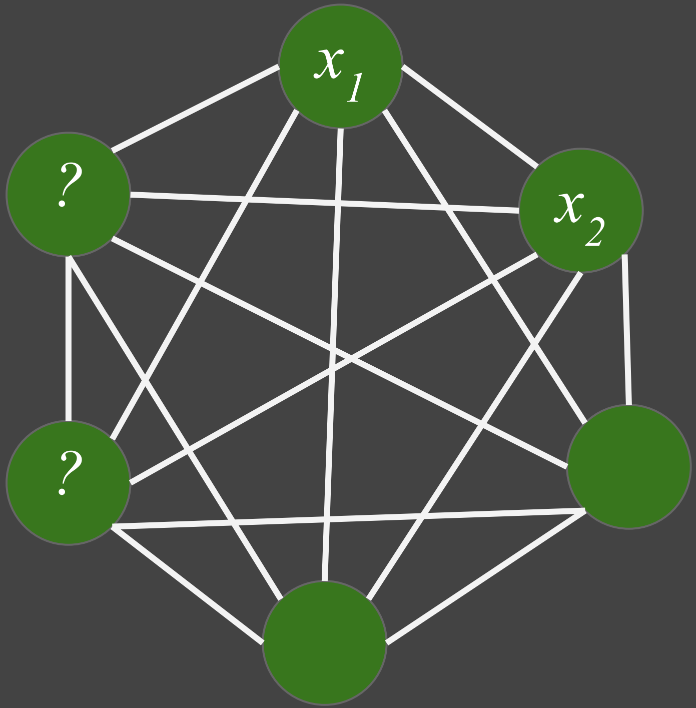
Probabilistic Graphical Models

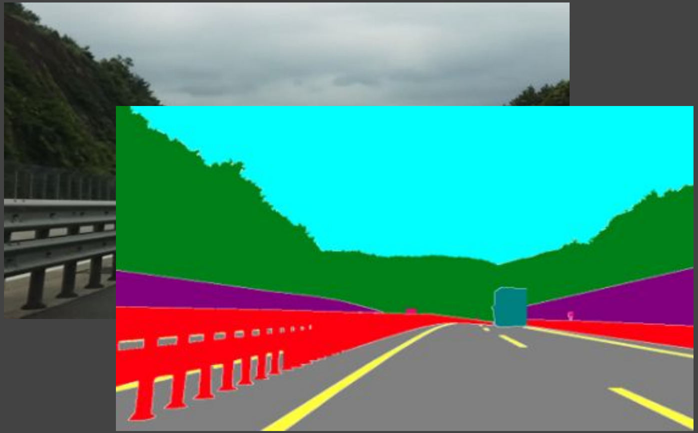
Lecture 2

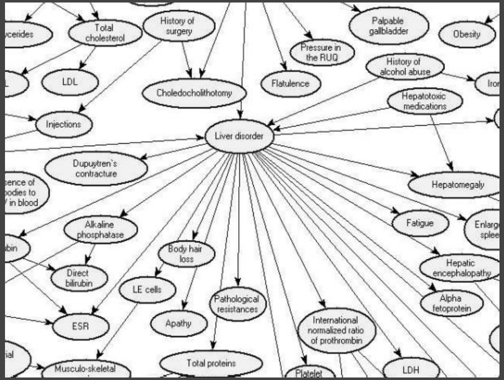
Probabilistic Modeling

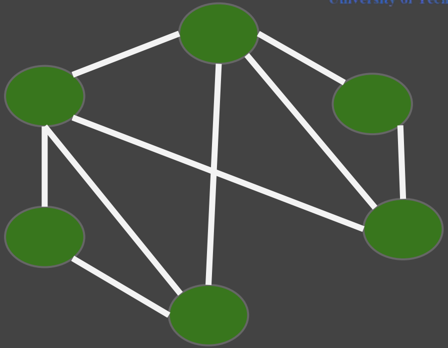
How to model?

