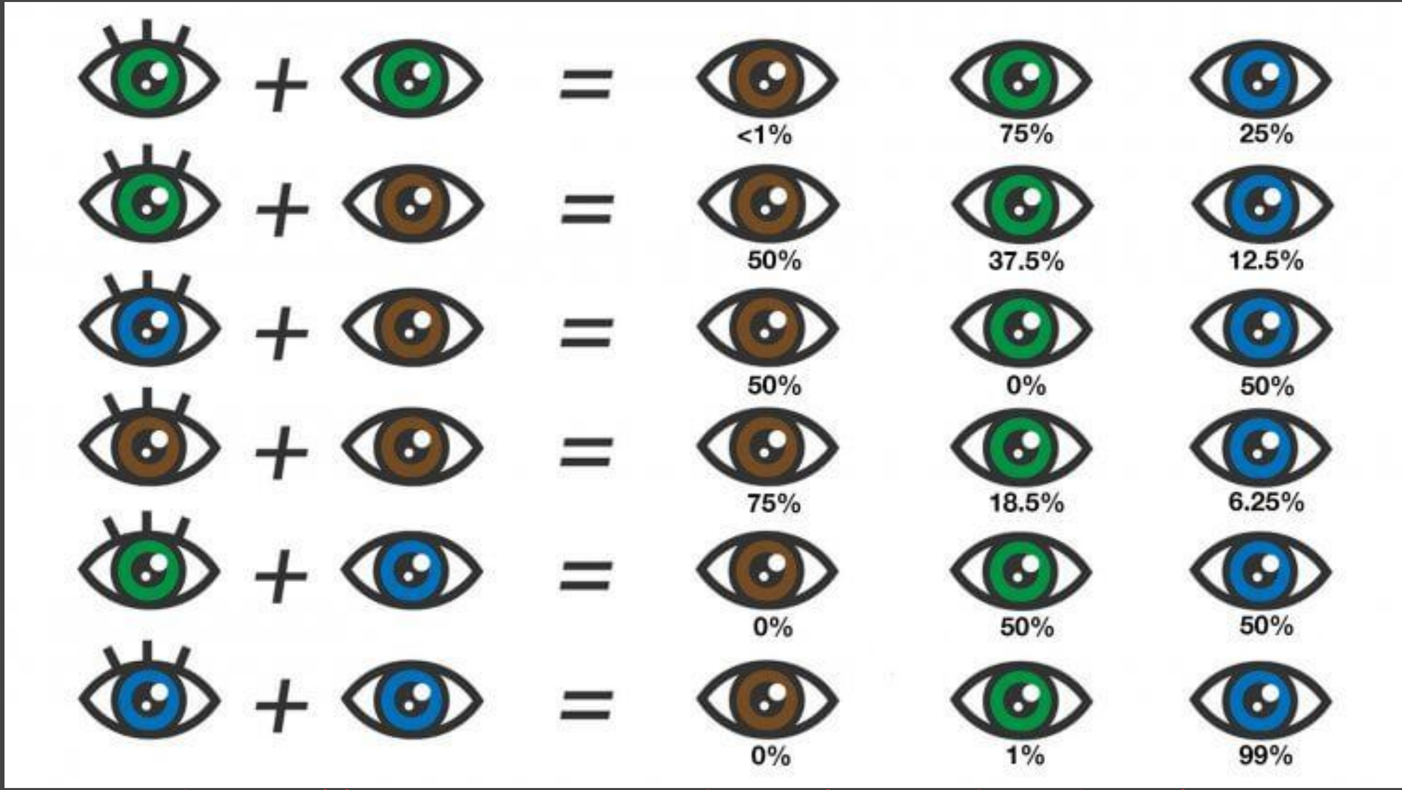
Probabilistic Graphical Models

Lectures 10

Parameter Sharing,
Representing Markov Random Fields

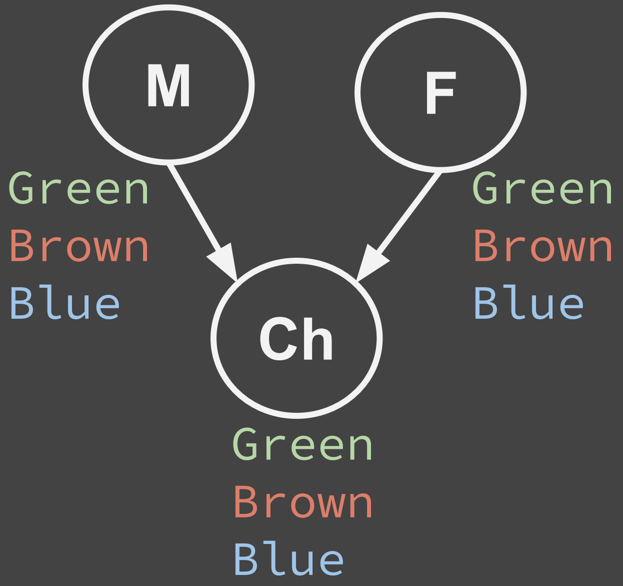
Example: Eye colour inheritance

K. N. Toosi
University of Technology

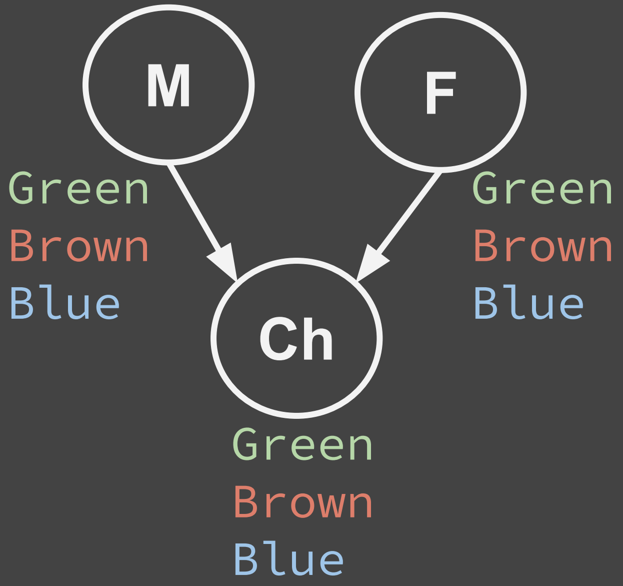


https://www.mamanatural.com/eye-color-chart/

Example: Eye colour inheritance

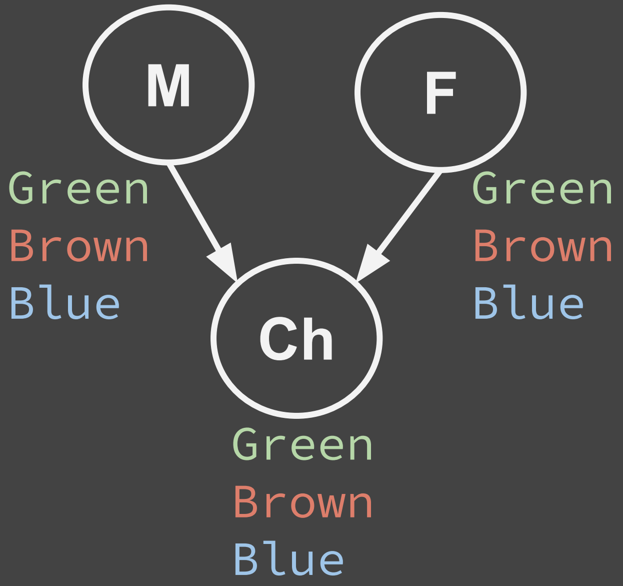


Example: Eye colour inheritance



