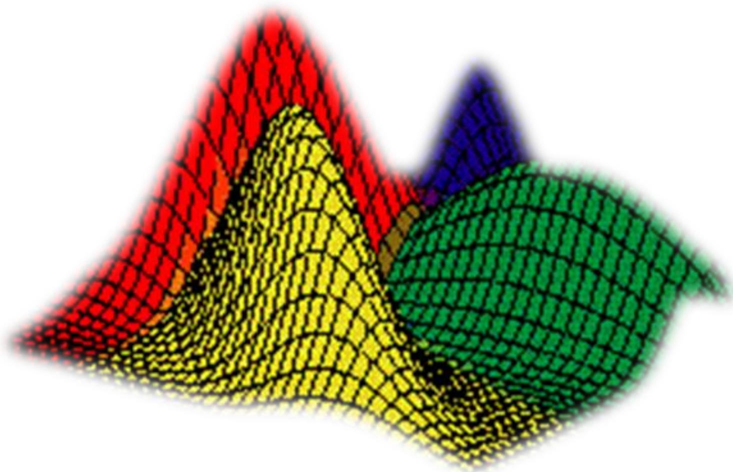


CSE 473
Pattern Recognition

Classifiers, Discriminant Functions and Decision Surfaces

Classifiers, Discriminant Functions and Decision Surfaces

- Remember the Bayesian classifier:

$$\text{Decide } \omega_i \text{ if } P(\omega_i|\mathbf{x}) > P(\omega_j|\mathbf{x}) \qquad \text{for all } j \neq i.$$

Classifiers, Discriminant Functions and Decision Surfaces

- Remember the Bayesian classifier:

$$\text{Decide } \omega_i \text{ if } P(\omega_i|\mathbf{x}) > P(\omega_j|\mathbf{x}) \qquad \text{for all } j \neq i.$$

$$g_i(x) > g_j(x)$$

Classifiers, Discriminant Functions and Decision Surfaces

- Remember the Bayesian classifier:

$$\text{Decide } \omega_i \text{ if } P(\omega_i|\mathbf{x}) > P(\omega_j|\mathbf{x}) \qquad \text{for all } j \neq i.$$

- This is equivalent to:

decide x to class $\boldsymbol{\omega_i}$

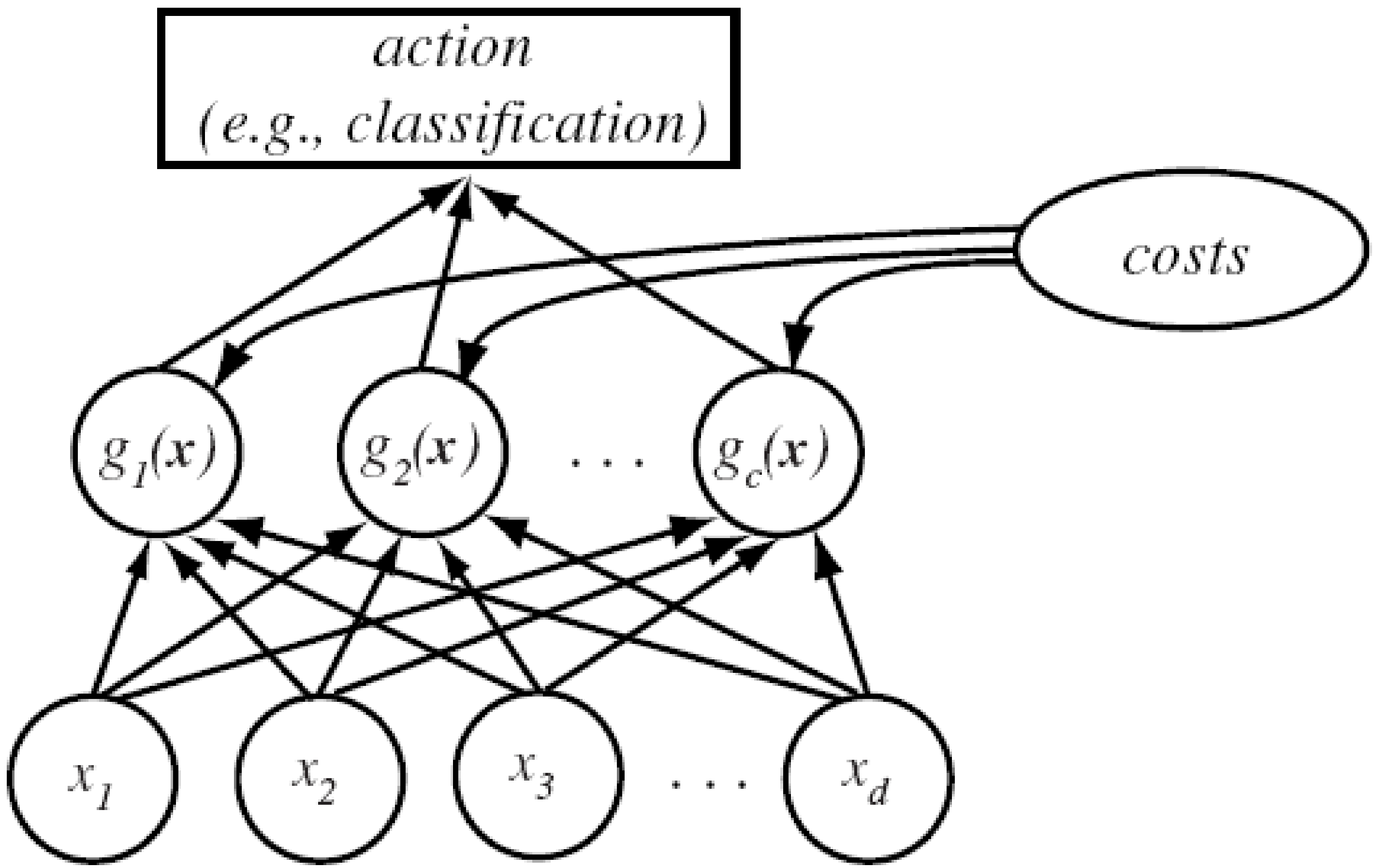
if $g_i(x) > g_j(x)\ \forall j \neq i$

where, $g_i(x) = P(\omega_i|\boldsymbol{x}),\ i = 1, \ldots, c$

Classifiers, Discriminant Functions and Decision Surfaces

Discriminant Functions

d- dimensional Feature vector



Classifiers, Discriminant Functions and Decision Surfaces

- *Based on minimum risk classification*
  - Let $g_i(x) = -R(\alpha_i \mid x)$

  (max. discriminant corresponds to min. risk!)

Classifiers, Discriminant Functions and Decision Surfaces

- For the minimum error rate, we take

$$g_i(x) = P(\omega_i \mid x)$$

(max. discrimination corresponds to max. posterior!)

Classifiers, Discriminant Functions and Decision Surfaces

- For the minimum error rate, we take

$$g_i(x) = P(\omega_i \mid x)$$

some alternate representations but giving similar results

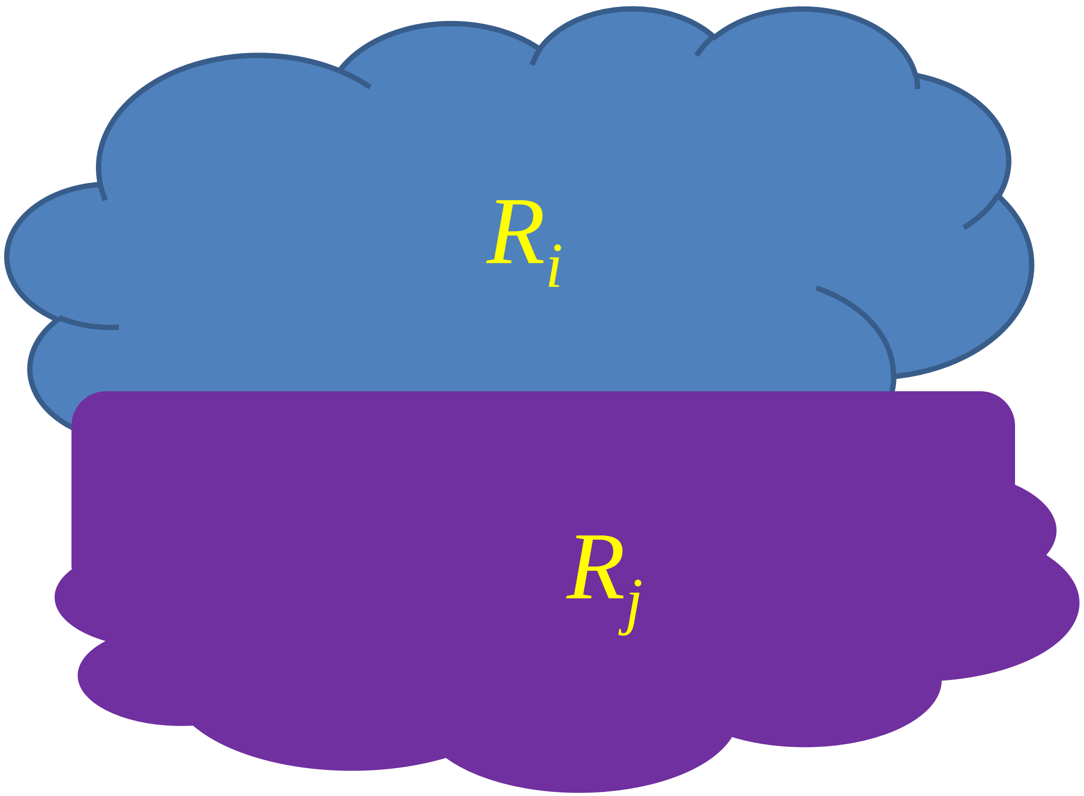
$$g_i(x) \equiv P(x \mid \omega_i)\, P(\omega_i)$$

$$g_i(x) \equiv \ln P(x \mid \omega_i) + \ln P(\omega_i)$$

(ln: natural logarithm!)