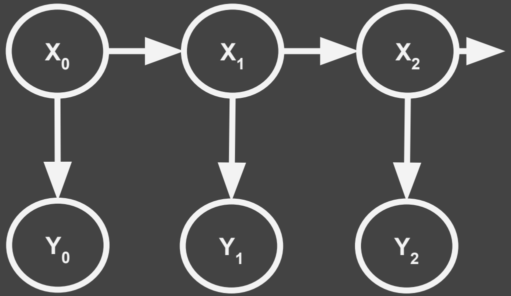


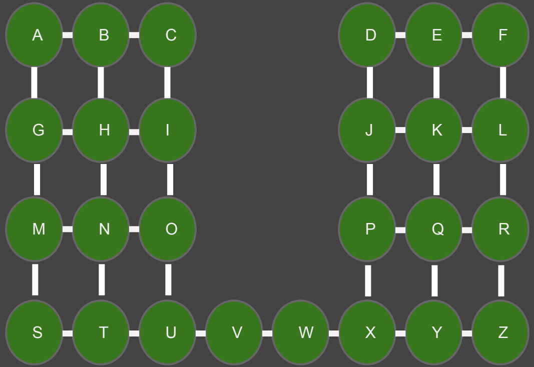
How to model dependency structures?











Mathematical Tools for modeling

K. N. Toosi
University of Technology

- Variables
- Functions, Time series, Signals & Systems
- (ordinary, differential, integral) equations

Example:

$$\mathbf{x} = \frac{\mathbf{f}}{2m} t^2 + \mathbf{v}_0 t + \mathbf{x}_0$$

Uncertainty

- Problems including uncertainty, hidden variables,
- Unknown factors,
- All factors cannot be exactly measured,
- Complex factors or relations that are too hard to model

Example:

$$\mathbf{x} = \frac{\mathbf{f}}{2m} t^2 + \mathbf{v}_0 t + \mathbf{x}_0$$

where

$$\mathbf{f} = \mathbf{f}_{\text{known}} + \mathbf{f}_{\text{wind}}$$

Probabilistic Models

K. N. Toosi
University of Technology

- Model probabilistic uncertainty/randomness
- Random Variables
- Distributions

Random Variables

- Profit = Revenue – Cost
  - Revenue = f(n)
  - Cost = g(n)
  - n : number of packages produced per day
- Ordinary (deterministic) Variables:
  - n = 244
- Random Variables:
  - n = 243 with a probability of 1/6
  - n = 244 with a probability of 3/6
  - n = 245 with a probability of 2/6

Random Variables

- Ordinary (deterministic) Variables:
  - n = 244
- Random Variables:
  - n = 243 with a probability of 1/6
  - n = 244 with a probability of 3/6
  - n = 245 with a probability of 2/6
- Mathematical Operations
  - addition, subtraction, multiplication, etc.
  - functions
  - reasoning

Representing Random Variables

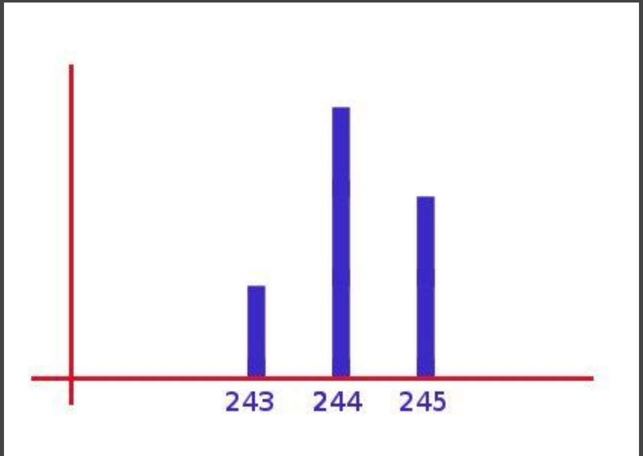
- Random Variables:
    - n = 243 with a probability of 1/6
    - n = 244 with a probability of 3/6
    - n = 245 with a probability of 2/6

- Probability Distribution
- Probability Mass Function
    - p = {(243, 1/6), (244, 3/6), (245, 2/6)}
    - p(243) = 1/6

Representing Random Variables

- Random Variables:
  - n = 243 with a probability of 1/6
  - n = 244 with a probability of 3/6
  - n = 245 with a probability of 2/6

