Mathematics for AI

Lecture 21

Joint distribution, Marginal Distribution, Conditional Distribution, Probabilistic Modeling, Generative vs Discriminative Models

Adding two random variables (general case)

K. N. Toosi
University of Technology

X: How many Mehran sells

Y: How " Milad sells

Z: How many both sell

$Z = X + Y = ?$ (what if they are not independent?)

$Z = f(X, Y) = ?$ p(X) & p(Y) won't help!

Y \ X	0	1	2
0	0.05	0.06	0.1
1	0.12	0.03	0.2
2	0.07	0.3	0.07

احتمال توأمان

Joint Distribution

$p(x, y)$

$X=0, Y=0 \quad p=0.05$

$X=1, Y=0 \quad p=0.0.6$

⋮

$X=2, Y=2 \quad p=0.07$

$Z = X + Y$

$$Z = \begin{cases} 0 & p = 0.05 \\ 1 & p = 0.06 + 0.12 \\ 2 & p = 0.1 + 0.03 + 0.07 \\ 3 & p = 0.3 + 0.2 \\ 4 & p = 0.07 \end{cases}$$

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

- 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 $Y = 1) = Pr(X = 6)\,P(Y = 1) = 1/6 * 1/6 = 1/36$

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) = ?

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

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

How the first and second scenarios differ?

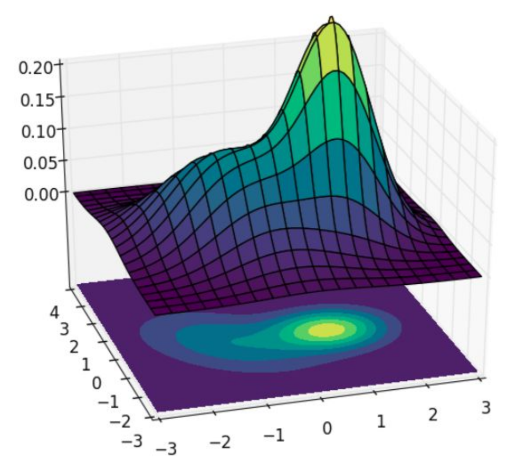
The joint probability distribution

K. N. Toosi University of Technology

- 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)

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)



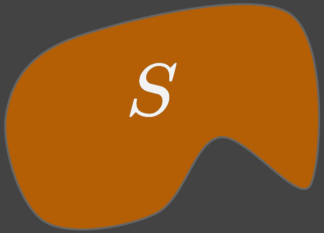
- Probability Density function (Continuous Variables)
  - p(x,y)

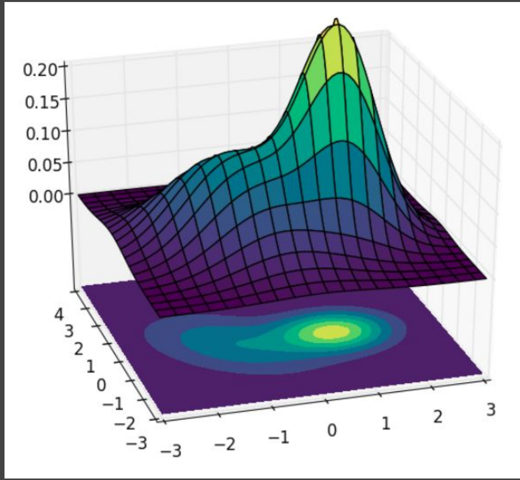
The joint probability distribution

K. N. Toosi
University of Technology

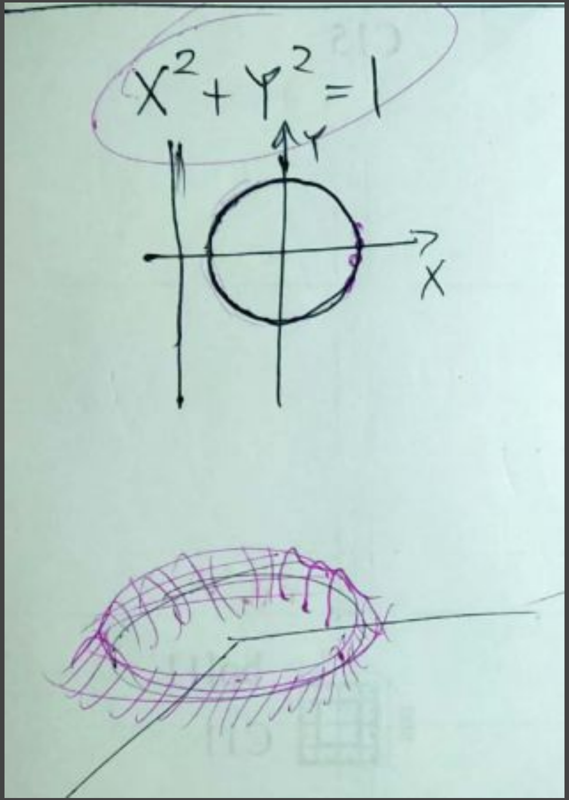
- Probability Density function (Continuous Variables)
  - p(x,y)