Decision Surface

- Let, R_i and R_j : two regions identifying classes ω_i and ω_j

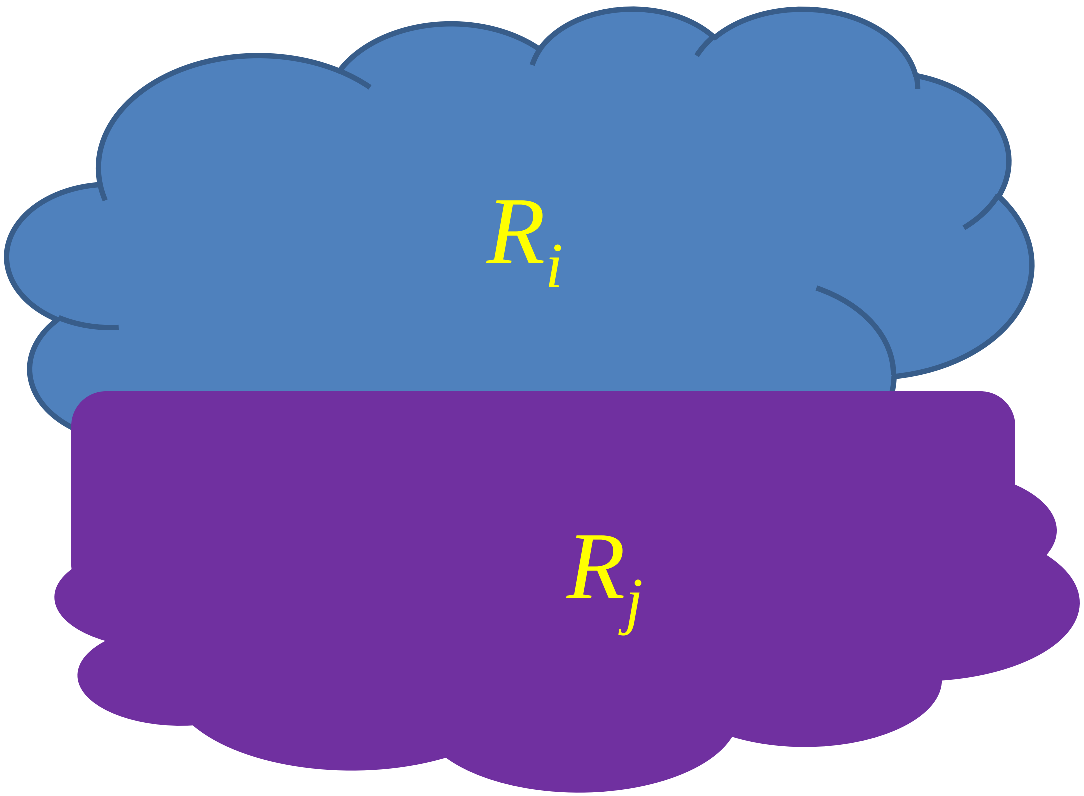


Decision Surface

- Let, R_i and R_j : two regions identifying classes ω_i and ω_j

Decision Rule

Decide ω_i if $P(\omega_i \,|x) > P(\omega_j \,|x)$

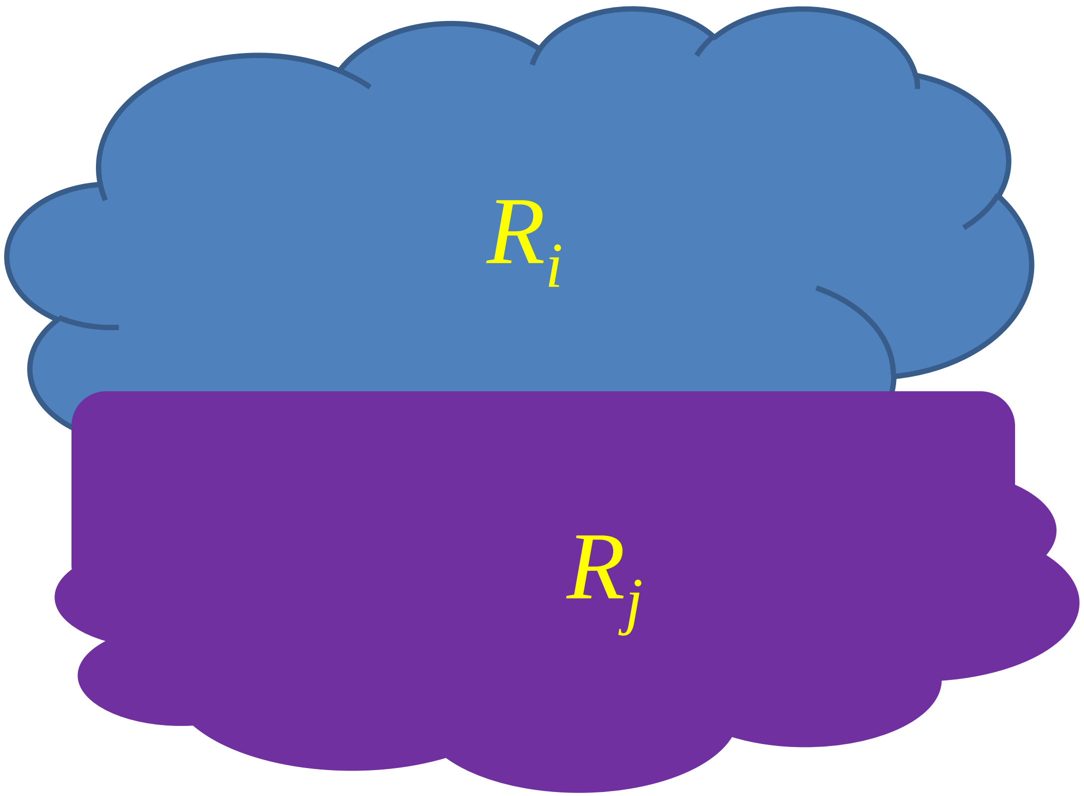


Decision Surface

- Let, R_i and R_j : two regions identifying classes ω_i and ω_j

Decision Rule

Decide ω_i if $P(\omega_i\,|x) - P(\omega_j\,|x) > 0$

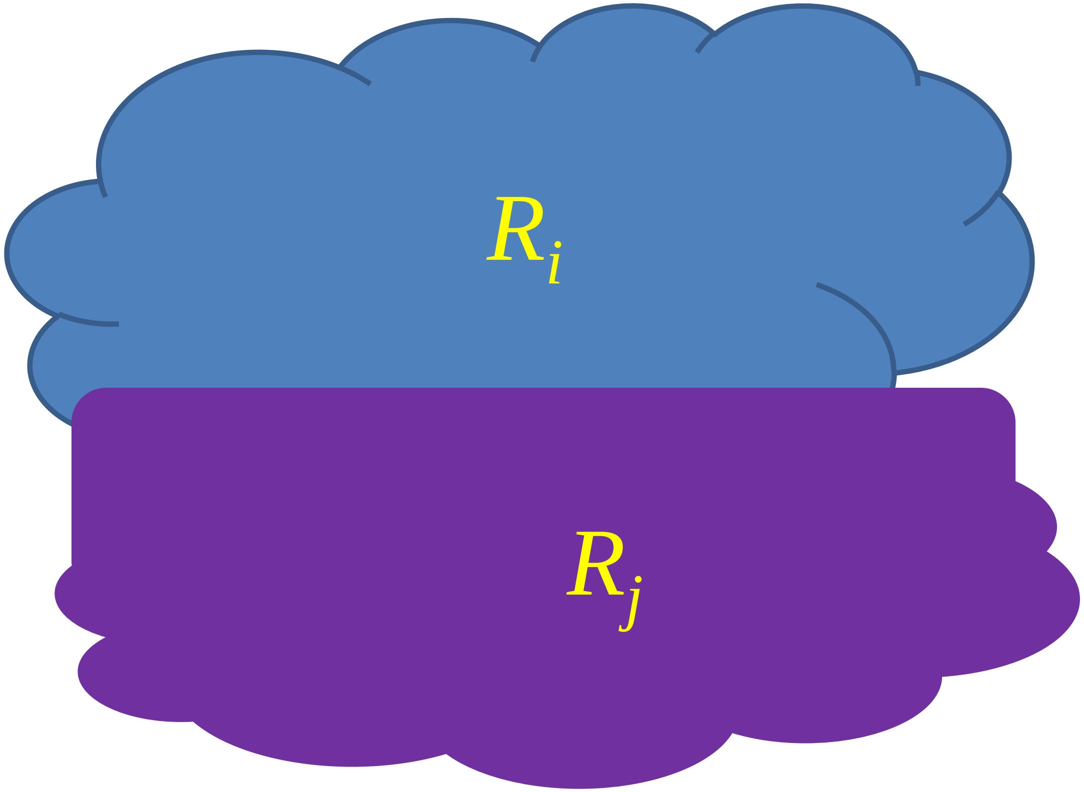


Decision Surface

- Let, R_i and R_j : two regions identifying classes ω_i and ω_j

Decision Rule

Decide ω_i if $g(x) \equiv P(\omega_i\,|x) - P(\omega_j\,|x) > 0$



Decision Surface

- Let, R_i, R_j : two regions identifying classes ω_i and ω_j

$$g(\underline{x}) \equiv P(\omega_i | \underline{x}) - P(\omega_j | \underline{x}) = 0$$

Decision Surface

- If R_i, R_j: two regions identifying classes ω_i and ω_j

$$g(\underline{x}) \equiv P(\omega_i|\underline{x}) - P(\omega_j|\underline{x}) = 0$$

Decision Surface

- If R_i, R_j: two regions identifying classes ω_i and ω_j

$$g(\underline{x}) \equiv P(\omega_i|\underline{x}) - P(\omega_j|\underline{x}) = 0$$

Decision Surface

- If R_i, R_j: two regions identifying classes ω_i and ω_j

$$g(\underline{x}) \equiv P(\omega_i|\underline{x}) - P(\omega_j|\underline{x}) = 0$$

$$R_i : P(\omega_i|\underline{x}) > P(\omega_j|\underline{x})$$

+

-

$$g(\underline{x}) = 0$$

$$R_j : P(\omega_j|\underline{x}) > P(\omega_i|\underline{x})$$

decision surface

Decision Surface in Multi-categories

- Feature space is divided into c decision regions

$$\text{if } g_i(x) > g_j(x) \;\; \forall j \neq i \text{ then } x \text{ is in } R_i$$

(R_i means: assign x to ω_i)

Decision Surface in Two-categories

- The two-category case
  - A classifier is a "*dichotomizer*" that has two discriminant functions g_1 and g_2

    Let $g(x) \equiv g_1(x) - g_2(x)$

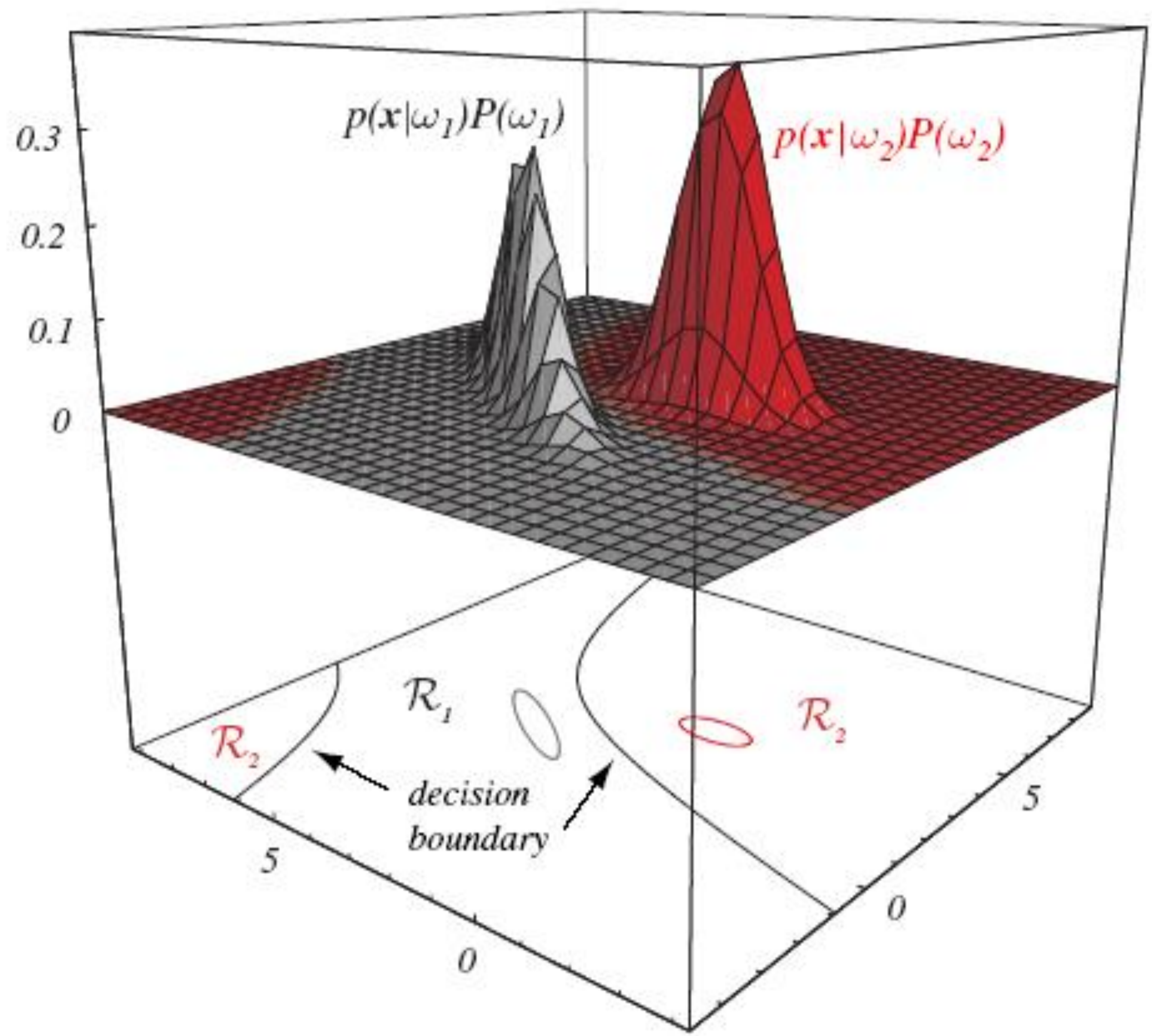
$$\text{Decide } \omega_1 \text{ if } g(x) > 0 \text{ ; Otherwise decide } \omega_2$$