- Probability Distribution
- Probability Mass Function
  - p = {(243, 1/6), (244, 3/6), (245, 2/6)}
  - p(243) = Pr(N=243) = 1/6



Representing Random Variables

K. N. Toosi
University of Technology

- Random Variables:
  - n = 243 with a probability of 1/6
  - n = 244 with a probability of 3/6
  - n = 245 with a probability of 2/6

- Ordinary variables as special case of random variables:

Representing Random Variables

- Random Variables:
  - n = 243 with a probability of 1/6
  - n = 244 with a probability of 3/6
  - n = 245 with a probability of 2/6

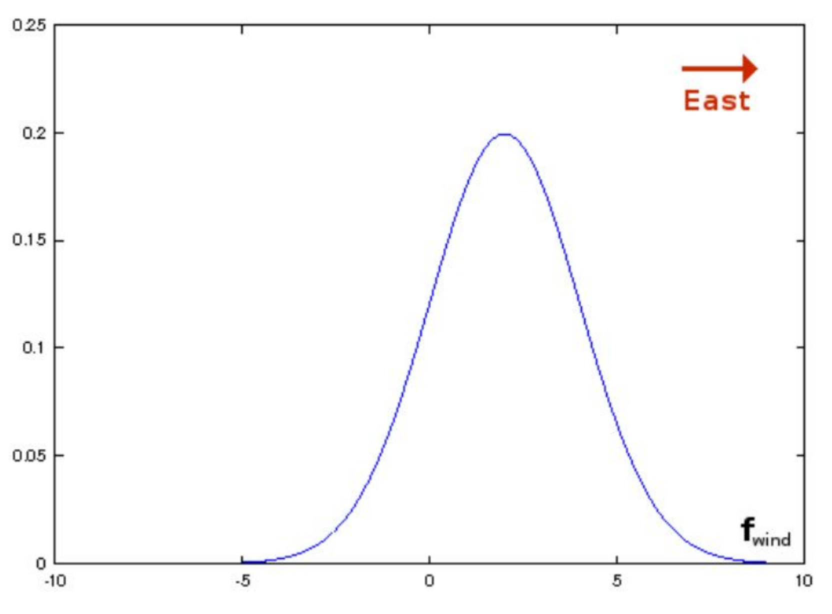
- Ordinary variables as special case of random variables:
  - n = 243 with a probability of 0
  - n = 244 with a probability of 1
  - n = 245 with a probability of 0

Continuous Random Variables

- $\mathbf{x} = \frac{\mathbf{f}}{2m} t^2 + \mathbf{v}_0 t + \mathbf{x}_0$
- $\mathbf{f} = \mathbf{f}_{\text{known}} + \mathbf{f}_{\text{wind}}$
- $\mathbf{f}_{\text{wind}} = ?$

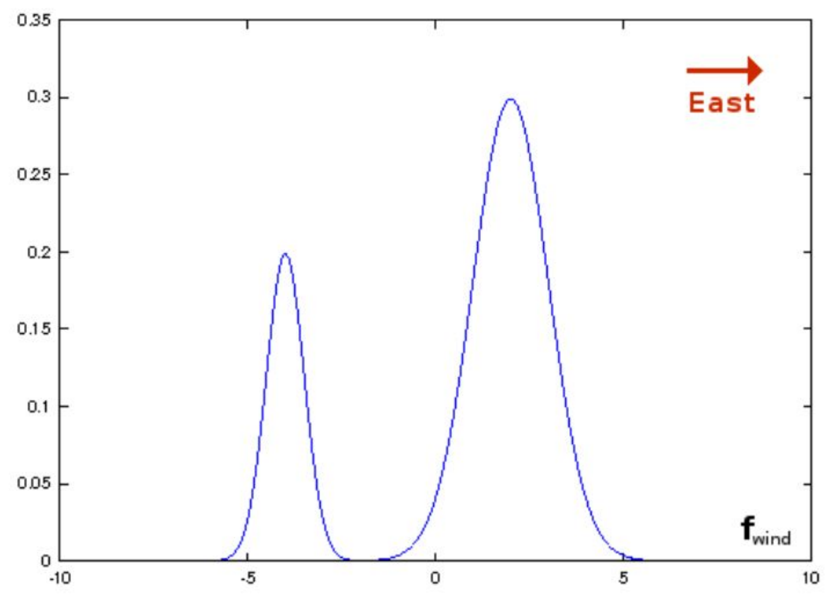
Continuous Random Variables

- $\mathbf{x} = \frac{\mathbf{f}}{2m} t^2 + \mathbf{v}_0 t + \mathbf{x}_0$
- $\mathbf{f} = \mathbf{f}_{\text{known}} + \mathbf{f}_{\text{wind}}$
- $\mathbf{f}_{\text{wind}} = ?$