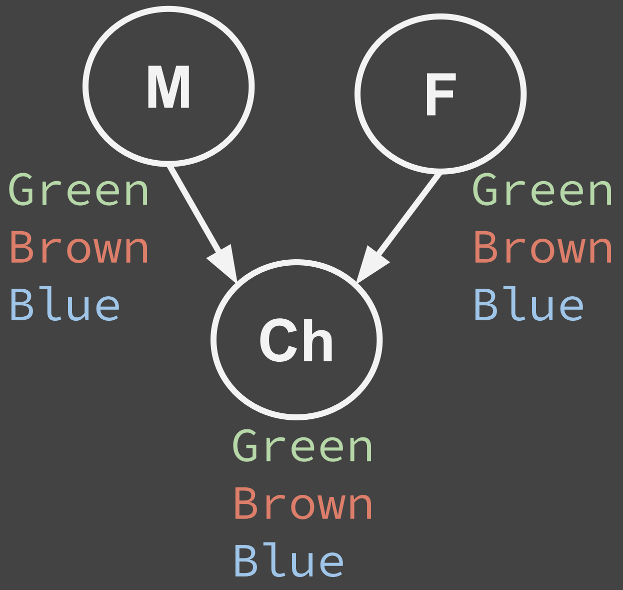
How many parameters?

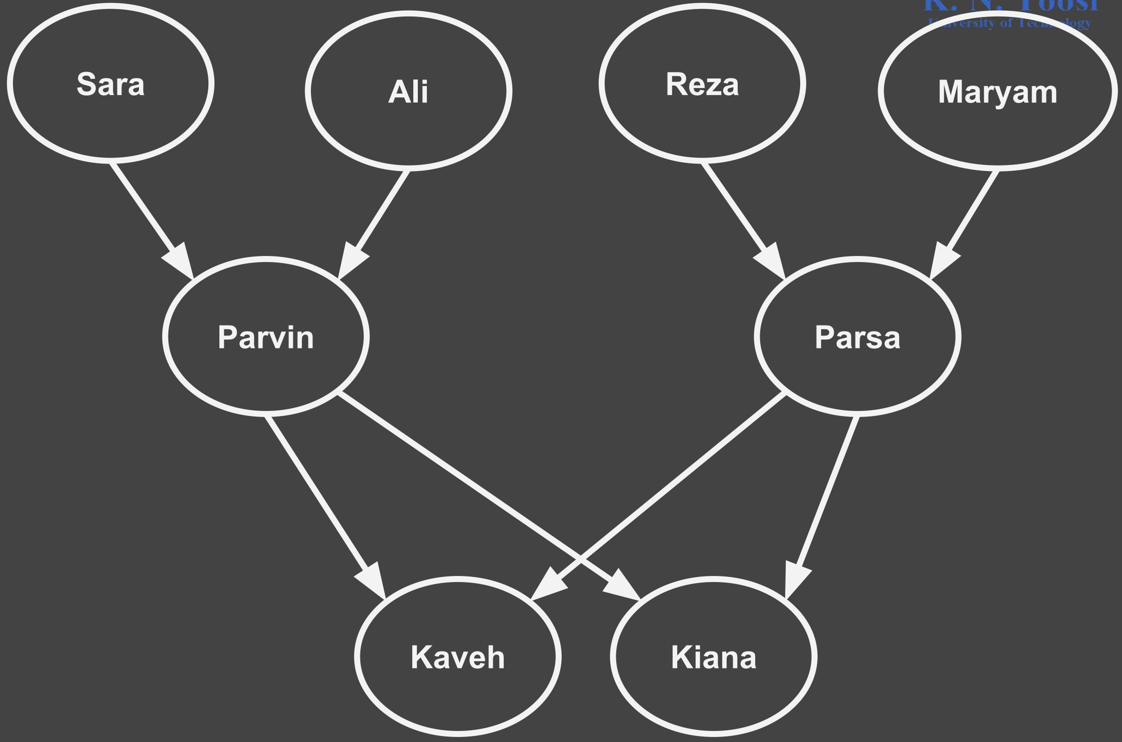
Example: Eye colour inheritance



How many parameters? 18

Example: Eye colour inheritance

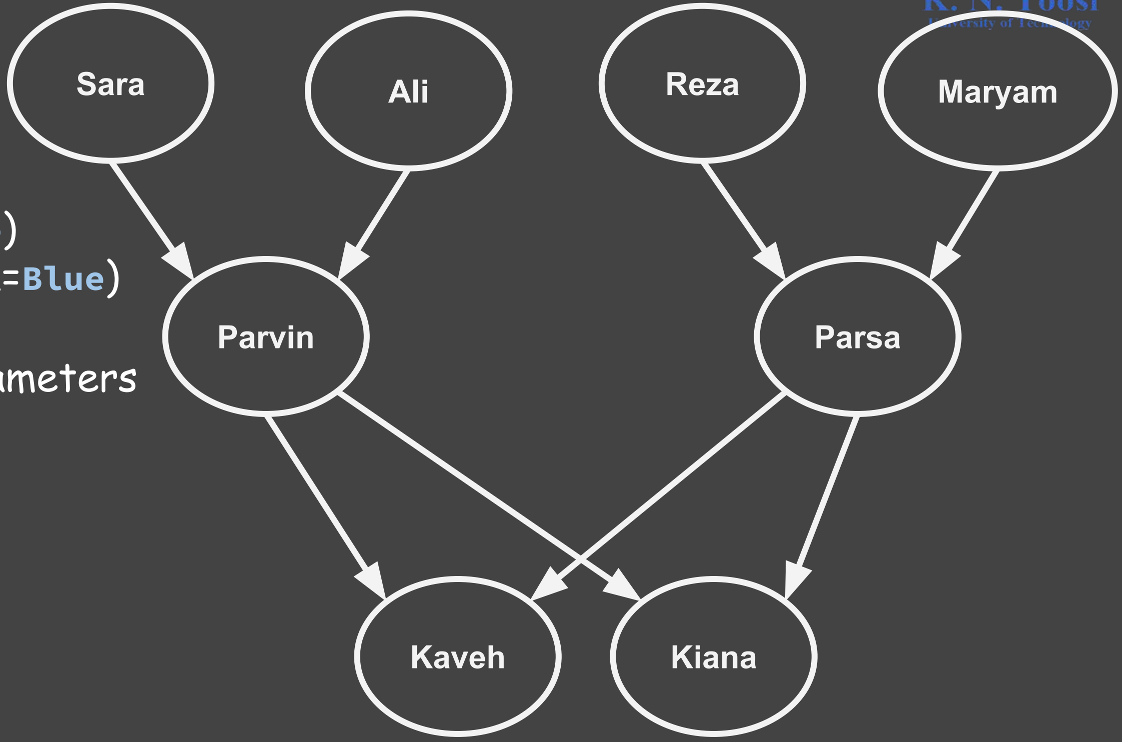




Example: Eye colour inheritance

$P(\text{Parvin}=\texttt{Blue} \mid \text{Sara}=\texttt{Green}, \text{Ali}=\texttt{Blue})$
$= P(\text{Kaveh}=\texttt{Blue} \mid \text{Parvin}=\texttt{Green}, \text{Parsa}=\texttt{Blue})$