– The computation of $g(x)$

$$g(x) \equiv P(\omega_1 \mid x) - P(\omega_2 \mid x)$$

$$= P(x \mid \omega_1)P(\omega_1) - P(x \mid \omega_2)P(\omega_2)$$

$$\equiv \ln \frac{P(x \mid \omega_1)}{P(x \mid \omega_2)} + \ln \frac{P(\omega_1)}{P(\omega_2)}$$



$p(\boldsymbol{x}|\omega_1)P(\omega_1)$
$p(\boldsymbol{x}|\omega_2)P(\omega_2)$
0.3
0.2
0.1
0
$\mathcal{R}_1$
$\mathcal{R}_2$
$\mathcal{R}_2$
decision boundary
5
0
5
0

The Normal Density

- Density which is analytically tractable
- Continuous density
- A lot of processes are asymptotically Gaussian
  - Handwritten characters, speech sounds, and many more
  - Any prototype corrupted by random process

The Normal Density

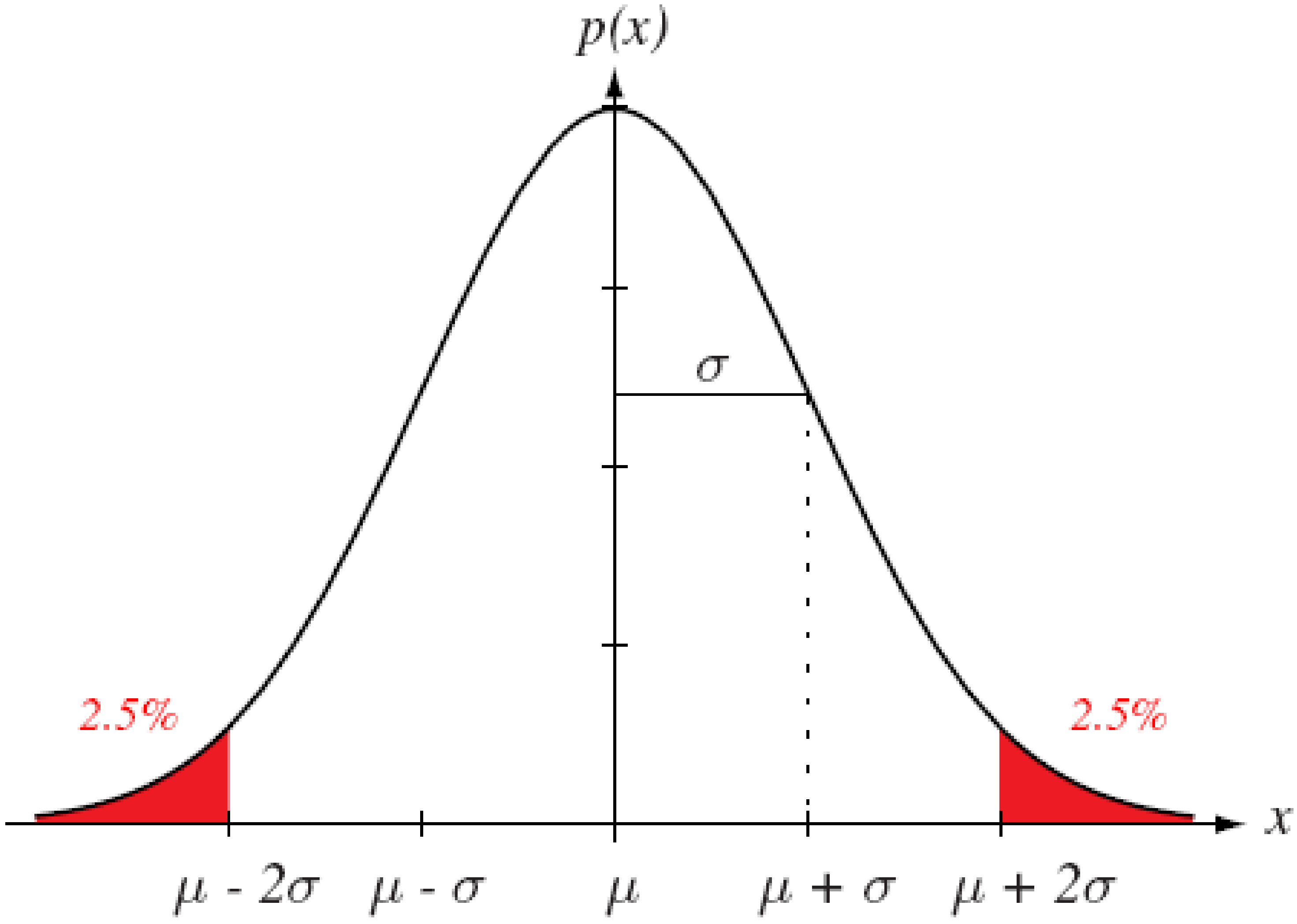
- Univariate density

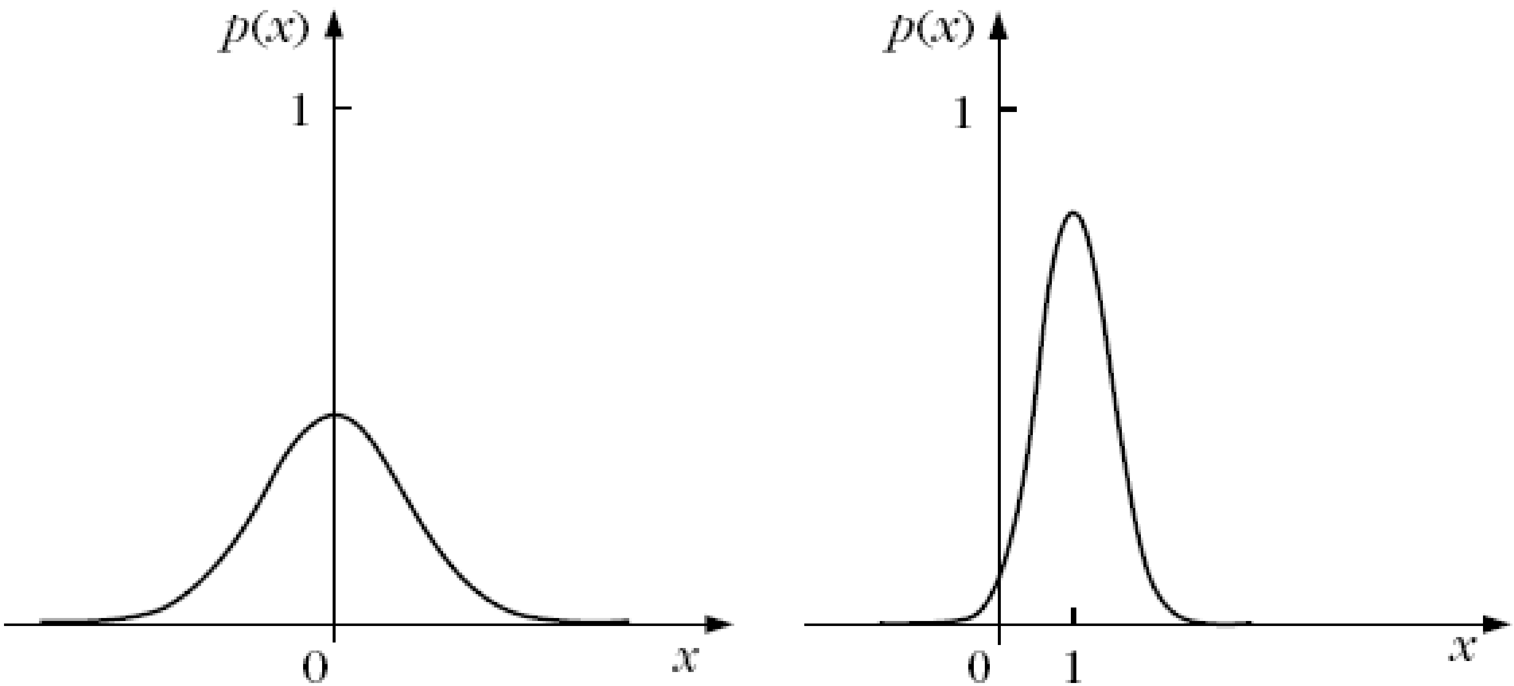
$$P(x) = \frac{1}{\sqrt{2\pi}\,\sigma} exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right],$$

Where:

μ = mean (or expected value) of x

σ^2 = expected squared deviation or variance


p(x)
σ
2.5%
2.5%
x
μ - 2σ
μ - σ
μ
μ + σ
μ + 2σ



(a) (b)

- Multivariate density

  - Multivariate normal density in *d* dimensions is:

$$P(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right]$$

where:

$x = (x_1, x_2, \ldots, x_d)^t$ (t stands for the transpose vector form)

$\mu = (\mu_1, \mu_2, \ldots, \mu_d)^t$ mean vector

$\Sigma = d*d$ covariance matrix

$|\Sigma|$ and Σ^{-1} are determinant and inverse respectively

- Multivariate density

  - Multivariate normal density in *d* dimensions is:

$$P(x) = \frac{1}{(2\pi)^{d/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)\right]$$

where:

$$\Sigma = E[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T]$$

$$\Sigma = E\left[\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}\begin{bmatrix} x_1 - \mu_1, & x_2 - \mu_2 \end{bmatrix}\right]$$

$$= \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

$$\Sigma = E[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})^T]$$

$$\Sigma = E\left[\begin{bmatrix} x_1 - \mu_1 \\ x_2 - \mu_2 \end{bmatrix}\begin{bmatrix} x_1 - \mu_1, & x_2 - \mu_2 \end{bmatrix}\right]$$