Continuous Random Variables

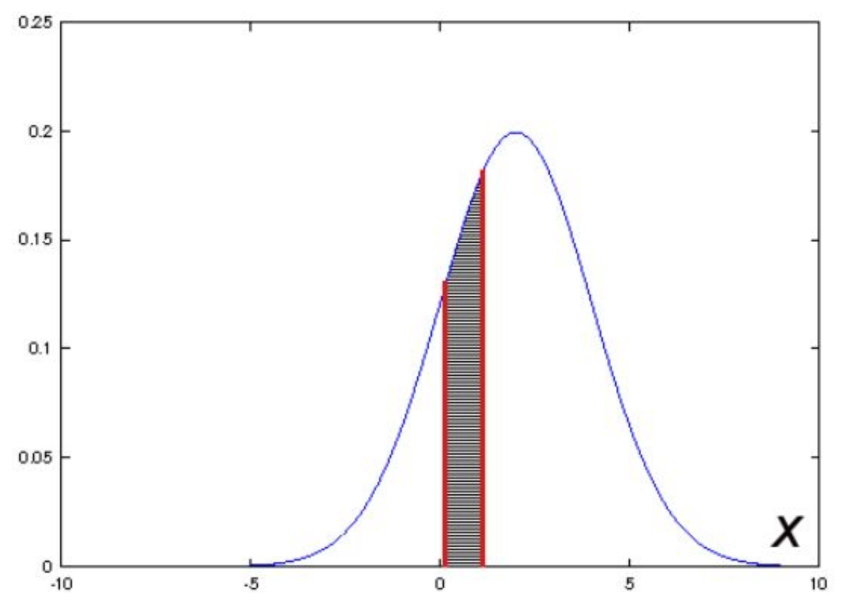
- $\mathbf{x} = \frac{\mathbf{f}}{2m} t^2 + \mathbf{v}_0 t + \mathbf{x}_0$
- $\mathbf{f} = \mathbf{f}_{\text{known}} + \mathbf{f}_{\text{wind}}$
- $\mathbf{f}_{\text{wind}} = ?$



Continuous Random Variables

- the probability of a variable being in a certain interval is of interest,
- Probability Density Function (PDF)

$$\Pr(\mathrm{a} \leq \mathrm{x} \leq \mathrm{b}) = \int_{x\in(a,b)} p(x)dx$$

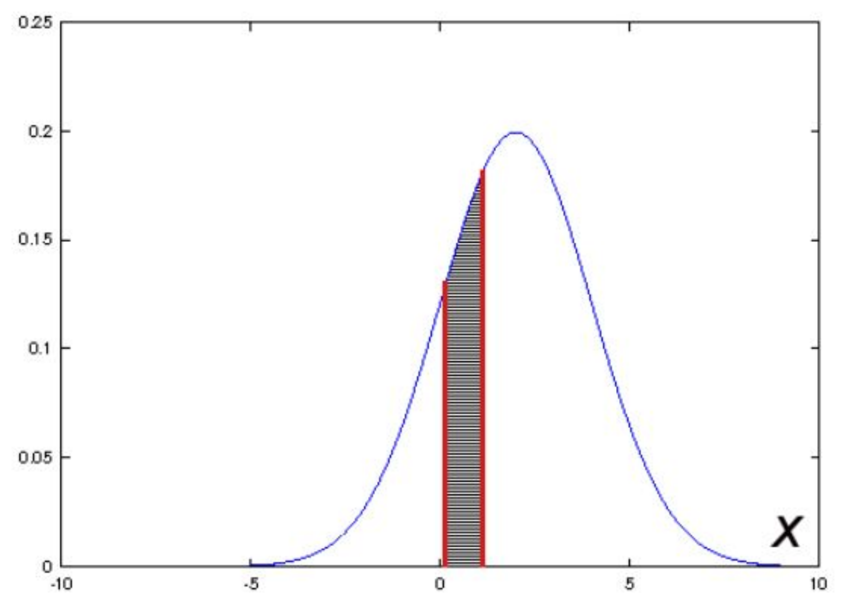


Continuous Random Variables

- In continuous case, mostly the probability of a variable being in a certain interval is of interest,
- Probability Density Function (PDF)

