes}) = ¼$

$P(\text{Race} = \text{white} \mid \text{Yes}) = ¾$

$P(\text{Race} = \text{black} \mid \text{No}) = ¾$

$P(\text{Race} = \text{white} \mid \text{No}) = ¼$

$P(\text{Yes}) = ½$

$P(\text{No}) = ½$

For attribute Income,

$P(\text{Income} = \text{high} \mid \text{Yes}) = ½$

$P(\text{Income} = \text{low} \mid \text{Yes}) = ½$

$P(\text{Income} = \text{high} \mid \text{No}) = 0$

$P(\text{Income} = \text{low} \mid \text{No}) = 1$

For attribute Child,

$P(\text{Child} = \text{yes} \mid \text{Yes}) = ¾$

$P(\text{Child} = \text{no} \mid \text{Yes}) = ¼$

$P(\text{Child} = \text{yes} \mid \text{No}) = 0$

$P(\text{Child} = \text{no} \mid \text{No}) = 1$

Naïve Bayes Classifier

$P(\text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no} \mid \text{Yes})$
$= 0.09375$

$P(\text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no} \mid \text{No})$
$= 0$

$P(\text{Yes} \mid \text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no})$

$$= \frac{P(\text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no} \mid \text{Yes})\, P(\text{Yes})}{P(\text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no})}$$

$$= \frac{0.09375 \times 0.5}{P(\text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no})}$$

$$= \frac{0.046875}{P(\text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no})}$$

Suppose there is a new person.

Race	Income	Child	Ins
white	high	no	

Naïve Bay

Insuran

$$P(\text{Yes} \mid \text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no}) = \frac{0.046875}{P(\text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no})}$$

$$P(\text{No} \mid \text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no}) = \frac{0}{P(\text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no})}$$

white	high	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

For attribute Race,

P(Race = black | Yes) = ¼

P(Race = white | Yes) = ¾

P(Race = black | No) = ¾

P(Race = white | No) = ¼

P(Yes) = ½

P(No) = ½

For attribute Income,

P(Income = high | Yes) = ½

P(Income = low | Yes) = ½

P(Income = high | No) = 0

P(Income = low | No) = 1

For attribute Child,

P(Child = yes | Yes) = ¾

P(Child = no | Yes) = ¼

P(Child = yes | No) = 0

P(Child = no | No) = 1

Naïve Bayes Classifier

P(Race = white, Income = high, Child = no| Yes)
= 0.09375

P(Race = white, Income = high, Child = no| No)
= 0

P(No | Race = white, Income = high, Child = no)

$$= \frac{P(\text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no} \mid \text{No})\, P(\text{No})}{P(\text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no})}$$

$$= \frac{0 \times 0.5}{P(\text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no})}$$

$$= \frac{0}{P(\text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no})}$$

Suppose there is a new person.

Race	Income	Child	Ins
white	high	no	

Naïve Bay

Insuran

$$P(\text{Yes} \mid \text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no}) = \frac{0.046875}{P(\text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no})}$$

$$P(\text{No} \mid \text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no}) = \frac{0}{P(\text{Race} = \text{white}, \text{Income} = \text{high}, \text{Child} = \text{no})}$$

white	high	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

For attribute Race,

P(Race = black | Yes) = ¼

P(Race = white | Yes) = ¾

P(Race = black | No) = ¾

P(Race = white | No) = ¼

P(Yes) = ½

P(No) = ½

Naïve Bayes Classifier

For attribute Income,

P(Income = high | Y

P(Income = low | Y

P(Income = high | N

P(Income = low | N

For attribute Child,

P(Child = yes | Yes)

P(Child = no | Yes)

P(Child = yes | No)

P(Child = no | No) = 1

Since P(Yes | Race = white, Income = high, Child = no) > P(No | Race = white, Income = high, Child = no).

we predict the following new person will buy an insurance.

Race	Income	Child	Insurance
white	high	no	?

Bayesian Classifier

- Naïve Bayes Classifier
- Bayesian Belief Networks

Bayesian Belief Network

- Naïve Bayes Classifier
  - Independent Assumption
- Bayesian Belief Network
  - Do not have independent assumption

Yes/No | Healthy/Unhealthy | Yes/No | High/Low | Yes/No | Yes/No

Exercise	Diet	Heartburn	Blood Pressure	Chest Pain	Heart Disease
Yes	Healthy	No	High	Yes	No
No	Unhealthy	Yes	Low	Yes	No
No	Healthy	Yes	High	No	Yes
…	…	…	…	…	…

Some attributes are dependent on other attributes.