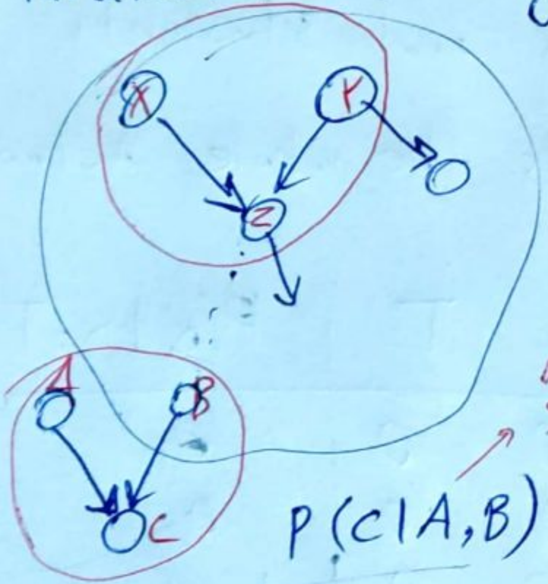
- The entire network has 18 parameters
- Parameter sharing



Parameter Sharing

K. N. Toosi
University of Technology

Parameter Sharing



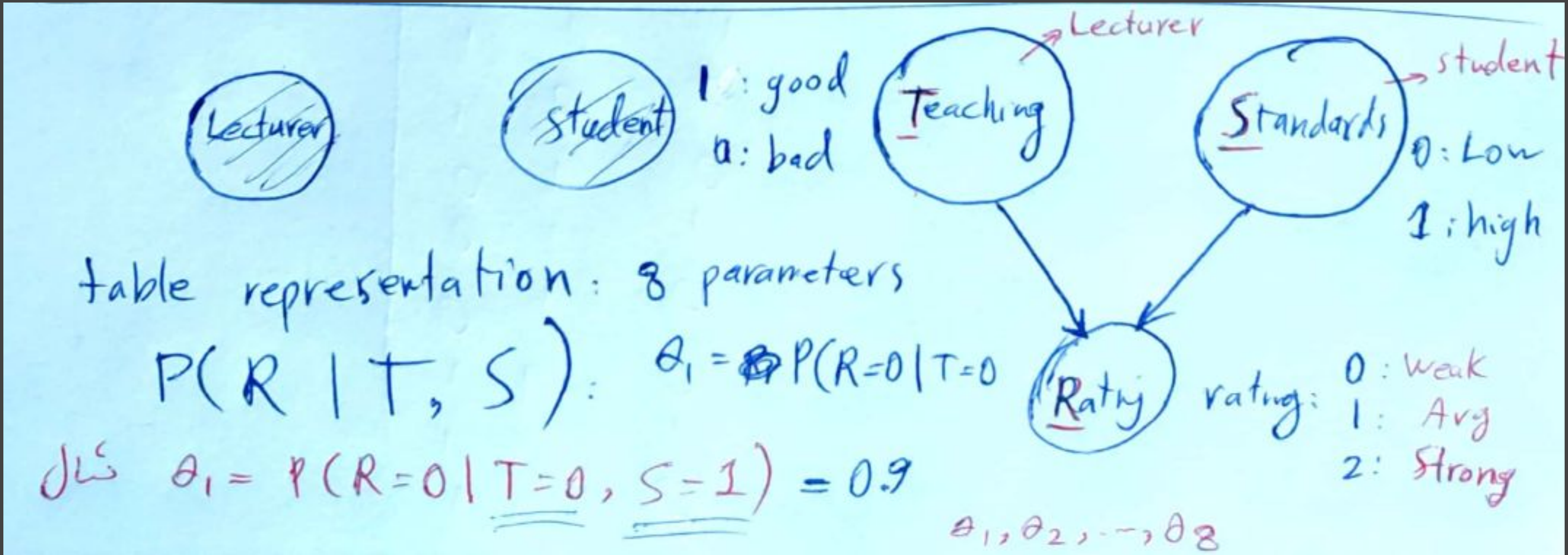
$$p(Z \mid X, Y) = f_\theta(Z, X, Y)$$

same process → $\theta \equiv \gamma$

$$p(C \mid A, B) = f_\gamma(\overset{C}{\cancel{Z}}, \overset{A}{\cancel{X}}, \overset{B}{\cancel{Y}})$$