$$\Pr(\mathrm{a} \leq \mathrm{x} \leq \mathrm{b}) = \int_{x \in (a,b)} p(x) dx$$

$$\Pr(x \in S) = \int_S p(x) dx$$



The joint probability distribution

- Question
  - $Pr(X = a) = 1/2$
    $Pr(Y = b) = 1/4$
  - what is $Pr(X = a$ AND $Y = b)$?

The joint probability distribution

- Question
  - $Pr(X = a) = 1/2$
    $Pr(Y = b) = 1/4$
  - what is $Pr(X = a$ AND $Y = b)$?

First Scenario: Roll a (fair) dice twice

- X: first number, Y: second number
- $Pr(X = 6$ AND $Y = 1) = ?$

The joint probability distribution

K. N. Toosi
University of Technology

- Question
    - Pr(X = a) = 1/2
      Pr(Y = b) = 1/4

    - what is Pr(X = a **AND** Y = b)?

First Scenario: Roll a (fair) dice twice

- X: first number, Y: second number
- Pr(X = 6 **AND** Y = 1) = Pr(X = 6) P(Y = 1) = ⅙ * ⅙ = 1/36

The joint probability distribution

K. N. Toosi University of Technology

- Question
  - Pr(X = a) = 1/2
    Pr(Y = b) = 1/4
  - what is Pr(X = a AND Y = b)?

Second Scenario (missing exam => failing course)

- Pr(miss exam session) = 0.01
  Pr(fail course) = 0.08
- Pr(miss exam session AND fail course) = ?

The joint probability distribution

K. N. Toosi
University of Technology

- Question
  - Pr(X = a) = 1/2
    Pr(Y = b) = 1/4
  - what is Pr(X = a AND Y = b)?

Second Scenario (missing exam => failing course)

- Pr(miss exam session) = 0.01
  Pr(fail course) = 0.08
- Pr(miss exam session AND fail course) = 0.01

The joint probability distribution

- Question
  - Pr(X = a) = 1/2
    Pr(Y = b) = 1/4
  - what is Pr(X = a AND Y = b)?

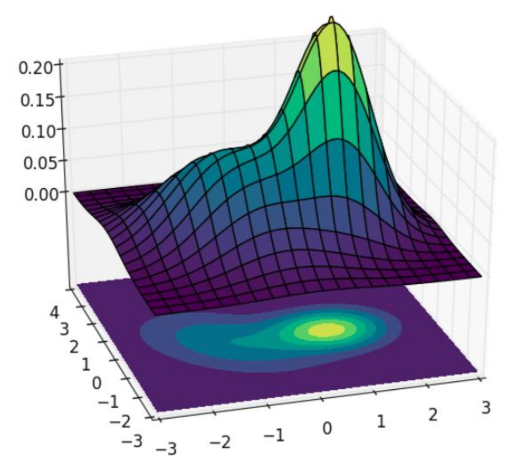
Second Scenario (missing exam => failing course)

- Pr(miss exam session) = 0.01
  Pr(fail course) = 0.08
- Pr(miss exam session AND fail course) = 0.01

How the first and second scenarios are different?