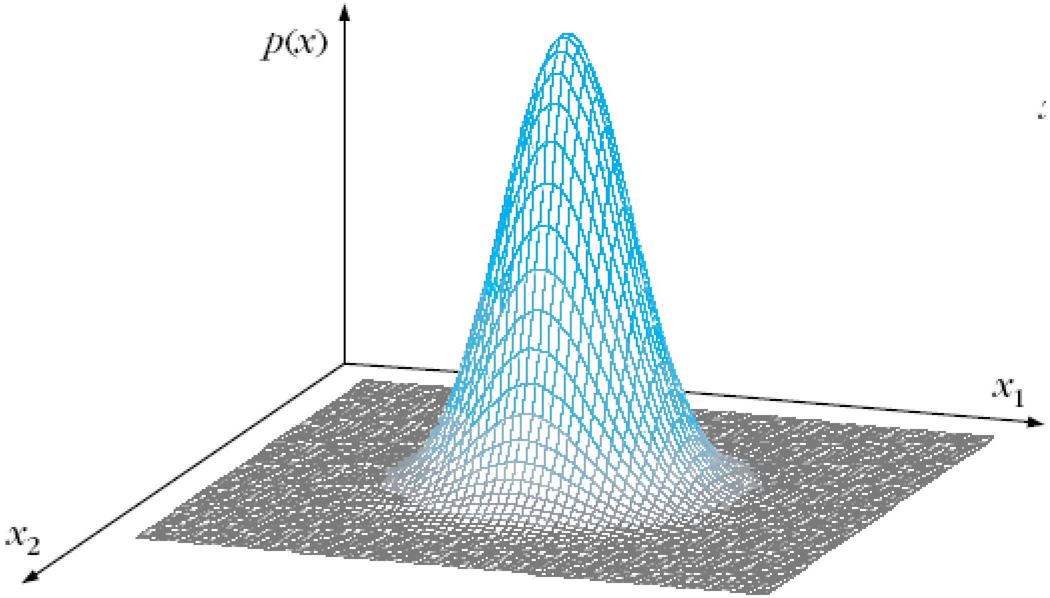
$$= \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}$$

$$E[x_i] = \mu_i, \; i = 1, 2$$

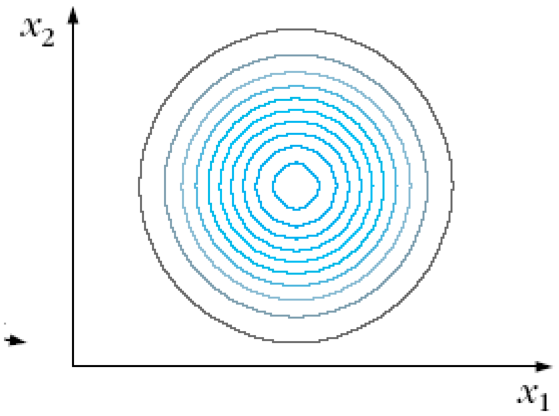
$$\sigma_{12} = E[(x_1 - \mu_1)(x_2 - \mu_2)]$$

2D Gaussian Example - 1

$$\Sigma = \begin{bmatrix} 3 & 0 \\ 0 & 3 \end{bmatrix}$$

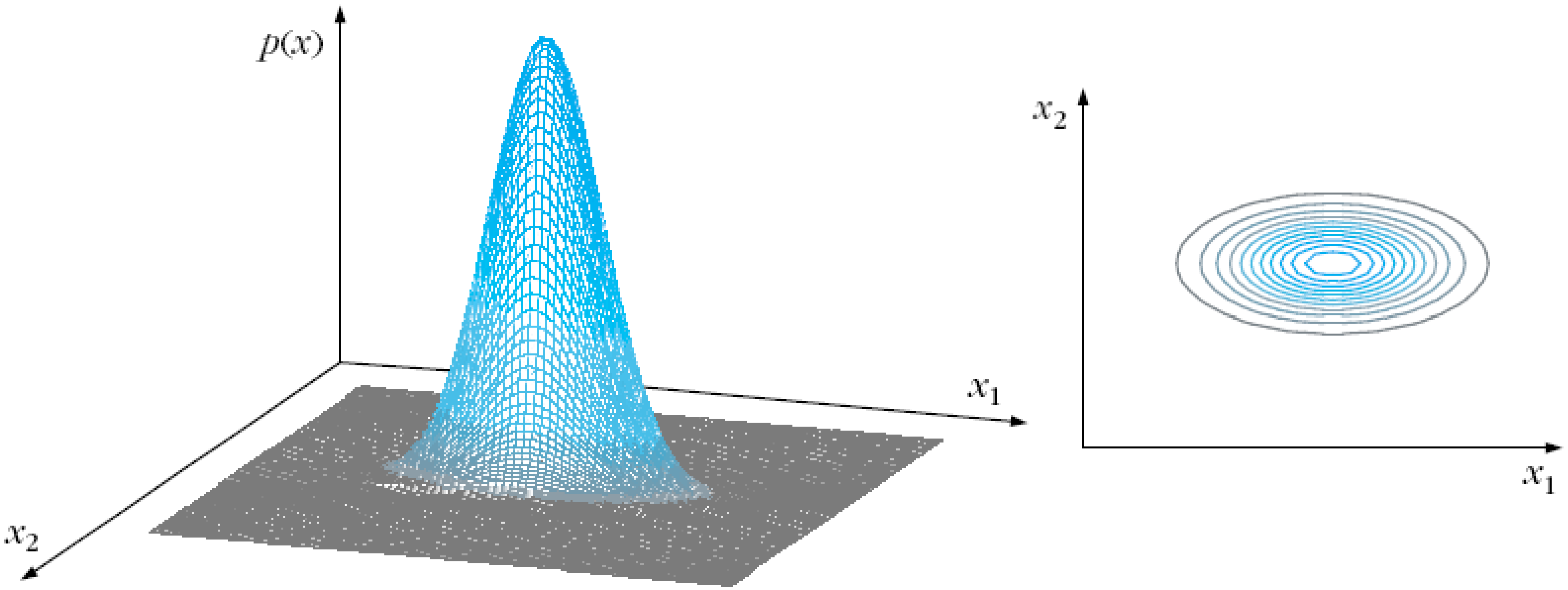


(a)



2D Gaussian Example - 2

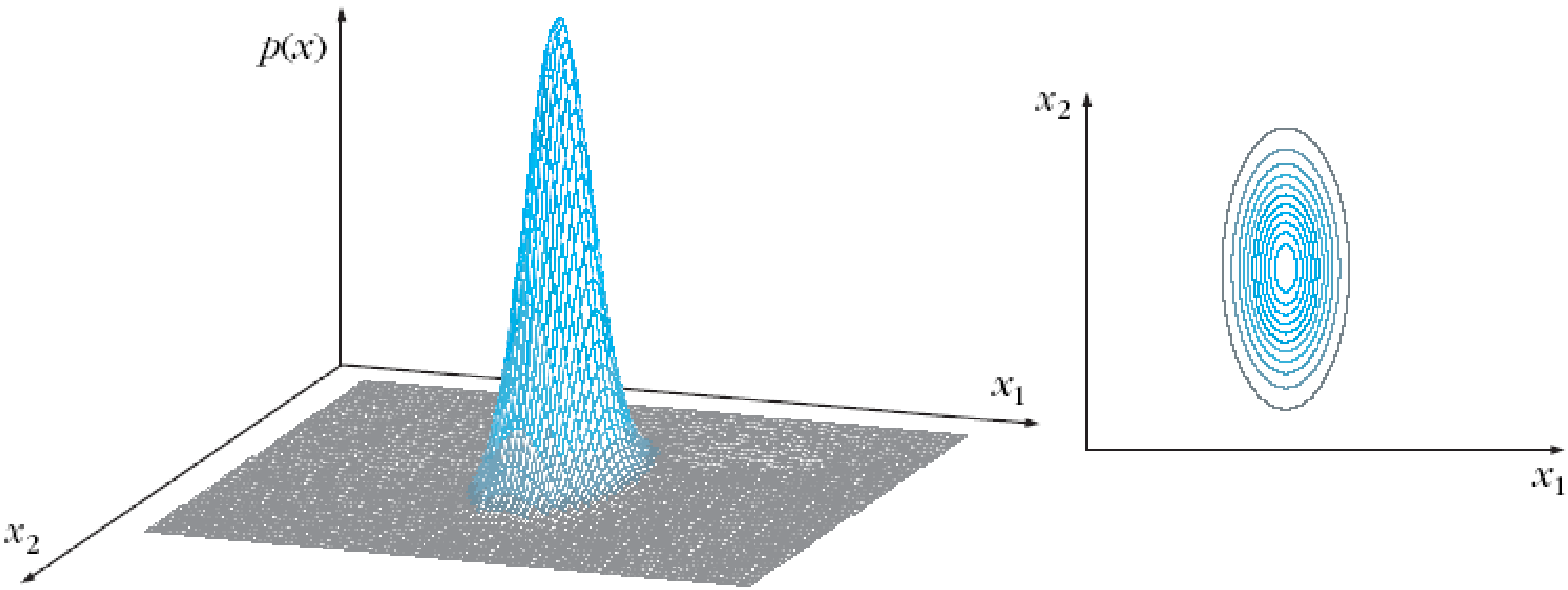
$$\sigma_1^2 >> \sigma_2^2$$



(a) (b)

2D Gaussian Example - 3

$$\sigma_1^2 << \sigma_2^2$$



(a)

(b)

Computer Exercise

- Use matlab to generate Gaussian plots
- Try with different Σ and б

Classification Example 2

Tid	**Refund**	**Marital Status**	**Taxable Income**	**Evade**
1	Yes	Single	125K	**No**
2	No	Married	100K	**No**
3	No	Single	70K	**No**
4	Yes	Married	120K	**No**
5	No	Divorced	95K	**Yes**
6	No	Married	60K	**No**
7	Yes	Divorced	220K	**No**
8	No	Single	85K	**Yes**
9	No	Married	75K	**No**
10	No	Single	90K	**Yes**

- A married person with income 120K did not refund the loan previously
- *Can we trust him*?

Bayesian Classifiers

- We have multiple attributes ($A_1, A_2, \ldots, A_n$)
  - Goal is to predict class C
  - Specifically, we want to find the value of C that maximizes $P(C| A_1, A_2, \ldots, A_n)$
- Can we estimate $P(C| A_1, A_2, \ldots, A_n)$ directly from data?

Bayesian Classifiers

- Approach:
  - compute the posterior probability $P(C \mid A_1, A_2, \ldots, A_n)$ for all values of C using the Bayes theorem

$$\boxed{P(C \mid A_1 A_2 \ldots A_n) = \frac{P(A_1 A_2 \ldots A_n \mid C) P(C)}{P(A_1 A_2 \ldots A_n)}}$$

  - Choose value of C that maximizes
    $P(C \mid A_1, A_2, \ldots, A_n)$

  - Equivalent to choosing value of C that maximizes
    $P(A_1, A_2, \ldots, A_n|C)\, P(C)$

- How to estimate $P(A_1, A_2, \ldots, A_n \mid C)$?

Naïve Bayes Classifier

- Assume independence among attributes A_i when class is given:
  - $P(A_1, A_2, \ldots, A_n | C_j) = P(A_1| C_j) P(A_2| C_j) \ldots P(A_n| C_j)$
  - Can estimate $P(A_i| C_j)$ for all A_i and C_j
  - *the new pattern* is classified to C_j if $P(C_j) \Pi P(A_i| C_j)$ is maximum

How to Estimate Probabilities from Data?

Tid	**Refund**	**Marital Status**	**Taxable Income**	**Evade**
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Class: $P(C) = N_c/N$
  - e.g., $P(No) = 7/10$,
    $P(Yes) = 3/10$
- For discrete attributes:

$$P(A_i \mid C_k) = |A_{ik}|/ N_c$$

  - where $|A_{ik}|$ is number of instances having attribute A_i and belongs to class C_k
  - Examples:

P(Status=Married|No) = 4/7
P(Refund=Yes|Yes)=0

How to Estimate Probabilities from Data?

- For continuous attributes:
  - Discretize the range into bins
    - one ordinal attribute per bin
  - Two-way split: (A < v) or (A > v)
    - choose only one of the two splits as new attribute
  - Probability density estimation:
    - Assume attribute follows a normal distribution
    - Use data to estimate parameters of distribution (e.g., mean and standard deviation)
    - Once probability distribution is known, can use it to estimate the conditional probability $P(A_i|c)$

How to Estimate Probabilities from Data?