see Koller, Sec 2, Template Models

Example

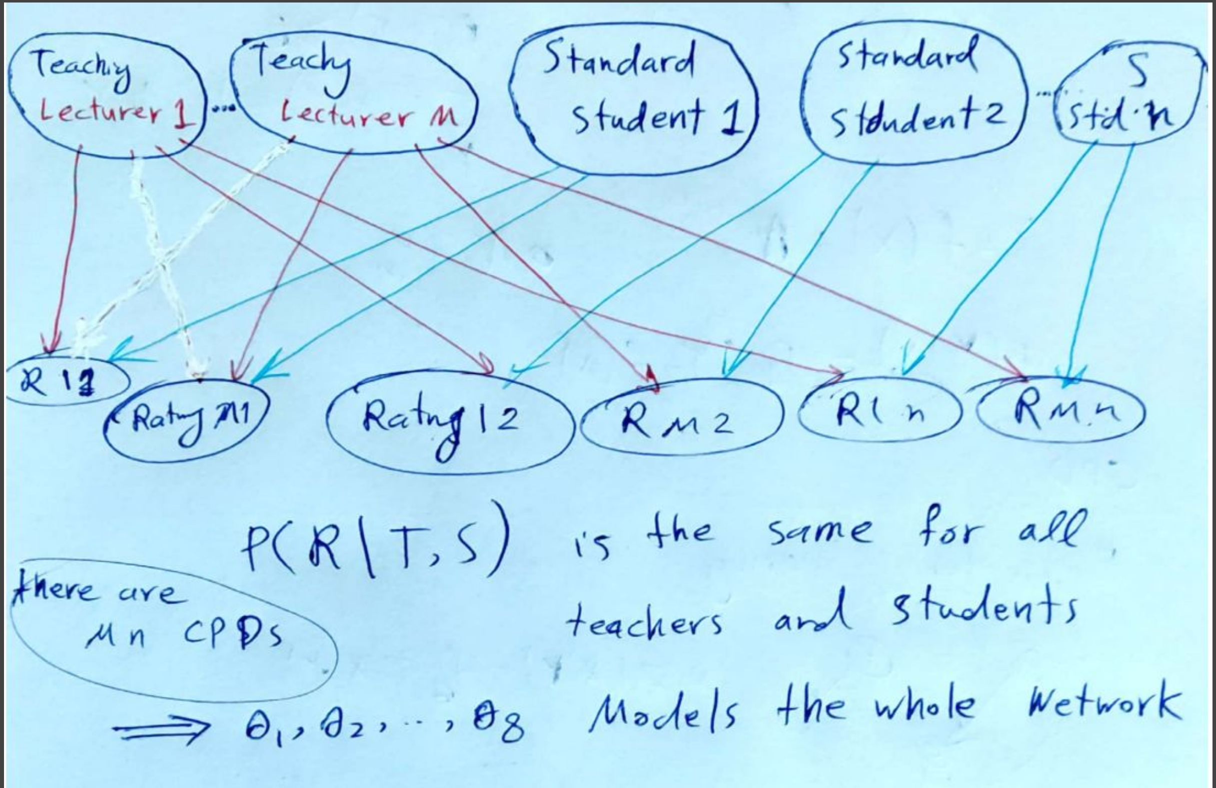

Lecturer
Student
1 : good
0 : bad
Teaching
Lecturer
Standards
student
0 : Low
1 : high
table representation : 8 parameters
P(R | T, S) : θ1 = P(R=0 | T=0
Rating
rating : 0 : Weak
1 : Avg
2 : Strong
مثال θ1 = P(R=0 | T=0 , S=1) = 0.9
θ1 , θ2 , ... , θ8

Example

$P(R|S,T)$

R	S	T	$P(R\|S,T)$
0	0	0	θ_1
1	0	0	θ_2
2	0	0	$1-\theta_1-\theta_2$
0	0	1	θ_3
1	0	1	θ_4
2	0	1	$1-\theta_3-\theta_4$
0	1	0	
1	1	0	
2	1	0	
0	1	1	
1	1	1	
2	1	1	

Example

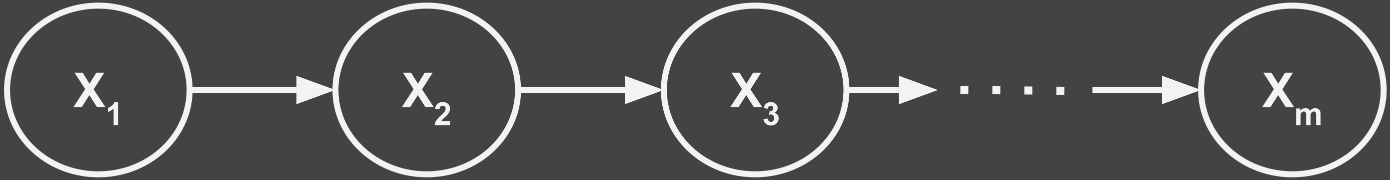

Teaching Lecturer 1
Teaching Lecturer M
Standard Student 1
Standard Student 2
S Std n
R 11
Rating M1
Rating 12
R M2
R 1n
R Mn
P(R | T, S) is the same for all teachers and students
there are Mn CPDs
⟹ θ1, θ2, …, θ8 Models the whole Network

Example: faulty push-button

$$X_1 \to X_2 \to X_3 \to \cdots \to X_m$$

- x_n device on or off (1/0) after n times pressing the button
- button works with probability p_t if device is on, and
  with probability q_t if device is off

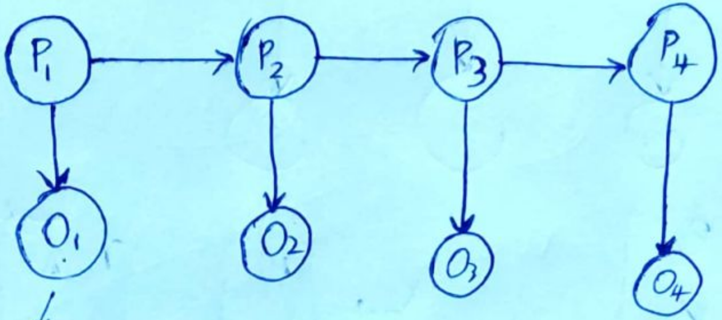
Example: faulty push-button



- x_n device on or off (1/0) after n times pressing the button
- button works with probability p_t if device is on, and
  with probability q_t if device is off
- button works with probability p if device is on, and
  with probability q if device is off

Temporal Models

P: position



GPS: Gassian Noise, standard deviation σ_g

$$P(O_t \mid P_t) = \mathcal{N}(O_t - P_t \,;\, \mu = 0 \,;\, \sigma = \sigma_g)$$

$$= \frac{1}{\bigcirc} \exp\left(-\frac{(O_t - P_t)^2}{2\sigma_g^2}\right)$$

Remember MRFs

K. N. Toosi
University of Technology

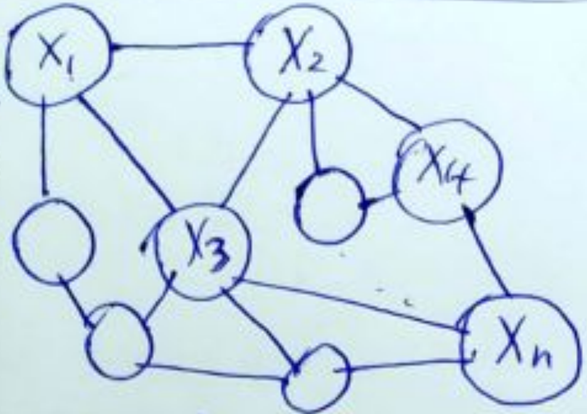
$V = \{1, 2, \ldots, n\}$ nodes/vetices

$\mathcal{E} = \{(1,2), (1,3), (2,3), (2,4) \ldots (4,n)\}$

edges

$I = \{3, 4, 7, 10\}$

$X_I = \{X_3, X_4, X_7, X_{10}\}$



$$P(X_1, X_2, \ldots, X_n) = \frac{1}{Z} \prod_{c \in C} \phi_c(X_c)$$

C: a subset of cliques

Remember MRFs

$$P(X_1, X_2, \cdots, X_n) = \frac{1}{Z} \prod_{c \in C} \phi_c(\underline{X}_c)$$

potential functions

C : a subset of cliques of the graph

E.g. $\rightarrow$ All cliques

$\rightarrow$ Maximal cliques

$\rightarrow$ nodes & edges

How to represent MRFs?

We may represent MRFs by representing the potentials $\varphi_c(X_c)$ (in addition to representing the graph itself).