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





Generalize Relations



Probabilistic Modelling

K. N. Toosi
University of Technology

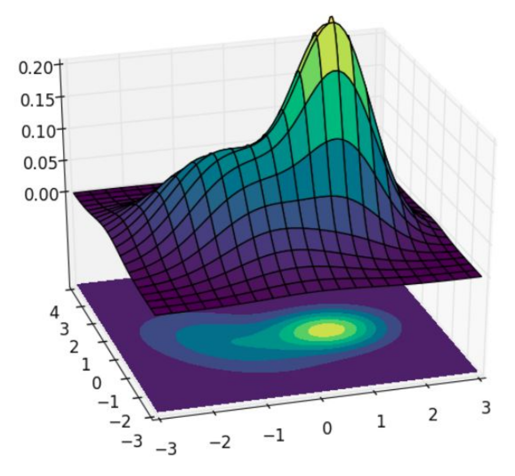
- System variables $X_1, X_2, \dots, X_N$
- Generative Model: Joint distribution $p(x_1, x_2, \dots, x_N)$
- If you have the joint distribution, you have everything

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

Remember: The joint probability distribution

K. N. Toosi
University of Technology

- The probability of co-occurrence.
- 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)

Remember: 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

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

Remember: 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

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

Generative Model

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

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

Example:

The joint probability of

- **having a rainfall in an hour**, and
- **the sky being cloudy at the moment**
  - p(r,c) = Pr(R = r, C = c)

r (rain)	c (cloudy)	$\Pr(R = r, C = c)$
0	0	0.75
0	1	0.10
1	0	0.05
1	1	0.10

Example:

K. N. Toosi
University of Technology

The joint probability of

- **having a rainfall in an hour** and
- **the sky being cloudy at the moment**
    - p(r,c) = Pr(R = r, C = c)

r (rain)	c (cloudy)	$Pr(R = r, C = c)$
0	0	0.75
0	1	0.10
1	0	0.05
1	1	0.10

Pr(R=r, C = c)

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

Question

K. N. Toosi
University of Technology

- Having the joint distribution $Pr(R = r, C = c)$, what is $Pr(R = r)$?

$$Pr(R = r) = ?$$

Question

- Having the joint distribution $Pr(R = r, C = c)$, what is $Pr(R = r)$?

$$Pr(R = r) = Pr\left((R = r \text{ AND } C = 0) \text{ OR } (R = r \text{ AND } C = 1) \right)$$

Question

K. N. Toosi
University of Technology

- Having the joint distribution $Pr(R = r, C = c)$, what is $Pr(R = r)$?

$$Pr(R = r) = Pr\ (\ (R = r \text{ AND } C = 0) \text{ OR } (R = r \text{ AND } C = 1)\ )$$

$$= Pr(R = r \text{ AND } C = 0) + Pr(R = r \text{ AND } C = 1) \qquad \text{(why?)}$$

Question

- Having the joint distribution $Pr(R = r, C = c)$, what is $Pr(R = r)$?

$$Pr(R = r) = Pr\left((R = r \text{ AND } C = 0) \text{ OR } (R = r \text{ AND } C = 1) \right)$$

$$= Pr(R = r \text{ AND } C = 0) + Pr(R = r \text{ AND } C = 1)$$ (why?)

$Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

$Pr(R = r)$

R=0	R=1
0.85	0.15

Marginal Distribution

K. N. Toosi
University of Technology

- Discrete: probability mass function p(m,n) = Pr(M=m, N=n)

$$p(m) = \Pr(M = m) = \sum_n p(m, n)$$

- Continuous: probability density function p(x,y)

$$p(x) = \int p(x, y)\, dy$$

Marginal Probability

K. N. Toosi
University of Technology

$P(x, y)$	$x = 0$	$x = 1$	$x = 2$	row sum
$y = 0$	0.32	0.03	0.01	**0.36**
$y = 1$	0.06	0.24	0.02	**0.32**
$y = 2$	0.02	0.03	0.27	**0.32**
col sum	**0.40**	**0.30**	**0.30**	checksum = 1.0

image from http://stats.stackexchange.com

Marginal Probability

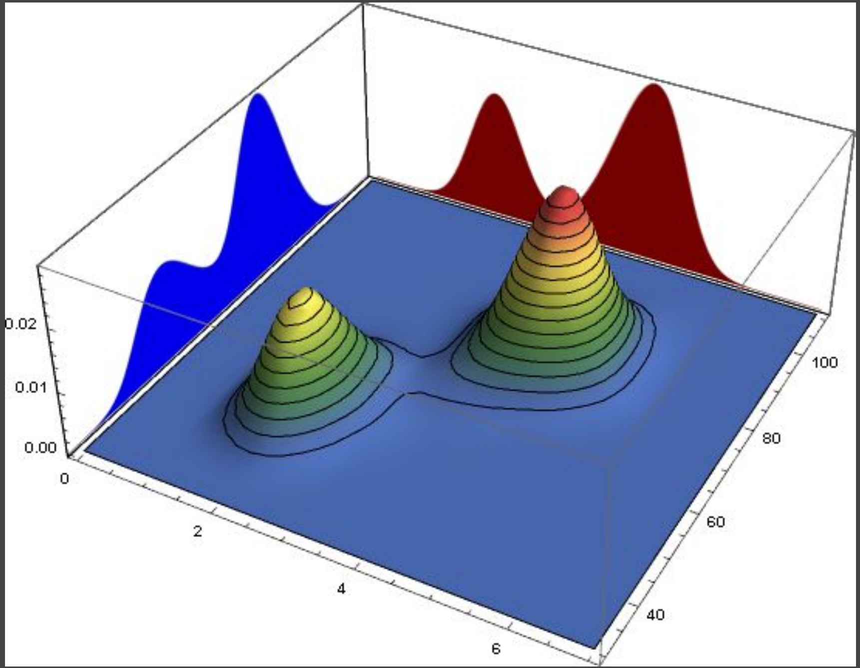


image from www.wolfram.com

Question

- What is the probability of having a rainfall today?

$\Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

Question

- What is the probability of having a rainfall today?

$$Pr(R = 1) = Pr(R = 1, C = 0) + Pr(R = 1, C = 1)$$

$\Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

Question

- What is the probability of having a rainfall today?

$$Pr(R = 1) = Pr(R = 1, C = 0) + Pr(R = 1, C = 1) = 0.05 + 0.10 = 0.15$$

$Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

$Pr(R = r)$

R=0	R=1
0.85	0.15

Question

- What is the probability of having a rainfall today?

$$Pr(R = 1) = Pr(R = 1, C = 0) + Pr(R = 1, C = 1) = 0.05 + 0.10 = 0.15$$

- If we know the sky is cloudy, what is the probability of having a rainfall today?

$Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

Question

- What is the probability of having a rainfall today?

$$Pr(R = 1) = Pr(R = 1, C = 0) + Pr(R = 1, C = 1) = 0.05 + 0.10 = 0.15$$

- If we know the sky is cloudy, what is the probability of having a rainfall today?

$Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

$Pr(R = 1 \mid C = 1) =$

Question

- What is the probability of having a rainfall today?

$$\Pr(R = 1) = \Pr(R = 1, C = 0) + \Pr(R = 1, C = 1) = 0.05 + 0.10 = 0.15$$

- If we know the sky is cloudy, what is the probability of having a rainfall today?

$\Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

$$\Pr(R = 1 \mid C = 1) =$$

Question

- What is the probability of having a rainfall today?

$$Pr(R = 1) = Pr(R = 1, C = 0) + Pr(R = 1, C = 1) = 0.05 + 0.10 = 0.15$$

- If we know the sky is cloudy, what is the probability of having a rainfall today?

$Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

$$Pr(R = 1 \mid C = 1) = 0.10 / (0.10 + 0.10) = 0.5$$

$$= \frac{Pr(R=1, C=1)}{Pr(R=1, C=1) + Pr(R=0, C=1)}$$

Question

- What is the probability of having a rainfall today?

$$Pr(R = 1) = Pr(R = 1, C = 0) + Pr(R = 1, C = 1) = 0.05 + 0.10 = 0.15$$

- If we know the sky is cloudy, what is the probability of having a rainfall today?