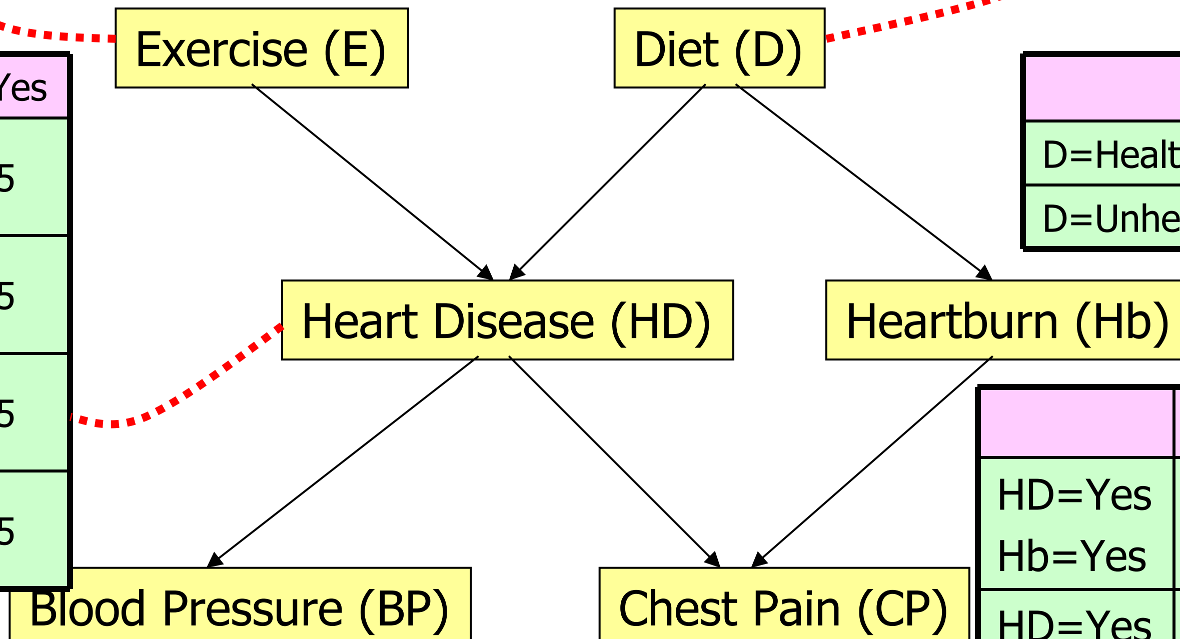
e.g., doing exercises may reduce the probability of suffering from Heart Disease

Exercise (E) → Heart Disease

Bayesian Belief Network

E = Yes
0.7

D = Healthy
0.25



	HD=Yes
E=Yes D=Healthy	0.25
E=Yes D=Unhealthy	0.45
E=No D=Healthy	0.55
E=No D=Unhealthy	0.75

	Hb=Yes
D=Healthy	0.85
D=Unhealthy	0.2

	BP=High
HD=Yes	0.85
HD=No	0.2

	CP=Yes
HD=Yes Hb=Yes	0.8
HD=Yes Hb=No	0.6
HD=No Hb=Yes	0.4
HD=No Hb=No	0.1

Let X, Y, Z be three random variables.
X is said to be **conditionally independent** of Y given Z if the following holds.