Pairwise MRFs

Pairwise MRFs $\longrightarrow$ $\mathcal{C}$ = set of all edges

OR

$\mathcal{C}$ = set of all nodes + all edges

$$P(X_1, X_2, \ldots, X_n) = \frac{1}{Z} \prod_{(i,j)\in\mathcal{E}} \phi_{ij}(X_i, X_j)$$

OR

$$P(X_1, X_2, \ldots, X_n) = \frac{1}{Z} \prod_{i=1}^{n} \underbrace{\phi_i(X_i)}_{\text{unary potentials}} \prod_{i,j\in\mathcal{E}} \underbrace{\phi_{ij}(X_i, X_j)}_{\text{binary (pairwise) potentials}}$$

Pairwise MRFs - No. of Parameters

$\phi_i(X_i)$ table representation $\quad X_i = \{1, 2, \dots, M\}$

How many parameters? M parameters

$$\begin{cases} \phi_i(1) = \theta_1 \\ \phi_i(2) = \theta_2 \\ \vdots \\ \phi_i(M) = \theta_M \end{cases}$$

$$(\theta_1, \theta_2, \dots \theta_M) \equiv (\alpha\theta_1, \alpha\theta_2, \dots, \alpha\theta_M)$$

$$P(X_1, X_2) = \frac{1}{Z} \phi_1(X_1)\, \phi_2(X_2)\, \phi_{12}(X_1, X_2) \qquad X$$

$$= \frac{1}{\sum_{X_1} \sum_{X_2} (\alpha\phi_1(X_1))\, \phi_2(X_2)\, \phi_{1,2}(X_1, X_2)} (\alpha\phi_1(X_1))\, \phi_2(X_2)\, \phi_{12}(-)$$

$M-1$ free parameters for $\phi_i(X_i)$

Pairwise MRFs - No. of Parameters

$\phi_{ij}(X_i, X_j)$ $\quad X_i \in \{1, 2, \dots, M\}$

$X_j \in \{1, 2, \dots, M\}$

$M^2 - 1$ free parameters for table representation of $\underline{\phi_{ij}(X_i, X_j)}$

Pairwise MRFs - No. of Parameters

$$P(X_1, X_2, \ldots X_n) = \frac{1}{Z} \prod_{(i,j)\in\mathcal{E}} \Phi_{ij}(X_i, X_j)$$

# of parameters $|\mathcal{E}|\,(M^2-1)$.

$= O(|\mathcal{E}|.\ \ )$

fully connected pairwise MRF $= O\left(\frac{n(n-1)}{2}\right) = O(n^2)$

General Case

$P(X_1, X_2, \ldots, X_n)$ $\quad M^n-1$ parameter $\quad O(M^n)$

$X_i \in \{1, 2, \ldots, M\}$

$O(n^2) \ll O(M^n)$

Exponential form

$X_i \in \{1, 2, \dots, M\}$ $\quad$ M might be very large

X_i might be continuous

$\phi_{ij}(X_i, X_j)$ or $\phi_c(X_c)$ must be parameterized

such that $\phi_c(X_c) > 0$ $\quad$ (or $\phi_c(X_c) \geq 0$)

Expenential form $\quad \phi_c(X_c) > 0$

$$\Rightarrow \phi_c(X_c) = e^{\theta_c(X_c)} \qquad \theta_c(X_c) = \ln(\phi_c(X_c))$$

Exponential form

$$P(X_1, X_2, \ldots, X_n) = \frac{1}{Z} \prod_{c \in C} \phi_c(X_c)$$

$$= \frac{1}{Z} \prod_{c \in C} e^{\theta_c(X_c)}$$

$$= \frac{1}{Z} e^{\sum_{c \in C} \theta_c(X_c)}$$

$$\phi_c(X_c) > 0 \qquad \theta_c(X_c) \in \mathbb{R}$$

log-linear representation

log-linear form: A way to parameterize $\phi_c(X_c)$

$$\theta_c(X_c) = w_1 \, f_c^1(X_c) + w_2 \, f_c^2(X_c) + \cdots + w_p \, f_c^p(X_c)$$

parameters

feature

$$\phi_c(X_c) = e^{\theta_c(X_c)} = \exp\left(\sum_{i=1}^{p} w_i \, f_c^i(X_c) \right)$$

log $\phi_c(X_c)$ is linear in $(w_1, w_2, \dots, w_n)$

parameters

$\Longrightarrow$ log-linear parameterization

Example: log-linear pairwise MRFs

$$p(X_1, X_2, \ldots, X_n) = \frac{1}{Z} \prod_{i=1}^{n} \phi_i(X_i) \prod_{i,j \in \mathcal{E}} \phi_{ij}(X_i, X_j)$$

$$= \frac{1}{Z} \exp\left(\sum_{i=1}^{n} \theta_i(X_i) + \sum_{i,j \in \mathcal{E}} \theta_{ij}(X_i, X_j) \right)$$

log-linear

Parameters

Parameters

$$= \frac{1}{Z} \exp\left(\sum_{i=1}^{n} \sum_{k=1}^{P} w_{i,k} \, f_i^k(X_i) + \sum_{i,j \in \mathcal{E}} \sum_{k=1}^{Q} w_{i,j,k} \, f_{ij}^k(X_i, X_j) \right)$$

Energy function

12 I

$$\underbrace{P(X_1, X_2, \ldots, X_n)}_{X} = \frac{1}{Z} \prod_{i=1}^{n} \phi_i(X_i) \prod_{(i,j)\in \mathcal{E}} \phi_{ij}(X_i, X_j)$$

$$P(X_1, \ldots, X_n) = P(X) = \frac{1}{Z} \exp\Big(\underbrace{\sum_{i=1}^{n} \theta_i(X_i) + \sum_{(i,j)\in \mathcal{E}} \theta_{ij}(X_i, X_j)}_{\theta(X_1, \ldots, X_n) \,=\, \theta(X)} \Big)$$

$$\boxed{\begin{aligned} E(X) &= -\theta(X) \\ E_i(X_i) &= -\theta_i(X_i) \\ E_{ij}(X_i, X_j) &= -\theta_{ij}(X_i, X_j) \end{aligned}}$$

$$= \frac{1}{Z} \exp\big(\theta(X)\big)$$

$$= \frac{1}{Z} \exp\big(-E(X)\big)$$

unary energy

$$= \frac{1}{Z} \exp\Big(-\sum_{i=1}^{n} E_i(X_i) - \sum_{(i,j)\in \mathcal{E}} E_{ij}(X_i, X_j)\Big)$$

binary energy function

$$E(X) = \sum_{i=1}^{n} E_i(X_i) + \sum_{(i,j)\in \mathcal{E}} E_{ij}(X_i, X_j)$$

log-linear models

Back to log-linear models

$$P(X_1 \text{—} X_n) = \frac{1}{Z} \exp\left(\sum_{i=1}^{n} \theta_i(X_i) + \sum_{(i,j)\in E} \theta_{ij}(X_i, X_j) \right)$$

$$= \frac{1}{Z} \exp\left(\sum_{i=1}^{n} \sum_{k=1}^{P} w_i^k f_i^k(X_i) + \sum_{(i,j)\in E} \sum_{k=1}^{Q} w_{ij}^k f_{ij}^k(X_i, X_j) \right)$$

How strong are log-linear models?