The joint probability distribution

- The probability of co-occurrence.
- If we have two random variables X,Y we cannot model the system using Pr(X=x) and Pr(Y=y).
- Probability mass function (Discrete Variables)
  - p(x,y) = Pr(X=x AND Y=y) = Pr(X=x, Y=y)
- Probability Density function (Continuous Variables)
  - p(x,y)

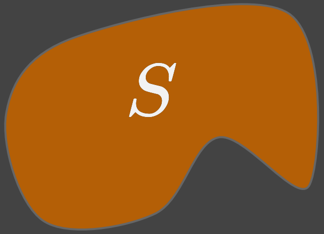


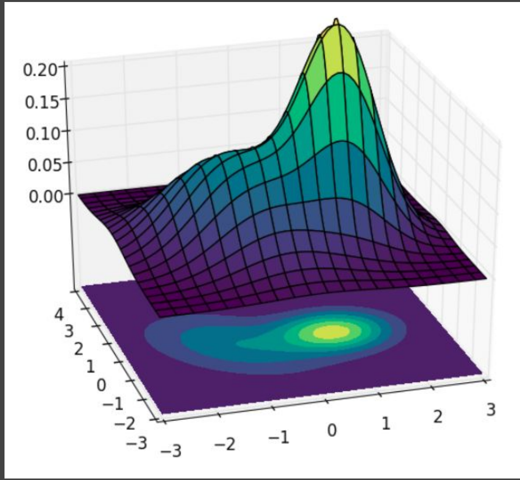
The joint probability distribution

K. N. Toosi
University of Technology

- Probability Density function (Continuous Variables)
  - p(x,y)

$$\Pr((x,y) \in S) = \int_S p(x,y)\,dx\,dy$$





Probabilistic Modelling

K. N. Toosi University of Technology

- System variables $X_1, X_2, \dots, X_N$
- Generative Model: Joint distribution $p(x_1, x_2, \dots, x_N)$
- If you have the joint distribution, you have everything

- Prediction:
  - Having $p(x,y,z) = Pr(X=x, Y=y, Z=z)$, predict x,y,z

Probabilistic Modelling

K. N. Toosi
University of Technology

- System variables $X_1, X_2, \dots, X_N$
- Generative Model: Joint distribution $p(x_1, x_2, \dots, x_N)$
- If you have the joint distribution, you have everything

- Prediction:
  - Having $p(x,y,z) = Pr(X=x, Y=y, Z=z)$, predict x,y,z
  - Find the most likely configuration of system variables

Probabilistic Modelling

K. N. Toosi
University of Technology

- System variables $X_1, X_2, \ldots, X_N$
- Generative Model: Joint distribution $p(x_1, x_2, \ldots, x_N)$
- If you have the joint distribution, you have everything

- Prediction:
  - Having p(x,y,z) = Pr(X=x, Y=y, Z=z), predict x,y,z
  - Find the most likely configuration of system variables

$$x^*, y^*, z^* = \arg\max_{x,y,z} p(x, y, z)$$

Probabilistic Modelling

K. N. Toosi University of Technology

- System variables $X_1, X_2, \ldots, X_N$
- Generative Model: Joint distribution $p(x_1, x_2, \ldots, x_N)$
- If you have the joint distribution, you have everything

- Prediction:
  - Having $p(x,y,z) = Pr(X=x, Y=y, Z=z)$
  - If we know $Z = z_0$, predict x,y

Probabilistic Modelling

K. N. Toosi
University of Technology

- System variables $X_1, X_2, \dots, X_N$
- Generative Model: Joint distribution $p(x_1, x_2, \dots, x_N)$
- If you have the joint distribution, you have everything

- Prediction:
  - Having $p(x,y,z) = Pr(X=x, Y=y, Z=z)$
  - If we know $Z = z_0$, predict x,y

$$x, y = \arg\max_{x,y} p(x, y, z_0)$$

Probabilistic Modelling

1. ?

Probabilistic Modelling

1. learning/modeling:
   - find $p(x_1, x_2, \ldots, x_m, y)$
2.