$$P(X \mid Y, Z) = P(X \mid Z)$$

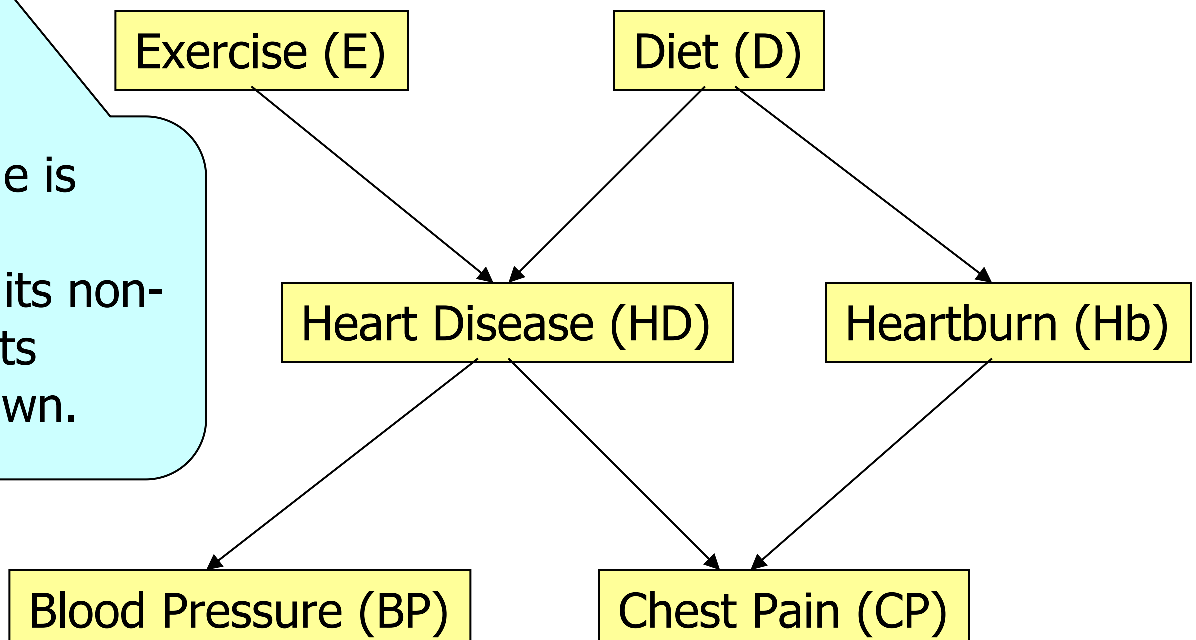
Lemma:
If X is conditionally independent of Y given Z,

$$P(X, Y \mid Z) = P(X \mid Z) \times P(Y \mid Z) ?$$

Let X, Y, Z be three random variables.
X is said to be **conditionally independent** of Y given Z if the following holds.

$$P(X \mid Y, Z) = P(X \mid Z)$$

Property: A node is **conditionally independent** of its non-descendants if its parents are known.



e.g., P(BP = High | HD = Yes, D = Healthy) = P(BP = High | HD = Yes)

“BP = High” is **conditionally independent** of “D = Healthy” given “HD = Yes”

e.g., P(BP = High | HD = Yes, CP=Yes) = P(BP = High | HD = Yes)

“BP = High” is **conditionally independent** of “CP = Yes” given “HD = Yes”

Yes/No

Healthy/Unhealthy

Yes/No

High/Low

Yes/No

Yes/No

Exercise	Diet	Heartburn	Blood Pressure	Chest Pain	Heart Disease
Yes	Healthy	No	High	Yes	No
No	Unhealthy	Yes	Low	Yes	No
No	Healthy	Yes	High	No	Yes
…	…	…	…	…	…

Suppose there is a new person and I want to know whether he is likely to have Heart Disease.

Exercise	Diet	Heartburn	Blood Pressure	Chest Pain	Heart Disease
?	?	?	?	?	?

Exercise	Diet	Heartburn	Blood Pressure	Chest Pain	Heart Disease
?	?	?	High	?	?

Exercise	Diet	Heartburn	Blood Pressure	Chest Pain	Heart Disease
Yes	Healthy	?	High	?	?

Suppose there is a new person and I want to know whether he is likely to have Heart Disease.

Exercise	Diet	Heartburn	Blood Pressure	Chest Pain	Heart Disease
?	?	?	?	?	?

$$P(HD = Yes) = \sum_{x\in\{Yes,\ No\}} \sum_{y\in\{Healthy,\ Unhealthy\}} P(HD=Yes|E=x,\ D=y) \times P(E=x,\ D=y)$$

$$= \sum_{x\in\{Yes,\ No\}} \sum_{y\in\{Healthy,\ Unhealthy\}} P(HD=Yes|E=x,\ D=y) \times P(E=x) \times P(D=y)$$

$$= 0.25 \times 0.7 \times 0.25 + 0.45 \times 0.7 \times 0.75 + 0.55 \times 0.3 \times 0.25 + 0.75 \times 0.3 \times 0.75$$

$$= 0.49$$

$$P(HD = No) = 1- P(HD = Yes)$$

$$= 1- 0.49$$

$$= 0.51$$

Suppose there is a new person and I want to know whether he is likely to have Heart Disease.