K. N. Toosi
University of Technology

$$P(X_1 \ldots X_n) = \frac{1}{Z}\exp\Big(\sum_{i=1}^{n}\theta_i(X_i) + \sum_{(i,j)\in E}\theta_{ij}(X_i,X_j)\Big)$$

$$= \frac{1}{Z}\exp\Big(\underline{\sum_{i=1}^{n}\sum_{k=1}^{p} w_i^k f_i^k(X_i)} + \sum_{(i,j)\in E}\sum_{k=1}^{q} w_{ij}^k f_{ij}^k(X_i,X_j)\Big)$$

$X_i \in \{0,1\}$

X_i	$\theta_i(X_i)$
0	$\theta_i(0) = \theta_{i0}$
1	$\theta_i(1) = \theta_{i1}$

$p = 2$

$$w_i^1 f_i^1(X_i) + w_i^2 f_i^2(X_i)$$

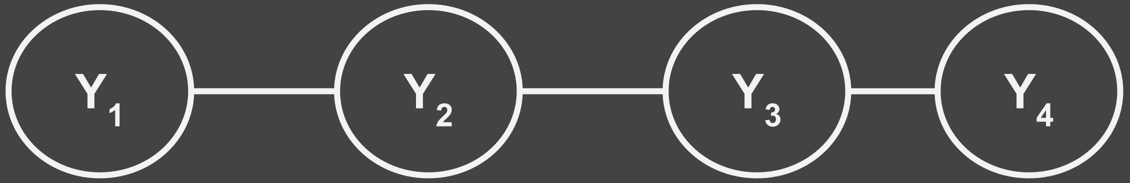
$$\theta_{i0}\, 1(X_i = 0) + \theta_{i1}\, 1(X_i = 1)$$

parameters features

X_i, X_j	$\theta_{ij}(X_i, X_j)$
0 0	$\cancel{\theta_{ij00}}$ α
0 1	$\cancel{\theta_{ij01}}$ β
1 0	θ γ
1 1	σ

$$\alpha\, 1(X_i = 0, X_j = 0) + \beta\, 1(X_i = 0, X_j = 1)$$
$$+ \gamma\, 1(X_i = 1, X_j = 0) + \sigma\, 1(X_i = 1, X_j = 1)$$

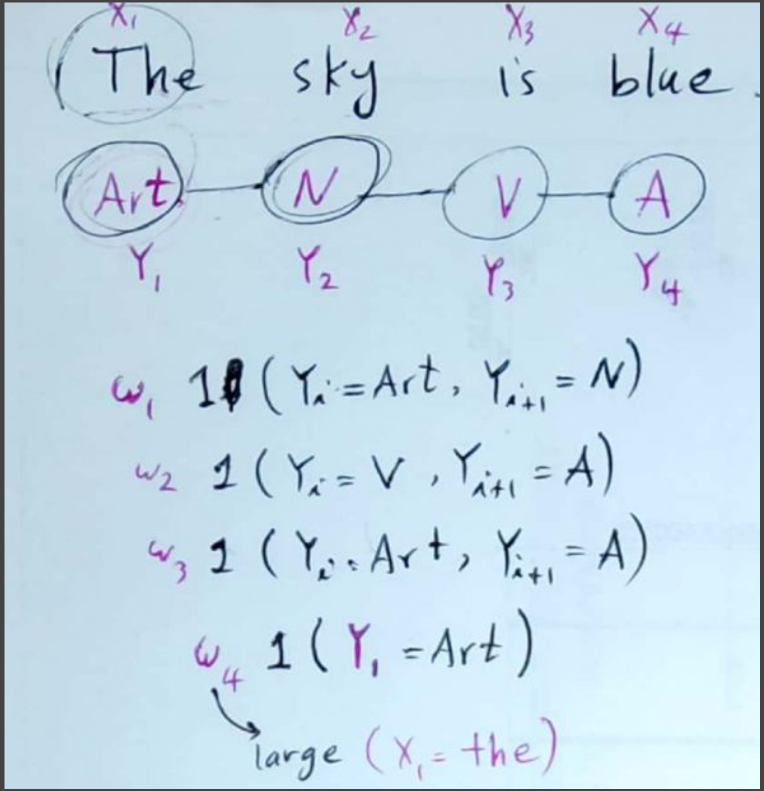
Example



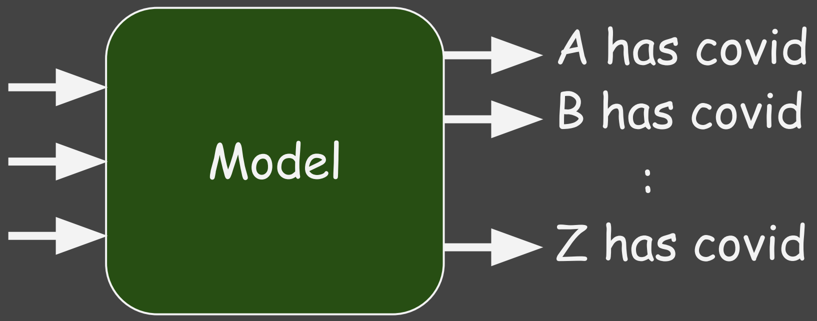
X_1	X_2	X_3	X_4
The	sky	is	blue

$$w_1 \, 1(Y_i = \text{Article},\ \ Y_{i+1} = \text{Noun})$$
$$w_2 \, 1(Y_i = \text{Verb},\ \ Y_{i+1} = \text{Adjective})$$
$$w_3 \, 1(Y_i = \text{Article},\ \ Y_{i+1} = \text{Adjective})$$
$$w_4 \, 1(Y_1 = \text{Article})$$

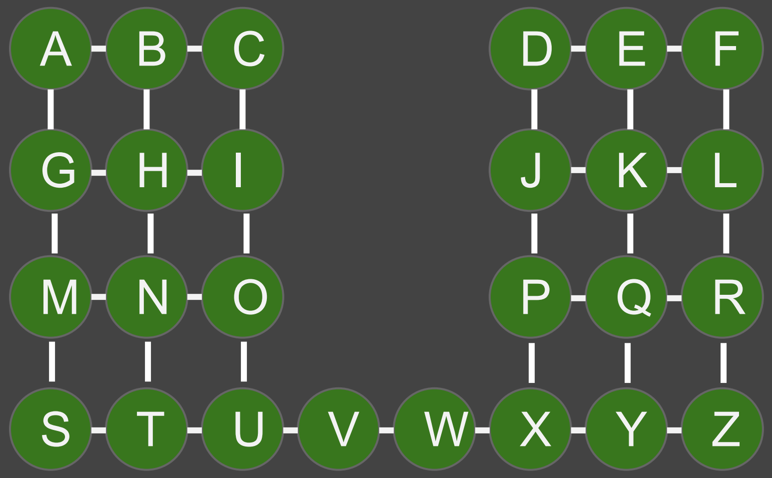
Example


X1 X2 X3 X4
The sky is blue.
Art N V A
Y1 Y2 Y3 Y4
w1 1(Yi = Art, Yi+1 = N)
w2 1(Yi = V, Yi+1 = A)
w3 1(Yi = Art, Yi+1 = A)
w4 1(Y1 = Art)
large (X1 = the)