Probabilistic Modelling

1. learning/modeling:
   - find $p(x_1, x_2, \ldots, x_m, y)$

2. prediction/testing

Probabilistic Modelling

1. learning/modeling:
   - find $p(x_1, x_2, \ldots, x_m, y)$

2. prediction/testing

Probabilistic Modelling

1. learning/modeling:
   - find $p(x_1, x_2, \ldots, x_m, y)$

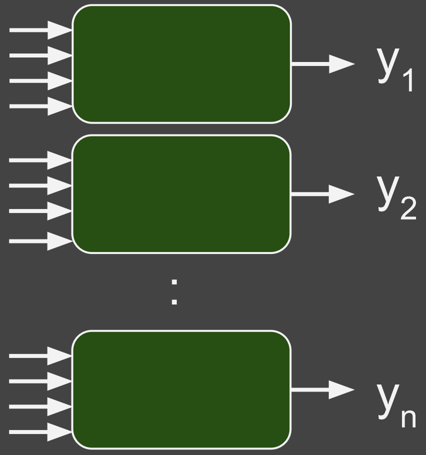
2. prediction/testing

$$y^* = \arg\max_y p(x_1, x_2, \ldots, x_m, y)$$

Probabilistic Modelling

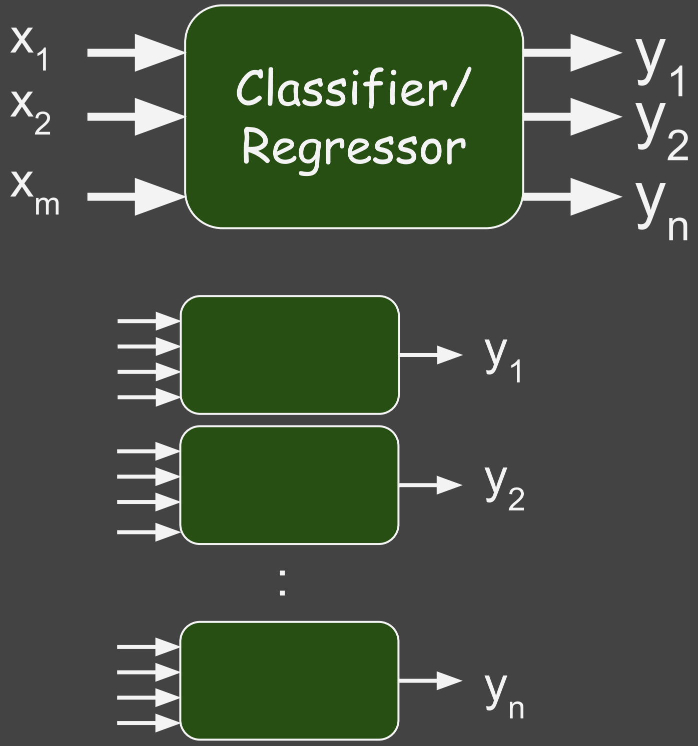
1. learning/modeling:
   -
2. prediction/testing

Probabilistic Modelling



1. learning/modeling:
   -
2. prediction/testing

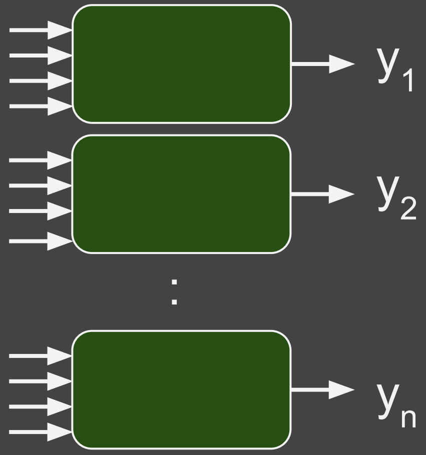
Probabilistic Modelling



1. learning/modeling:
   - find $p(x_1, x_2, \ldots, x_m, y_1)$
   - find $p(x_1, x_2, \ldots, x_m, y_2)$
   - :
   - find $p(x_1, x_2, \ldots, x_m, y_n)$

2. prediction/testing

Probabilistic Modelling



1. learning/modeling:
   - find $p(x_1, x_2, \ldots, x_m, y_1)$
   - find $p(x_1, x_2, \ldots, x_m, y_2)$
   - :
   - find $p(x_1, x_2, \ldots, x_m, y_n)$

2. prediction/testing

$$y_1^* = \arg\max_{y_1} p(x_1, \ldots, x_m, y_1)$$
$$y_2^* = \arg\max_{y_2} p(x_1, \ldots, x_m, y_2)$$
$$y_n^* = \arg\max_{y_n} p(x_1, \ldots, x_m, y_n)$$

Probabilistic Modelling

1. learning/modeling:
   -
2. prediction/testing

Probabilistic Modelling

1. learning/modeling:
    - find $p(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n)$
2. prediction/testing