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Normal distribution:

$$P(A_i \mid c_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(A_i-\mu_{ij})^2}{2\sigma_{ij}^2}}$$

  – One for each (A_i, c_j) pair

- For (Income, Class=No):
  – If Class=No
    - sample mean = 110K
    - sample variance = 2975

How to Estimate Probabilities from Data?

Tid	Refund	Marital Status	Taxable Income	Evade
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

- Normal distribution:

$$P(A_i \mid c_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} e^{-\frac{(A_i-\mu_{ij})^2}{2\sigma_{ij}^2}}$$

  – One for each (A_i, c_j) pair

- For (Income, Class=No):
  – If Class=No
    - sample mean = 110K
    - sample variance = 2975

$$P(Income = 120 \mid No) = \frac{1}{\sqrt{2\pi}(54.54)} e^{-\frac{(120-110)^2}{2(2975)}} = 0.0072$$

Example of Naïve Bayes Classifier

Given a Test Record: $X = (\text{Refund} = \text{No}, \text{Married}, \text{Income} = 120\text{K})$

naive Bayes Classifier:

P(Refund=Yes|No) = 3/7
P(Refund=No|No) = 4/7
P(Refund=Yes|Yes) = 0
P(Refund=No|Yes) = 1
P(Marital Status=Single|No) = 2/7
P(Marital Status=Divorced|No)=1/7
P(Marital Status=Married|No) = 4/7
P(Marital Status=Single|Yes) = 2/7
P(Marital Status=Divorced|Yes)=1/7
P(Marital Status=Married|Yes) = 0

For taxable income:
If class=No: sample mean=110
sample variance=2975
If class=Yes: sample mean=90
sample variance=25

- P(X|Class=No) = P(Refund=No|Class=No)
  $\times$ P(Married| Class=No)
  $\times$ P(Income=120K| Class=No)
  = 4/7 $\times$ 4/7 $\times$ 0.0072 = 0.0024

- P(X|Class=Yes) = P(Refund=No| Class=Yes)
  $\times$ P(Married| Class=Yes)
  $\times$ P(Income=120K| Class=Yes)
  = 1 $\times$ 0 $\times$ 1.2 $\times$ 10^{-9} = 0

Since P(X|No)P(No) > P(X|Yes)P(Yes)

Therefore P(No|X) > P(Yes|X)
=> Class = No

Example-2 of Naïve Bayes Classifier

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

Example-2 of Naïve Bayes Classifier

Name	Give Birth	Can Fly	Live in Water	Have Legs	Class
human	yes	no	no	yes	mammals
python	no	no	no	no	non-mammals
salmon	no	no	yes	no	non-mammals
whale	yes	no	yes	no	mammals
frog	no	no	sometimes	yes	non-mammals
komodo	no	no	no	yes	non-mammals
bat	yes	yes	no	yes	mammals
pigeon	no	yes	no	yes	non-mammals
cat	yes	no	no	yes	mammals
leopard shark	yes	no	yes	no	non-mammals
turtle	no	no	sometimes	yes	non-mammals
penguin	no	no	sometimes	yes	non-mammals
porcupine	yes	no	no	yes	mammals
eel	no	no	yes	no	non-mammals
salamander	no	no	sometimes	yes	non-mammals
gila monster	no	no	no	yes	non-mammals
platypus	no	no	no	yes	mammals
owl	no	yes	no	yes	non-mammals
dolphin	yes	no	yes	no	mammals
eagle	no	yes	no	yes	non-mammals

A: attributes

M: mammals

N: non-mammals

$$P(A|M) = \frac{6}{7} \times \frac{6}{7} \times \frac{2}{7} \times \frac{2}{7} = 0.06$$

$$P(A|N) = \frac{1}{13} \times \frac{10}{13} \times \frac{3}{13} \times \frac{4}{13} = 0.0042$$

$$P(A|M)P(M) = 0.06 \times \frac{7}{20} = 0.021$$

$$P(A|N)P(N) = 0.004 \times \frac{13}{20} = 0.0027$$

Give Birth	Can Fly	Live in Water	Have Legs	Class
yes	no	yes	no	?

P(A|M)P(M) > P(A|N)P(N)

=> Mammals

Naïve Bayes (Summary)

- Robust to isolated noise points
- Handle missing values by ignoring the instance during probability estimate calculations
- Robust to irrelevant attributes
- Independence assumption may not hold for some attributes
  - Use other techniques such as Bayesian Belief Networks (BBN)

Naïve Bayes (Issues)

- Over simplification
  - Use other techniques such as Bayesian Belief Networks (BBN)

Bayesian Belief Networks

- Let we have *l* random variables
- The joint probability is given by,

$$p(x_1, x_2, \ldots, x_\ell) = p(x_\ell \mid x_{\ell-1}, \ldots, x_1) \cdot p(x_{\ell-1} \mid x_{\ell-2}, \ldots, x_1) \cdot \ldots$$

$$\ldots \cdot p(x_2 \mid x_1) \cdot p(x_1)$$

Bayesian Belief Networks

The formula

$$p(x_1, x_2, \ldots, x_\ell) = p(x_\ell \mid x_{\ell-1}, \ldots, x_1) \cdot p(x_{\ell-1} \mid x_{\ell-2}, \ldots, x_1) \cdot \ldots$$

$$\ldots \cdot p(x_2 \mid x_1) \cdot p(x_1)$$

can be written as

$$p(x_1, x_2, \ldots, x_\ell) = p(x_1) \cdot \prod_{i=2}^{\ell} p(x_i \mid A_i)$$

where

$$A_i \subseteq \{x_{i-1}, x_{i-2}, \ldots, x_1\}$$

– For example, if ℓ=6, then we could assume:

$$p(x_6 \mid x_5, \ldots, x_1) = p(x_6 \mid x_5, x_4)$$

Then:

$$A_6 = \{x_5, x_4\} \subseteq \{x_5, \ldots, x_1\}$$

– Simialrly, if we assume

$$p(x_5|x_4, \ldots, x_1) = p(x_5|x_4)$$

$$p(x_4|x_3, x_2, x_1) = p(x_4|x_2, x_1)$$

$$p(x_3|x_2, x_1) = p(x_3|x_2)$$

$$p(x_2|x_1) = p(x_2)$$

Then:

$$A_5 = \{x_4\},\ A_4 = \{x_2, x_1\},\ A_3 = \{x_2\}, A_2 = \emptyset$$

– Similalrly, if we assume

$$p(x_5|x_4,\ldots,x_1) = p(x_5|x_4)$$

$$p(x_4|x_3,x_2,x_1) = p(x_4|x_2,x_1)$$

$$p(x_3|x_2,x_1) = p(x_3|x_2)$$

$$p(x_2|x_1) = p(x_2)$$

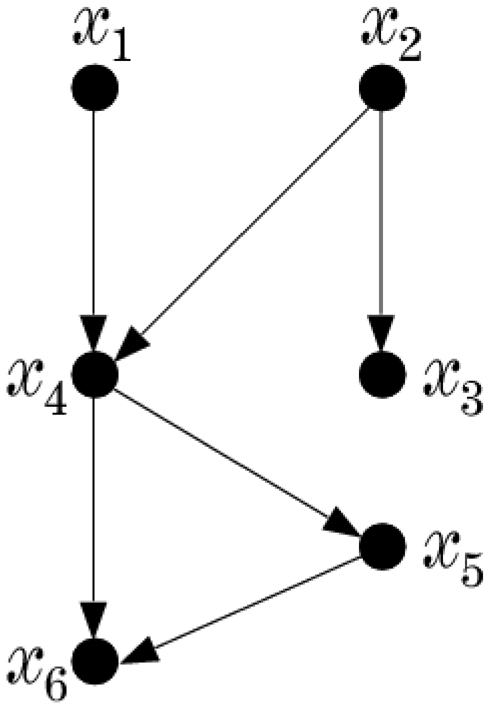
Then:

$$A_5 = \{x_4\},\ A_4 = \{x_2, x_1\},\ A_3 = \{x_2\}, A_2 = \emptyset$$

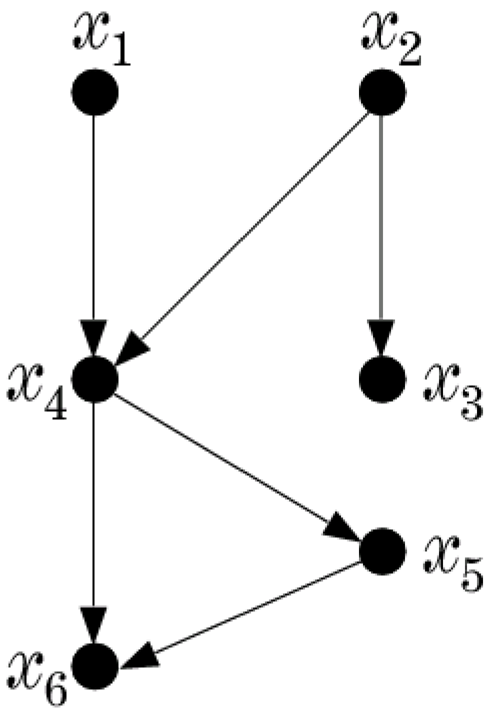
– The above is a generalization of the Naïve – Bayes. For the Naïve – Bayes the assumption is:

$$A_i = \emptyset, \text{ for } i=1, 2, \ldots, \ell$$

– A graphical way to portray conditional dependencies

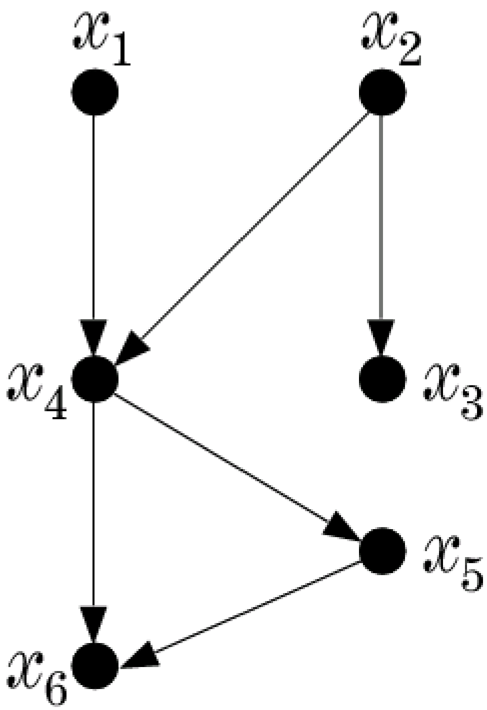


– A graphical way to portray conditional dependencies



- According to this figure we have :
  - x_6 is conditionally dependent on x_4, x_5.
  - x_5 on x_4
  - x_4 on x_1, x_2
  - x_3 on x_2
  - x_1, x_2 are conditionally independent on other variables.

– A graphical way to portray conditional dependencies



- According to this figure we have :
  - x_6 is conditionally dependent on x_4, x_5.
  - x_5 on x_4
  - x_4 on x_1, x_2
  - x_3 on x_2
  - x_1, x_2 are conditionally independent on other variables.