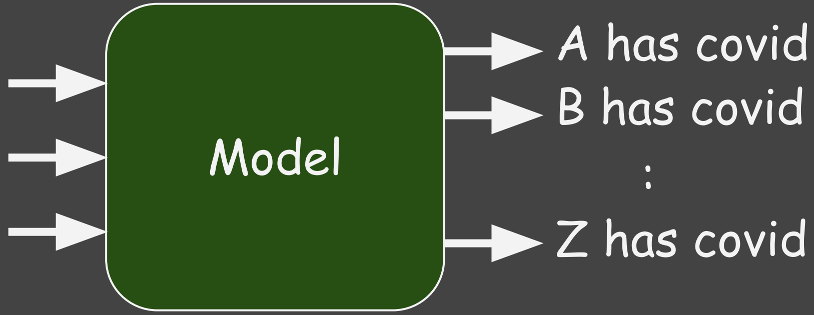
Using structure

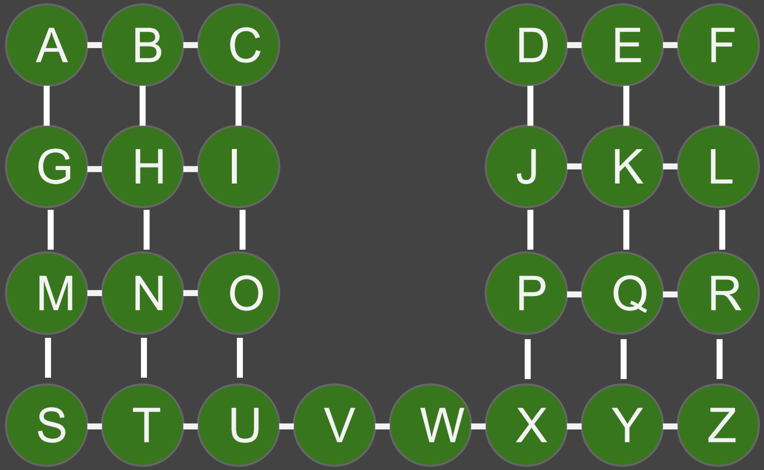


Features?



Using structure





Features?

f(A,B) = 1(A = B)

Example: Ising Model

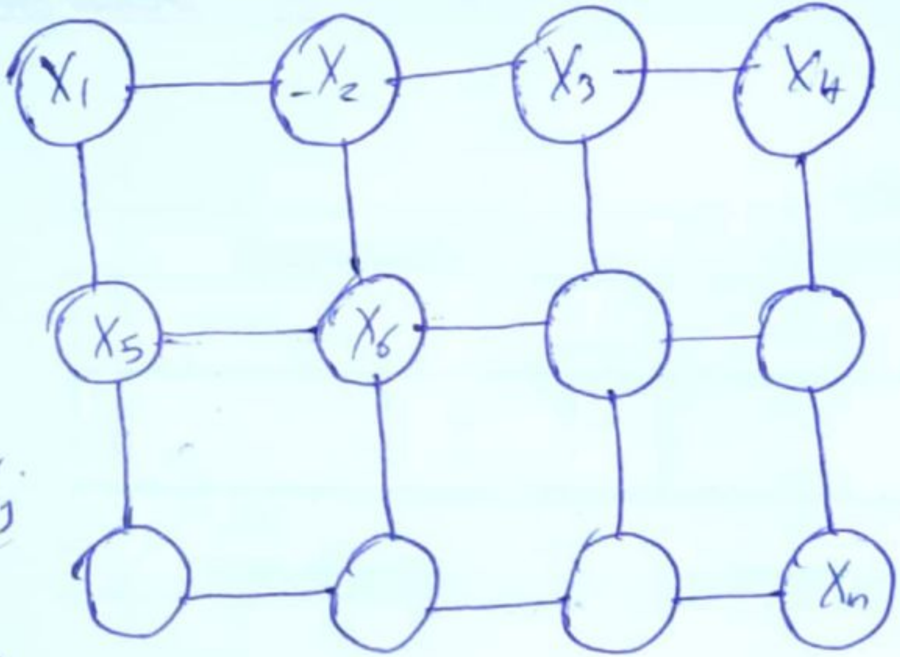
Ising Model

$X_i \in \{-1, 1\}$

$$\boxed{e^{-\frac{1}{T}E(x)}}$$

$$P(X_1 \ldots X_n) = P(X) = e^{-E(X)}$$

$$E(X_1 \ldots X_n) = -\sum_{i=1}^{n} w_i X_i - \sum_{i,j \in E} w_{ij} X_i X_j$$



Example: Ising Model

$$P(X_1 — X_n) = P(X) = e^{-E(X)}$$

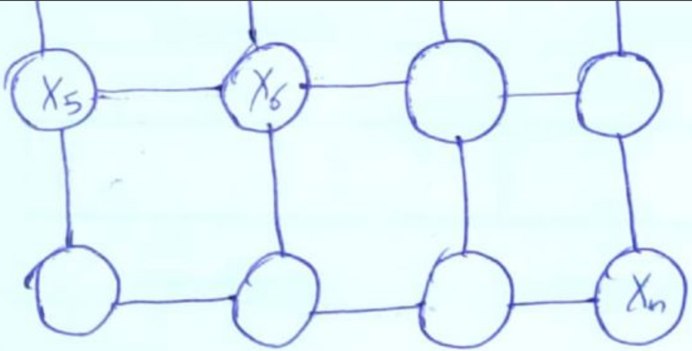
$$E(X_1 — X_n) = -\sum_{i=1}^{n} w_i X_i - \sum_{i,j \in E} w_{ij} X_i X_j$$

$w_{ij} > 0$ X_i, X_j tend to have similar spin

$w_{ij} < 0$ X_i, X_j tend to have ~~differen~~ opposite spins

$w_i > 0$ X_i tend to ~~0~~ = 1

$w_i < 0$ X_i tend to = -1



~~xxxx~~ M×N gride

node $MN = n$

edge $M(N-1) + N(M-1) = 2MN - M - N$

Shared edge parameters

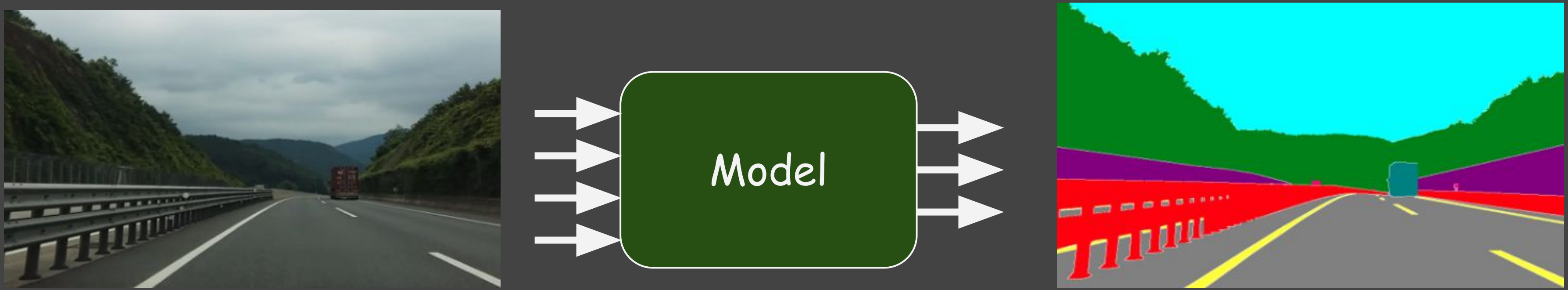
Homogeneous & isotropic systems

$$E(X_1, \dots, X_n) = -\sum_{i=1}^{n} w_i X_i - \sum_{(i,j)\in \mathcal{E}} w\, X_i X_j$$