Probabilistic Modelling

1. learning/modeling:
   - find $p(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n)$
2. prediction/testing

$$y_1^*, y_2^*, \ldots, y_n^* = \arg\max_{y_1,\ldots,y_n} p(x_1, \ldots, x_m, y_1, \ldots, y_n)$$

Probabilistic Modelling

1. learning/modeling:
   - find $p(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n)$
2. prediction/testing

$$y_1^*, y_2^*, \ldots, y_n^* = \arg\max_{y_1,\ldots,y_n} p(x_1, \ldots, x_m, y_1, \ldots, y_n)$$

$y_i \in \{0,1\}$, $n = 100$

-

Probabilistic Modelling

1. learning/modeling:
   - find $p(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n)$
2. prediction/testing

$$y_1^*, y_2^*, \ldots, y_n^* = \arg\max_{y_1,\ldots,y_n} p(x_1, \ldots, x_m, y_1, \ldots, y_n)$$

$y_i \in \{0,1\}$, $n = 100$

- argmax over $2^{100} \approx 10^{30}$ different combinations of $y_1, y_2, \ldots, y_n$
-

Probabilistic Modelling

K. N. Toosi
University of Technology

1. learning/modeling:
   - find $p(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n)$
2. prediction/testing

$$y_1^*, y_2^*, \ldots, y_n^* = \arg\max_{y_1, \ldots, y_n} p(x_1, \ldots, x_m, y_1, \ldots, y_n)$$

$y_i \in \{0,1\}$, $n = 100$

- argmax over $2^{100} \approx 10^{30}$ different combinations of $y_1, y_2, \ldots, y_n$
- 10^{12} = 1000G iterations per second $\rightarrow$ 32 billion years!

Model Representation

$$p(x_1, x_2, \ldots, x_m) \qquad x_i \in 0, 1$$

- How to store in computer?

Tabular representation

K. N. Toosi
University of Technology

$$p(x_1, x_2, \ldots, x_m) \qquad x_i \in 0, 1$$

- m = 3 ---> 8 entries
-

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.10
0	0	1	0.01
0	1	0	0.08
0	1	1	0.21
1	0	0	0.30
1	0	1	0.14
1	1	0	0.14
1	1	1	0.02

Tabular representation

$$p(x_1, x_2, \ldots, x_m) \qquad x_i \in 0, 1$$

- m = 3 ---> 8 entries
- 2^m entries
-

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.10
0	0	1	0.01
0	1	0	0.08
0	1	1	0.21
1	0	0	0.30
1	0	1	0.14
1	1	0	0.14
1	1	1	0.02

K. N. Toosi
University of Technology

Tabular representation

$$p(x_1, x_2, \ldots, x_m) \qquad x_i \in 0, 1$$

- m = 3 ---> 8 entries
- 2^m entries
- m = 100 --->

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.10
0	0	1	0.01
0	1	0	0.08
0	1	1	0.21
1	0	0	0.30
1	0	1	0.14
1	1	0	0.14
1	1	1	0.02

K. N. Toosi University of Technology

Tabular representation

$$p(x_1, x_2, \ldots, x_m) \qquad x_i \in 0, 1$$

- m = 3 ---> 8 entries
- 2^m entries
- m = 100 ---> 2^{100} entries ≈ 10^{21} G entries

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.10
0	0	1	0.01
0	1	0	0.08
0	1	1	0.21
1	0	0	0.30
1	0	1	0.14
1	1	0	0.14
1	1	1	0.02

K. N. Toosi
University of Technology

Tabular representation

$$p(x_1, x_2, \ldots, x_m) \qquad x_i \in 0, 1$$

- m = 3 ---> 8 entries
- 2^m entries
- m = 100 ---> 2^{100} entries ≈ 10^{21} G entries
- 1 byte per entry -> 10^{21} GB of RAM!

x_1	x_2	x_3	$p(x_1,x_2,x_3)$
0	0	0	0.10
0	0	1	0.01
0	1	0	0.08
0	1	1	0.21
1	0	0	0.30
1	0	1	0.14
1	1	0	0.14
1	1	1	0.02

K. N. Toosi
University of Technology