Exercise	Diet	Heartburn	Blood Pressure	Chest Pain	Heart Disease
?	?	?	High	?	?

$$P(BP = High) = \sum_{x \in \{Yes,\ No\}} P(BP = High|HD=x) \times P(HD = x)$$

$$= 0.85 \times 0.49 + 0.2 \times 0.51$$

$$= 0.5185$$

$$P(HD = Yes|BP = High) = \frac{P(BP = High|HD=Yes) \times P(HD = Yes)}{P(BP = High)}$$

$$= \frac{0.85 \times 0.49}{0.5185}$$

$$= 0.8033$$

$$P(HD = No|BP = High) = 1 - P(HD = Yes|BP = High)$$

$$= 1 - 0.8033$$

$$= 0.1967$$

Suppose there is a new person and I want to know whether he is likely to have Heart Disease.

Exercise	Diet	Heartburn	Blood Pressure	Chest Pain	Heart Disease
Yes	Healthy	?	High	?	?

$$P(HD = Yes \mid BP = High, D = Healthy, E = Yes)$$

$$= \frac{P(BP = High \mid HD = Yes, D = Healthy, E = Yes)}{P(BP = High \mid D = Healthy, E = Yes)} \times P(HD = Yes \mid D = Healthy, E = Yes)$$

$$= \frac{P(BP = High \mid HD = Yes)\, P(HD = Yes \mid D = Healthy, E = Yes)}{\sum_{x \in \{Yes, No\}} P(BP=High \mid HD=x)\, P(HD=x \mid D=Healthy, E=Yes)}$$