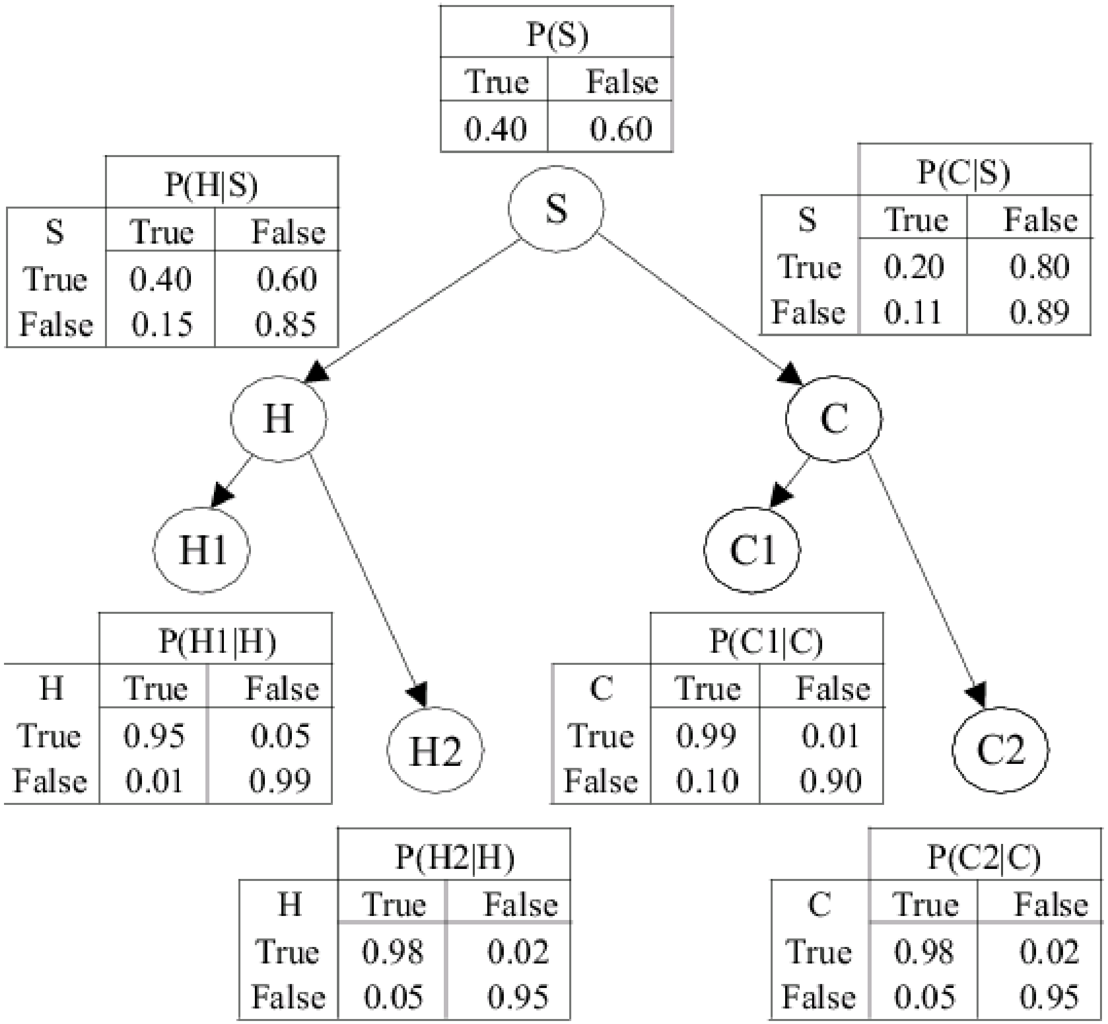
- For this case:

$$p(x_1, x_2, \ldots, x_6) = p(x_6 \mid x_5, x_4) \cdot p(x_5 \mid x_4) \cdot p(x_3 \mid x_2) \cdot p(x_2) \cdot p(x_1)$$

- Bayesian Networks
  - a directed acyclic graph (DAG)
  - the nodes correspond to random variables
  - arc represents parent-child (*dependence*) relationship

- A Bayesian Network is specified by:
  - The prior probabilities of its root nodes.
  - The conditional probabilities of the non-root nodes, given their parents, for ALL possible combinations.

– A Bayesian Network from a medical application



P(S)	
True	False
0.40	0.60

	P(H|S)	
S	True	False
True	0.40	0.60
False	0.15	0.85

	P(C|S)	
S	True	False
True	0.20	0.80
False	0.11	0.89

	P(H1|H)	
H	True	False
True	0.95	0.05
False	0.01	0.99

	P(C1|C)	
C	True	False
True	0.99	0.01
False	0.10	0.90

	P(H2|H)	
H	True	False
True	0.98	0.02
False	0.05	0.95

	P(C2|C)	
C	True	False
True	0.98	0.02
False	0.05	0.95

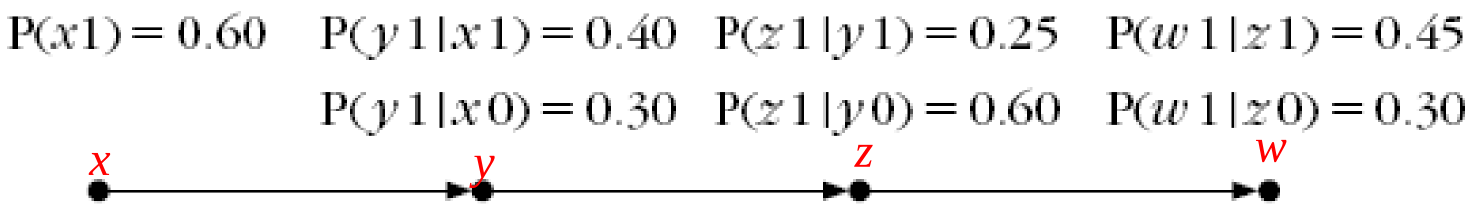
- This BBN models conditional dependencies concerning smokers' (S), tendencies to develop cancer (C) and heart disease (H), together with variables corresponding to heart (H1, H2) and cancer (C1, C2) medical tests

- any joint probability can be obtained by multiplying the prior (root nodes) and the conditional (non-root nodes) probabilities.

- Training: given a topology, probabilities are estimated from training data. There are also methods that learn the topology.

- Probability Inference: Given a pattern (evidence), the goal is to compute the conditional probabilities for some of the other variables (class)

- Example: Consider the Bayesian network of the figure:



- Random variables: *x*, *y*, *w*, *z*
- *x*0 means *x* = 0
- *x*1 means *x* = 1

- We can calculate the other probabilities

$P(x1)=0.60$ $\quad P(y1|x1)=0.40$ $\quad P(z1|y1)=0.25$ $\quad P(w1|z1)=0.45$

$P(y1|x0)=0.30$ $\quad P(z1|y0)=0.60$ $\quad P(w1|z0)=0.30$

$P(x0)=0.40$ $\quad P(y0|x1)=0.60$

$P(y0|x0)=0.70$

$P(y1)=0.36$

$P(y0)=0.64$

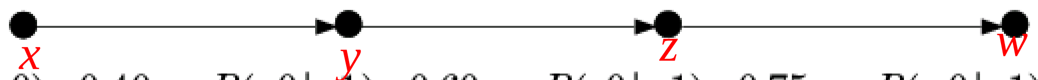
Example: $p(y_1)$:

$$P(y1)=\sum_{x} P(y1,x) = P(y1,x1)+P(y1,x0)$$

$$P(y1) = P(y1|x1)P(x1) + P(y1|x0)P(x0) = (0.4)(0.6) + (0.3)(0.4) = 0.36$$

- We can calculate the other probabilities

$P(x1)=0.60$	$P(y1\mid x1)=0.40$	$P(z1\mid y1)=0.25$	$P(w1\mid z1)=0.45$
	$P(y1\mid x0)=0.30$	$P(z1\mid y0)=0.60$	$P(w1\mid z0)=0.30$



$P(x0)=0.40$	$P(y0\mid x1)=0.60$	$P(z0\mid y1)=0.75$	$P(w0\mid z1)=0.55$
	$P(y0\mid x0)=0.70$	$P(z0\mid y0)=0.40$	$P(w0\mid z0)=0.70$
	$P(y1)=0.36$	$P(z1)=0.47$	$P(w1)=0.37$
	$P(y0)=0.64$	$P(z0)=0.53$	$P(w0)=0.63$

- Given this info, we can answer any probabilistic query:

$P(x1)=0.60 \quad P(y1|x1)=0.40 \quad P(z1|y1)=0.25 \quad P(w1|z1)=0.45$

$P(y1|x0)=0.30 \quad P(z1|y0)=0.60 \quad P(w1|z0)=0.30$

$P(x0)=0.40 \quad P(y0|x1)=0.60 \quad P(z0|y1)=0.75 \quad P(w0|z1)=0.55$

$P(y0|x0)=0.70 \quad P(z0|y0)=0.40 \quad P(w0|z0)=0.70$

$P(y1)=0.36 \quad P(z1)=0.47 \quad P(w1)=0.37$

$P(y0)=0.64 \quad P(z0)=0.53 \quad P(w0)=0.63$

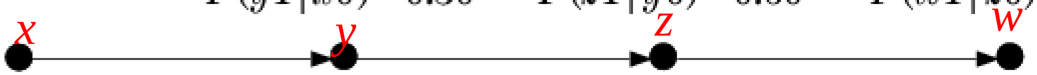
a) If x is measured to be $x=1$ ($x1$), compute $P(z1|x1)$ *and* $P(w0|x1)$.

b) If w is measured to be $w=1$ ($w1$) compute $P(z1|w1)$].

a) If x is measured to be $x=1$ ($x1$), compute P(z1|x1) and $P(w0|x1)$.

$$P(x1)=0.60 \quad P(y1|x1)=0.40 \quad P(z1|y1)=0.25 \quad P(w1|z1)=0.45$$

$$P(y1|x0)=0.30 \quad P(z1|y0)=0.60 \quad P(w1|z0)=0.30$$



$$P(x0)=0.40 \quad P(y0|x1)=0.60 \quad P(z0|y1)=0.75 \quad P(w0|z1)=0.55$$

$$P(y0|x0)=0.70 \quad P(z0|y0)=0.40 \quad P(w0|z0)=0.70$$

$$P(y1)=0.36 \quad P(z1)=0.47 \quad P(w1)=0.37$$

$$P(y0)=0.64 \quad P(z0)=0.53 \quad P(w0)=0.63$$

$$P(z1|x1) = P(z1|y1, x1)P(y1|x1) + P(z1|y0, x1)P(y0|x1)$$

$$= P(z1|y1)P(y1|x1) + P(z1|y0)P(y0|x1)$$

$$= (0.25)(0.4) + (0.6)(0.6) = 0.46$$

a) If x is measured to be $x=1$ ($x1$), compute P(z1|x1) and $P(w0|x1)$.

$$x \longrightarrow y \longrightarrow z \longrightarrow w$$

x	y	z	w
$P(x1)=0.60$	$P(y1\,|\,x1)=0.40$	$P(z1\,|\,y1)=0.25$	$P(w1\,|\,z1)=0.45$
	$P(y1\,|\,x0)=0.30$	$P(z1\,|\,y0)=0.60$	$P(w1\,|\,z0)=0.30$
$P(x0)=0.40$	$P(y0\,|\,x1)=0.60$	$P(z0\,|\,y1)=0.75$	$P(w0\,|\,z1)=0.55$
	$P(y0\,|\,x0)=0.70$	$P(z0\,|\,y0)=0.40$	$P(w0\,|\,z0)=0.70$
	$P(y1)=0.36$	$P(z1)=0.47$	$P(w1)=0.37$
	$P(y0)=0.64$	$P(z0)=0.53$	$P(w0)=0.63$

$$P(w0|x1) = P(w0|z1,\, x1)P(z1|x1) + P(w0|z0,\, x1)P(z0|x1)$$

$$= P(w0|z1)P(z1|x1) + P(w0|z0)P(z0|x1)$$

$$= (0.55)(0.46) + (0.7)(0.54) = 0.63$$

b) If w is measured to be w=1 (w1) compute P(z1|w1|)].