$\Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

$$Pr(R = 1 \mid C = 1) = 0.10 / (0.10 + 0.10) = 0.5$$

$$= \frac{\Pr(R{=}1, C{=}1)}{\Pr(R{=}1, C{=}1) + \Pr(R{=}0, C{=}1)}$$

$$= \frac{\Pr(R{=}1, C{=}1)}{\Pr(C{=}1)}$$

Question

K. N. Toosi
University of Technology

- What is the probability of having a rainfall today?

$$Pr(R = 1) = Pr(R = 1, C = 0) + Pr(R = 1, C = 1) = 0.05 + 0.10 = 0.15$$

- If we know the sky is cloudy, what is the probability of having a rainfall today?

$Pr(R = r, C = c)$

	R=0	R=1
C=0	0.75	0.05
C=1	0.10	0.10

$$Pr(R = 1 \mid C = 1) = 0.10 / (0.10 + 0.10) = 0.5$$

$$= \frac{\Pr(R{=}1, C{=}1)}{\Pr(R{=}1, C{=}1) + \Pr(R{=}0, C{=}1)}$$

$$= \frac{\Pr(R{=}1, C{=}1)}{\Pr(C{=}1)}$$

Conditional Distribution

K. N. Toosi
University of Technology

- Discrete: joint PMF p(m,n) = Pr(M=m, N=n)

$$\Pr(N = n_0 \mid M = m) = \frac{\Pr(N{=}n_0, M{=}m)}{\sum_n \Pr(N{=}n, M{=}m)}$$

$$= \frac{\Pr(N{=}n_0, M{=}m)}{\Pr(M{=}m)}$$

- Continuous: joint PDF p(x,y)

Conditional Distribution

K. N. Toosi
University of Technology

- Discrete: joint PMF p(m,n) = Pr(M=m, N=n)

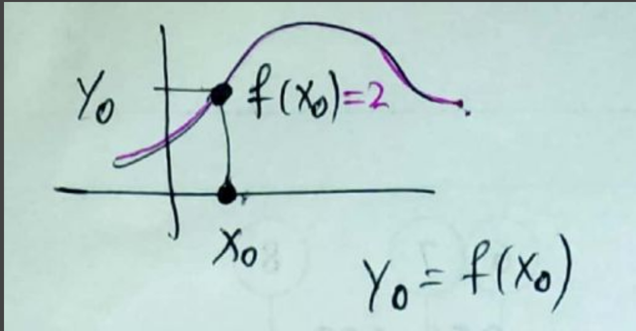
$$\Pr(N = n_0 \mid M = m) = \frac{\Pr(N{=}n_0, M{=}m)}{\sum_n \Pr(N{=}n, M{=}m)}$$

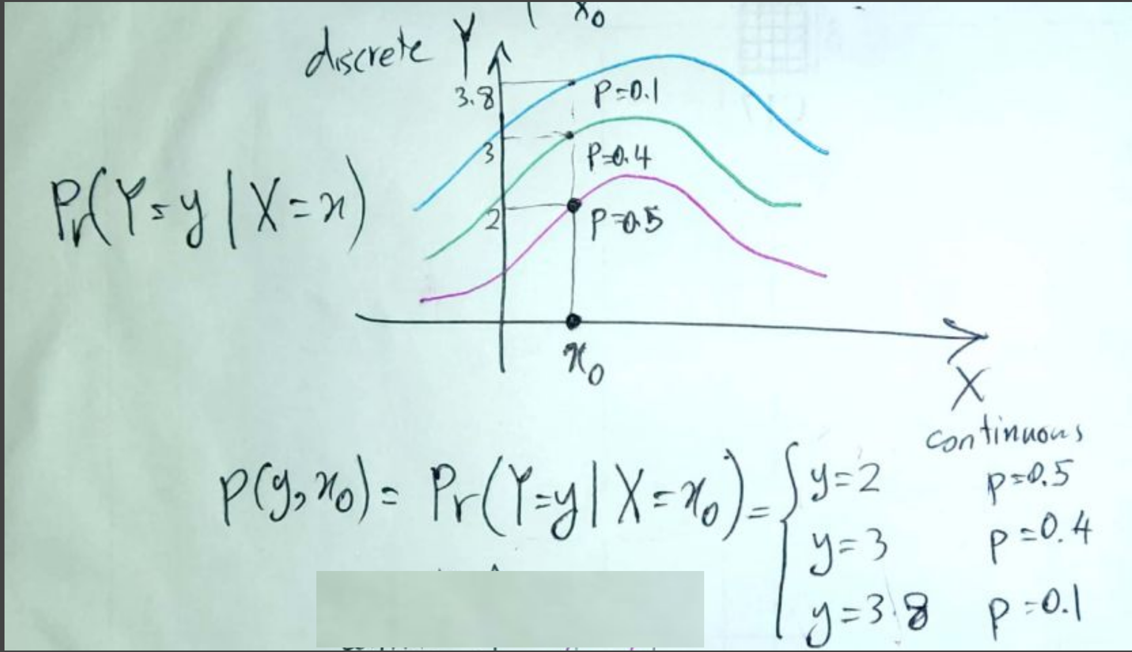
$$= \frac{\Pr(N{=}n_0, M{=}m)}{\Pr(M{=}m)}$$