$$= \frac{0.85 \times 0.25}{0.85 \times 0.25 + 0.2 \times 0.75}$$

$$= 0.5862$$

$$P(HD = No \mid BP = High, D = Healthy, E = Yes)$$

$$= 1 - P(HD = Yes \mid BP = High, D = Healthy, E = Yes)$$

$$= 1 - 0.5862$$

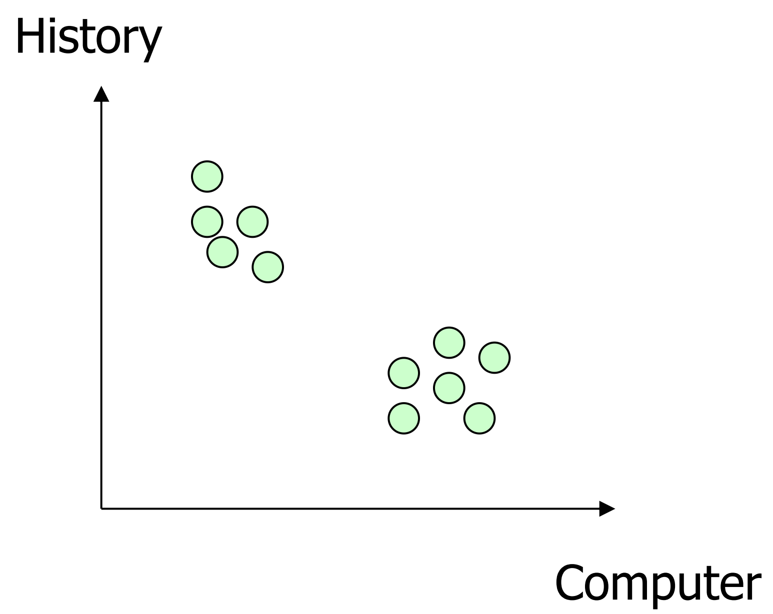
$$= 0.4138$$

Classification Methods

- Decision Tree
- Bayesian Classifier
- Nearest Neighbor Classifier

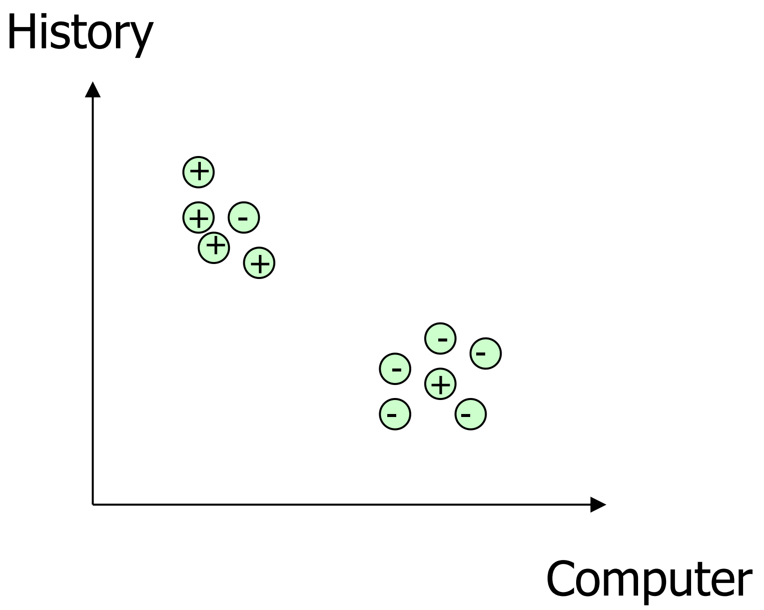
Nearest Neighbor Classifier

Computer	History
100	40
90	45
20	95
…	…



Nearest Neighbor Classifier

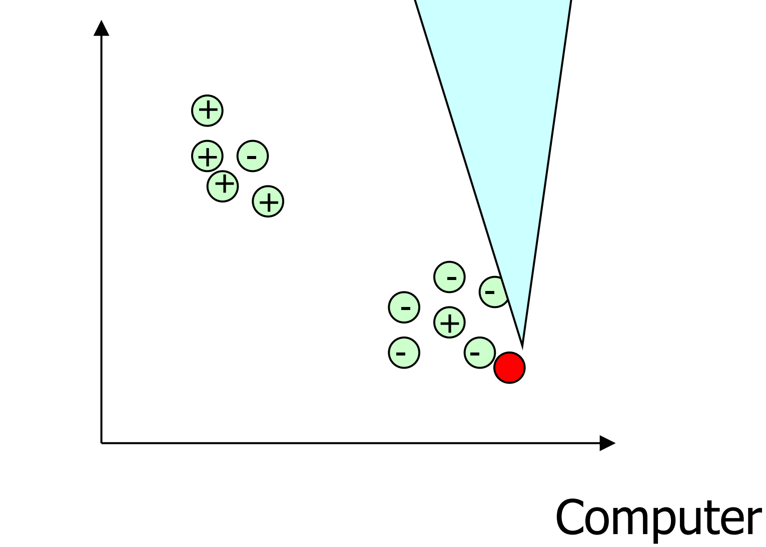
Computer	History	Buy Book?
100	40	No (-)
90	45	Yes (+)
20	95	Yes (+)
...	...	...



Nearest Neighbor Classifier

Nearest Neighbor Classifier:
Step 1: Find the nearest neighbor
Step 2: Use the "label" of this neighbor

Computer	History	Buy Book?
100	40	No (-)
90	45	Yes (+)
20	95	Yes (+)
...	...	...



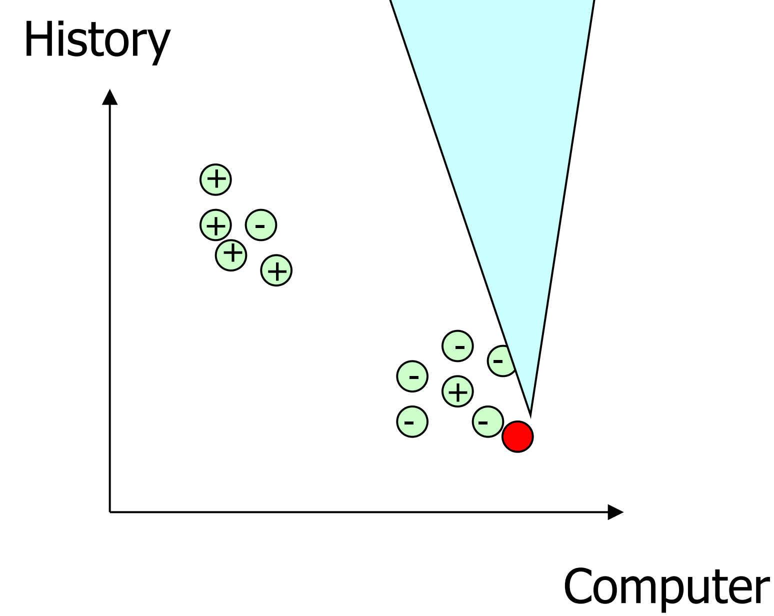
Suppose there is a new person

Computer	History	Buy Book?
95	35	?

Nearest Neig

k-Nearest Neighbor Classifier:
Step 1: Find k nearest neighbors
Step 2: Use the majority of the labels of the neighbors

Computer	History	Buy Book?
100	40	No (-)
90	45	Yes (+)
20	95	Yes (+)
…	…	…



Suppose there is a new person

Computer	History	Buy Book?
95	35	?