$P(x1)=0.60$ $\quad$ $P(y1|x1)=0.40$ $\quad$ $P(z1|y1)=0.25$ $\quad$ $P(w1|z1)=0.45$

$P(y1|x0)=0.30$ $\quad$ $P(z1|y0)=0.60$ $\quad$ $P(w1|z0)=0.30$

$P(x0)=0.40$ $\quad$ $P(y0|x1)=0.60$ $\quad$ $P(z0|y1)=0.75$ $\quad$ $P(w0|z1)=0.55$

$P(y0|x0)=0.70$ $\quad$ $P(z0|y0)=0.40$ $\quad$ $P(w0|z0)=0.70$

$P(y1)=0.36$ $\quad$ $P(z1)=0.47$ $\quad$ $P(w1)=0.37$

$P(y0)=0.64$ $\quad$ $P(z0)=0.53$ $\quad$ $P(w0)=0.63$

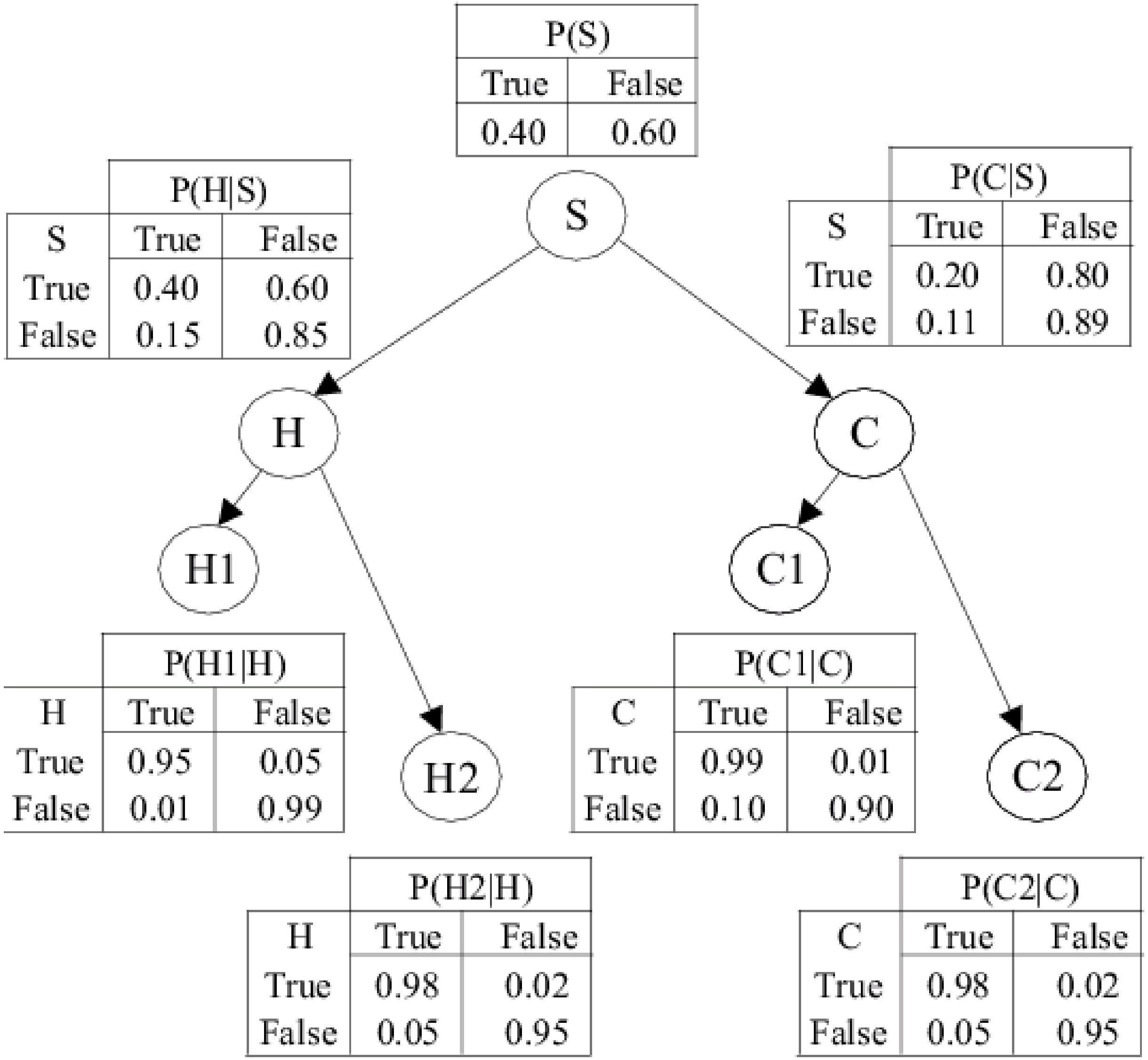
$$P(z1|w1) = \frac{P(w1|z1)P(z1)}{P(w1)} = \frac{(0.45)(0.47)}{0.37} = 0.57$$

c) CAN WE CALCULATE P(x0|w1|)]?

$P(x1)=0.60$	$P(y1\,|\,x1)=0.40$	$P(z1\,|\,y1)=0.25$	$P(w1\,|\,z1)=0.45$
	$P(y1\,|\,x0)=0.30$	$P(z1\,|\,y0)=0.60$	$P(w1\,|\,z0)=0.30$

$P(x0)=0.40$	$P(y0\,|\,x1)=0.60$	$P(z0\,|\,y1)=0.75$	$P(w0\,|\,z1)=0.55$
	$P(y0\,|\,x0)=0.70$	$P(z0\,|\,y0)=0.40$	$P(w0\,|\,z0)=0.70$
	$P(y1)=0.36$	$P(z1)=0.47$	$P(w1)=0.37$
	$P(y0)=0.64$	$P(z0)=0.53$	$P(w0)=0.63$

What’s about more complex networks?



P(S)	
True	False
0.40	0.60

	P(H|S)	
S	True	False
True	0.40	0.60
False	0.15	0.85

	P(C|S)	
S	True	False
True	0.20	0.80
False	0.11	0.89

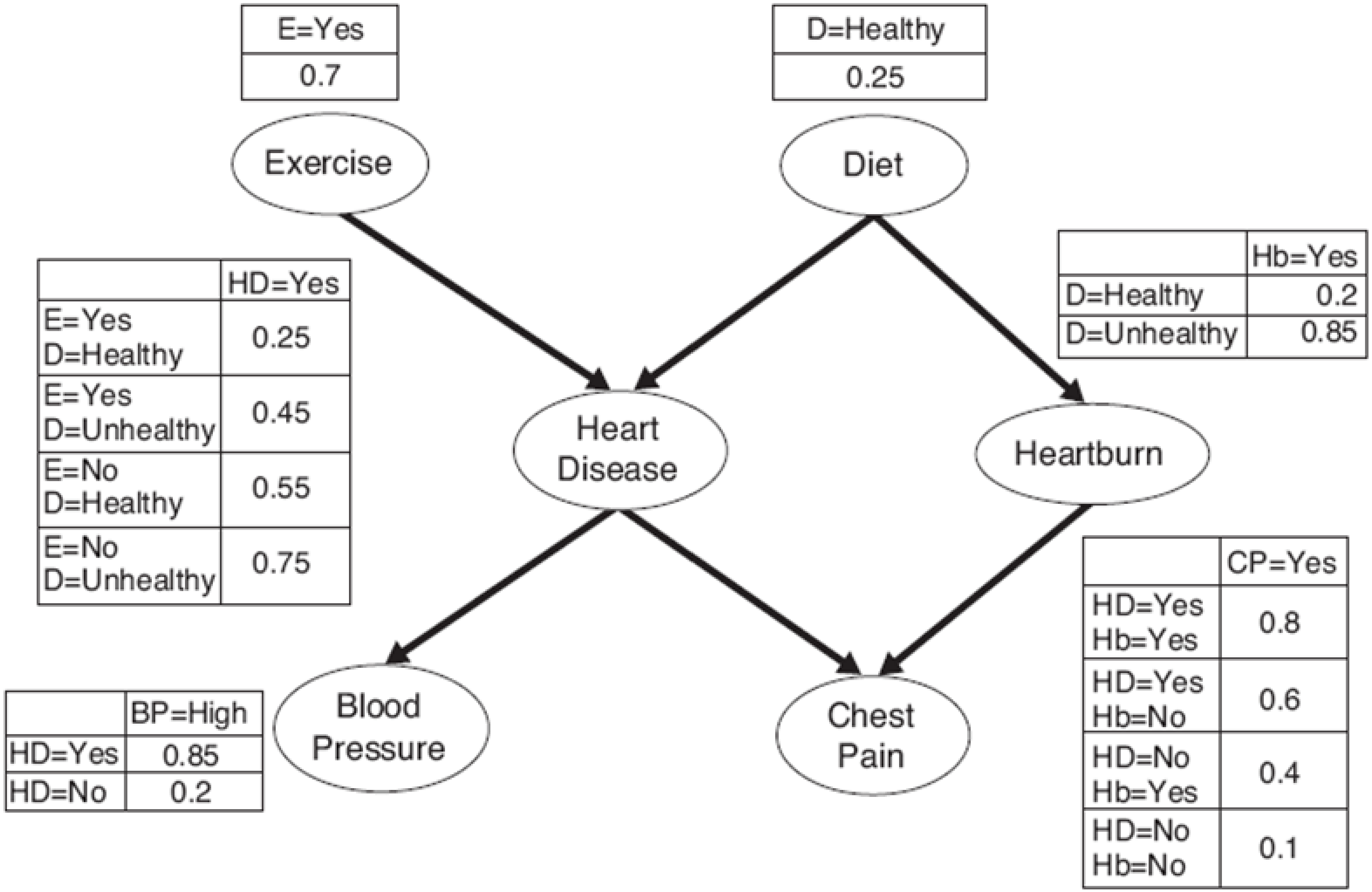
	P(H1|H)	
H	True	False
True	0.95	0.05
False	0.01	0.99