- Continuous: joint PDF p(x,y)

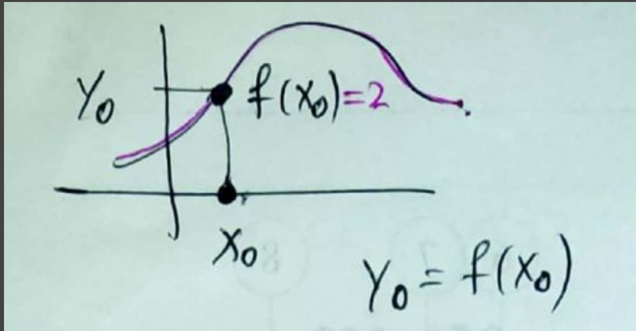
$$p(y \mid x) = \frac{p(x,y)}{\int p(x,y)\, dy} = \frac{p(x,y)}{p(x)}$$

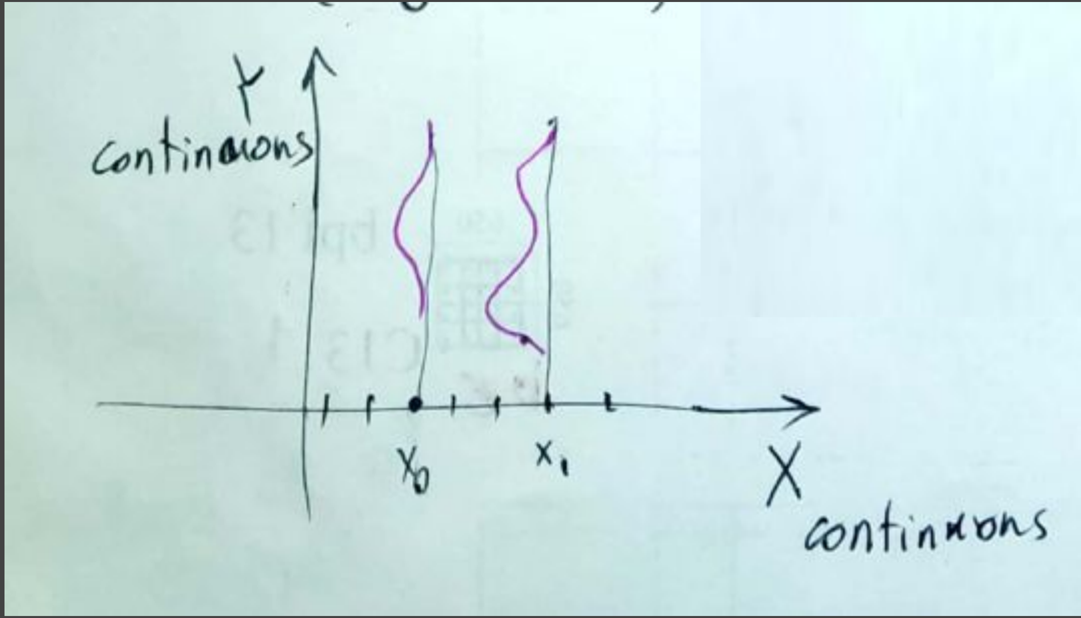
Generalize Functions





Generalize Functions





Discriminative Model

K. N. Toosi
University of Technology

Generative: $p(x_1, x_2, \ldots, x_m, y_1, y_2, \ldots, y_n)$

Discreminative: $p(y_1, y_2, \ldots, y_n \mid x_1, x_2, \ldots, x_m)$

Discriminative Model

K. N. Toosi
University of Technology

Generative: $p(X, Y)$

Discreminative: $p(Y|X)$

$$P(X,Y) = P(Y|X)\,P(X)$$

$$P(Y|X) = \frac{P(X,Y)}{P(X)} = \frac{P(X,Y)}{\sum_{Y'} P(X,Y')}$$