	P(C1|C)	
C	True	False
True	0.99	0.01
False	0.10	0.90

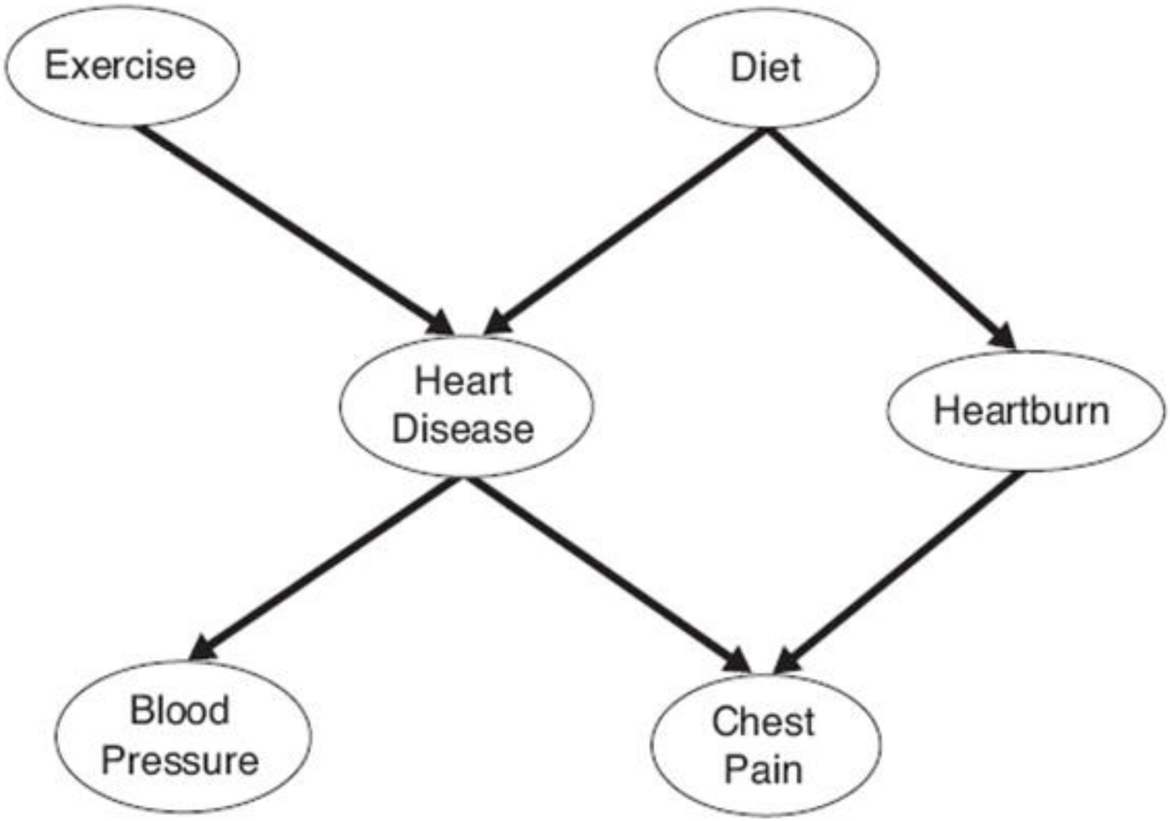
	P(H2|H)	
H	True	False
True	0.98	0.02
False	0.05	0.95

	P(C2|C)	
C	True	False
True	0.98	0.02
False	0.05	0.95

What's about more complex networks?

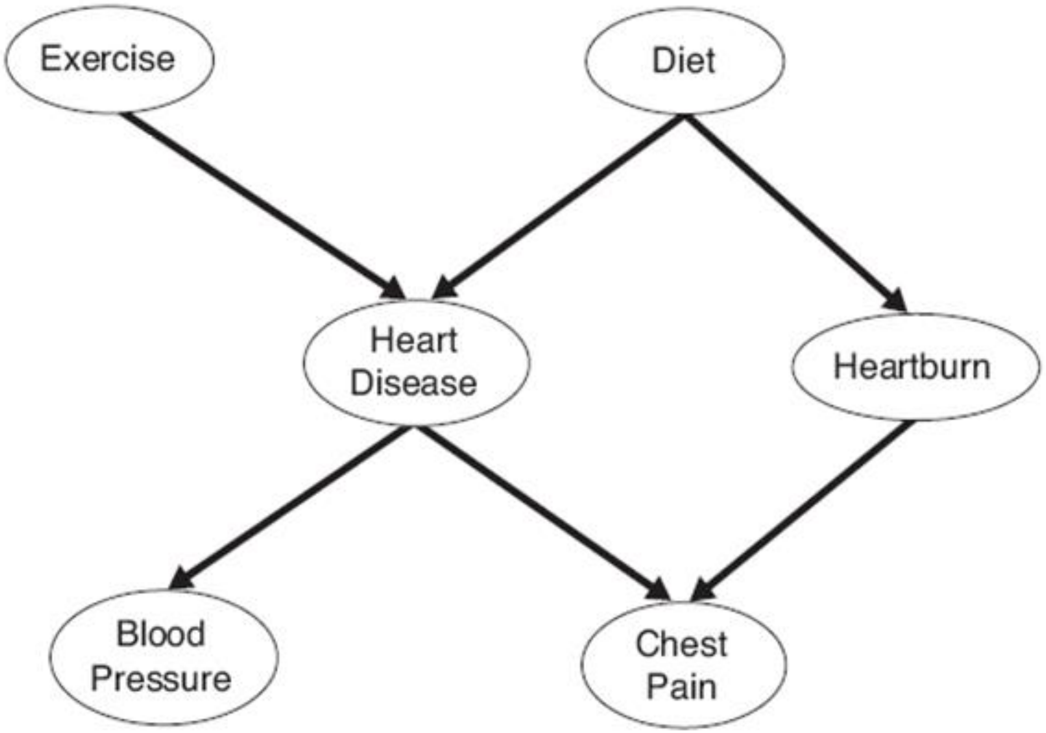


We will study this graph



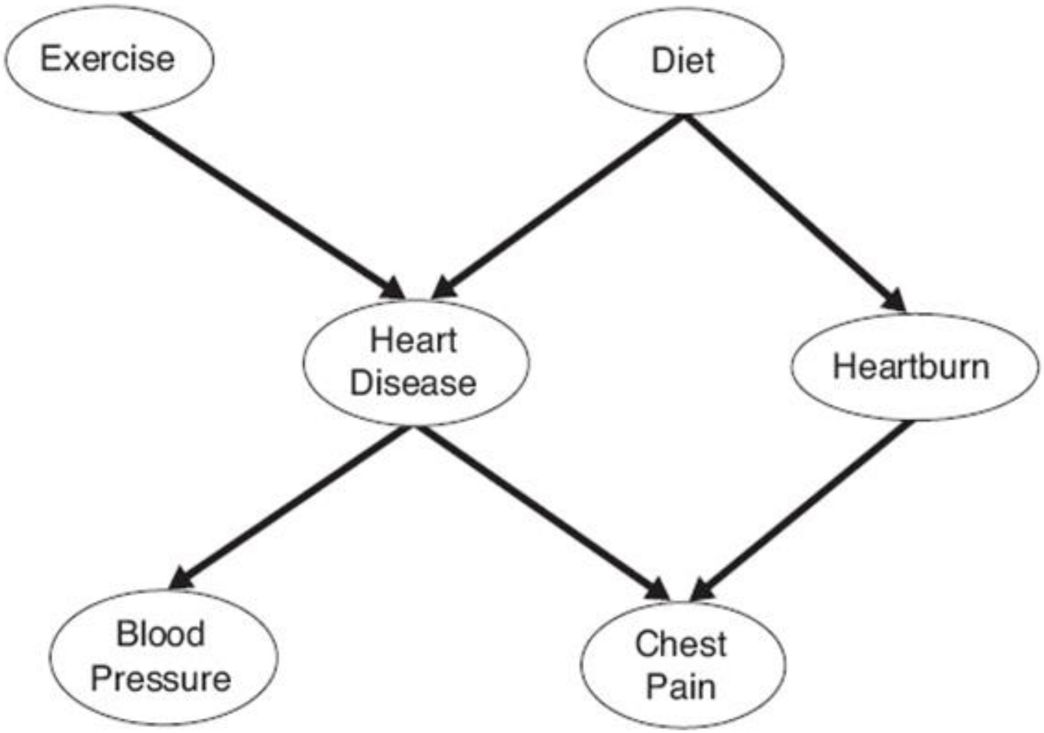
We can show:

- $P(D|E)=P(D)$
- $P(Hb|HD, E, D)= P(Hb|D)$
- $P(CP|Hb, HD, E, D)= P(CP|\mathrm{Hb}, \mathrm{HD})$
- $P(BP|CP, Hb, HD, E, D)= P(BP|\mathrm{HD})$
- However, $P(\mathrm{HD}|\mathrm{E},\mathrm{D})$ cannot be simplified



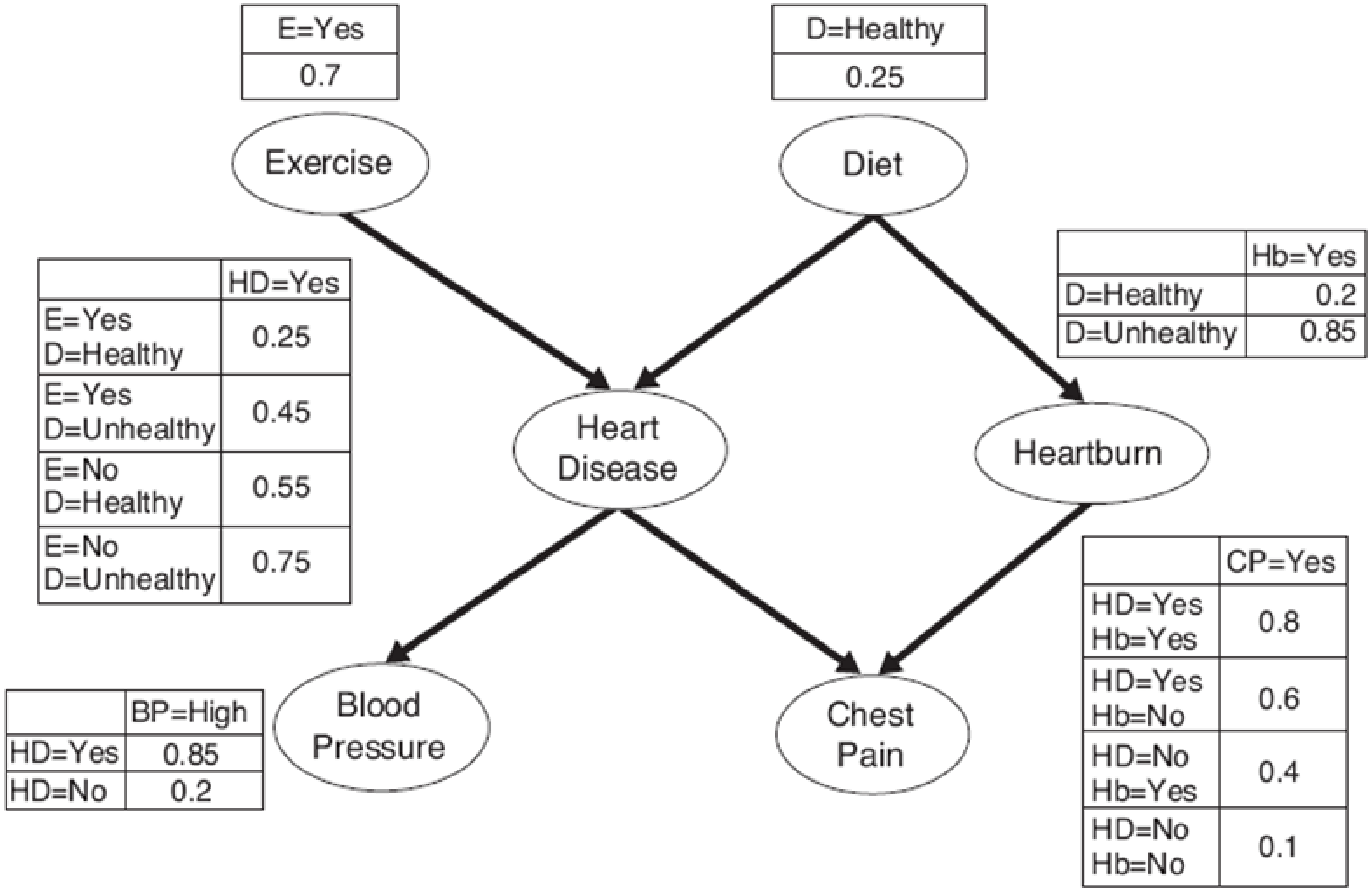
Exercise:

- $P\,(CP|HD, BP, E, D) = ?$



Exercise:

- *P (CP|HD, BP, E, D)*= No simplification



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	0.2
D=Unhealthy	0.85

	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

Calculate P(HD=yes) ?