parameter sharing

Image semantic segmentation

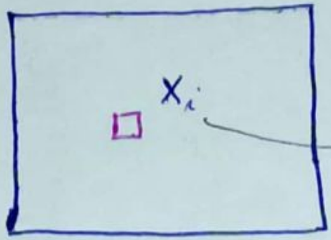
K. N. Toosi
University of Technology



https://sites.google.com/site/highwaydrivingdataset

Example: Image Segmentation

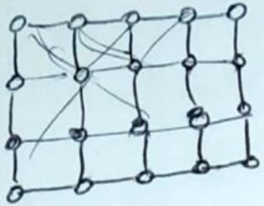
Image



Segmentation map

$X_i \in \{$ road, sky,
mountain,
Vegetation,
sidewalk, ... $\}$
Vehicle
Pedestrian, ~

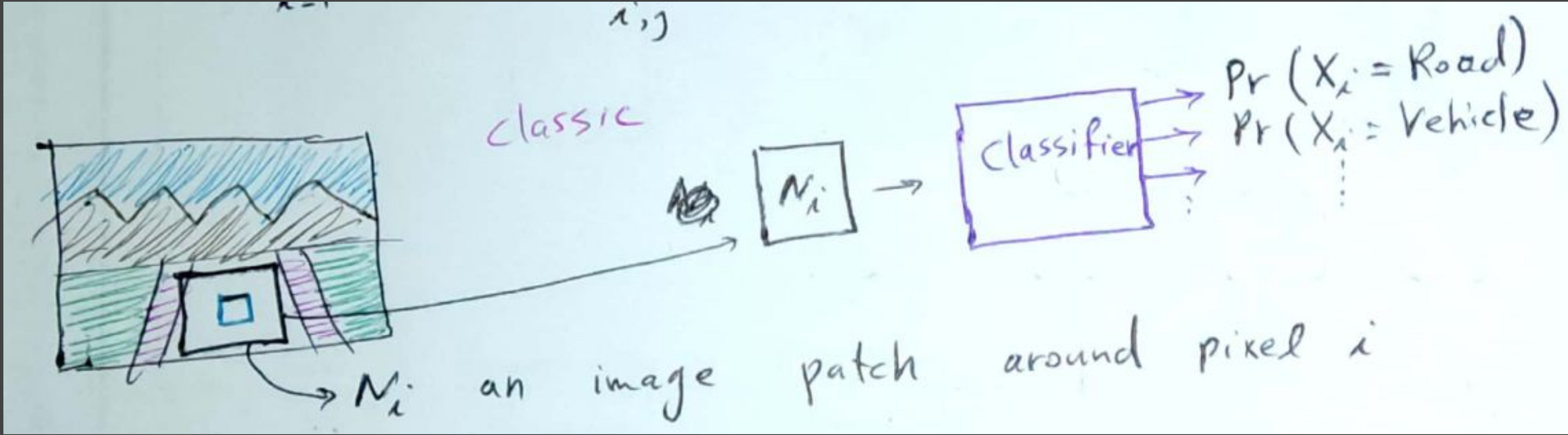
$$P(X_1 — X_n) = \frac{1}{Z} e^{-E(X_1 — X_n)}$$



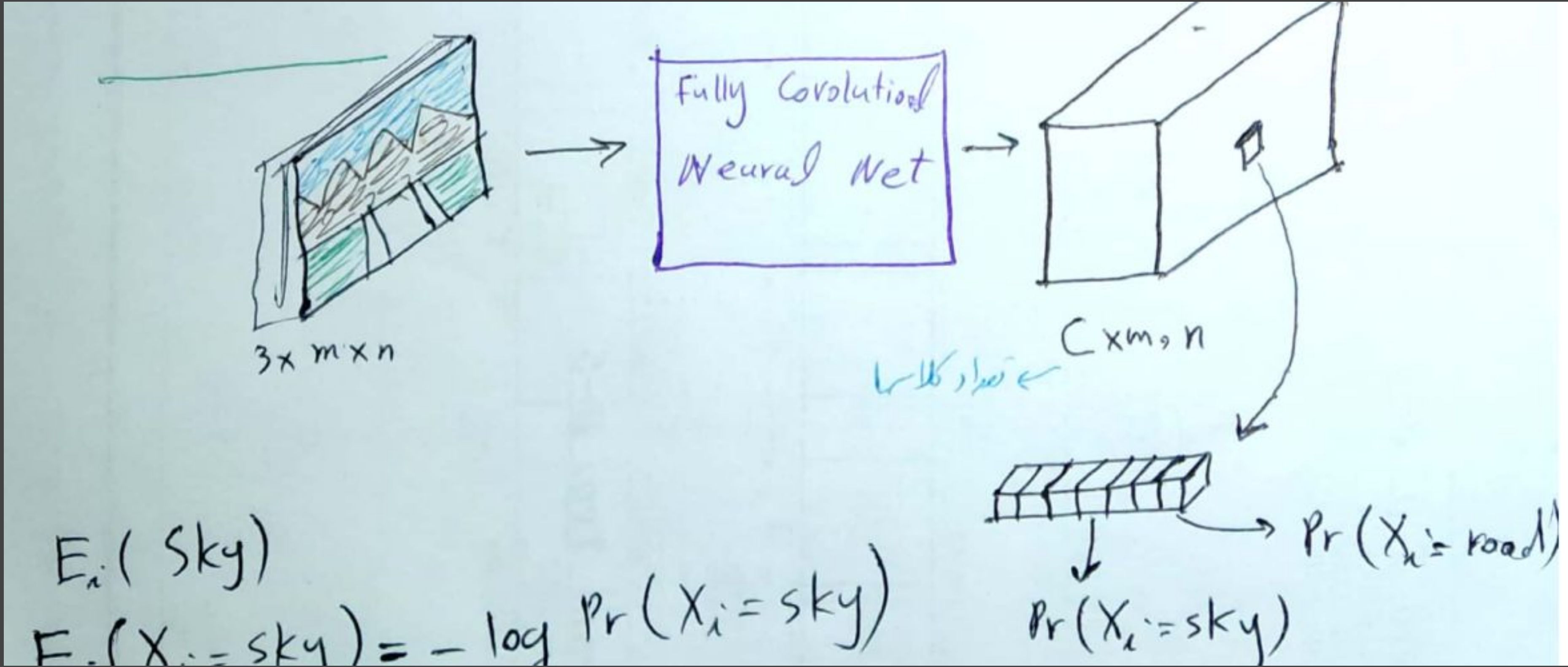
$$E(X_1, \dots, X_n) =$$

$$E(X_1 — X_n) = \sum_{i=1}^{n} E_i(X_i) + \sum_{i,j} E_{ij}(X_i, X_j)$$

Segmentation: How to represent unary energy terms $E_i(X_i)$?

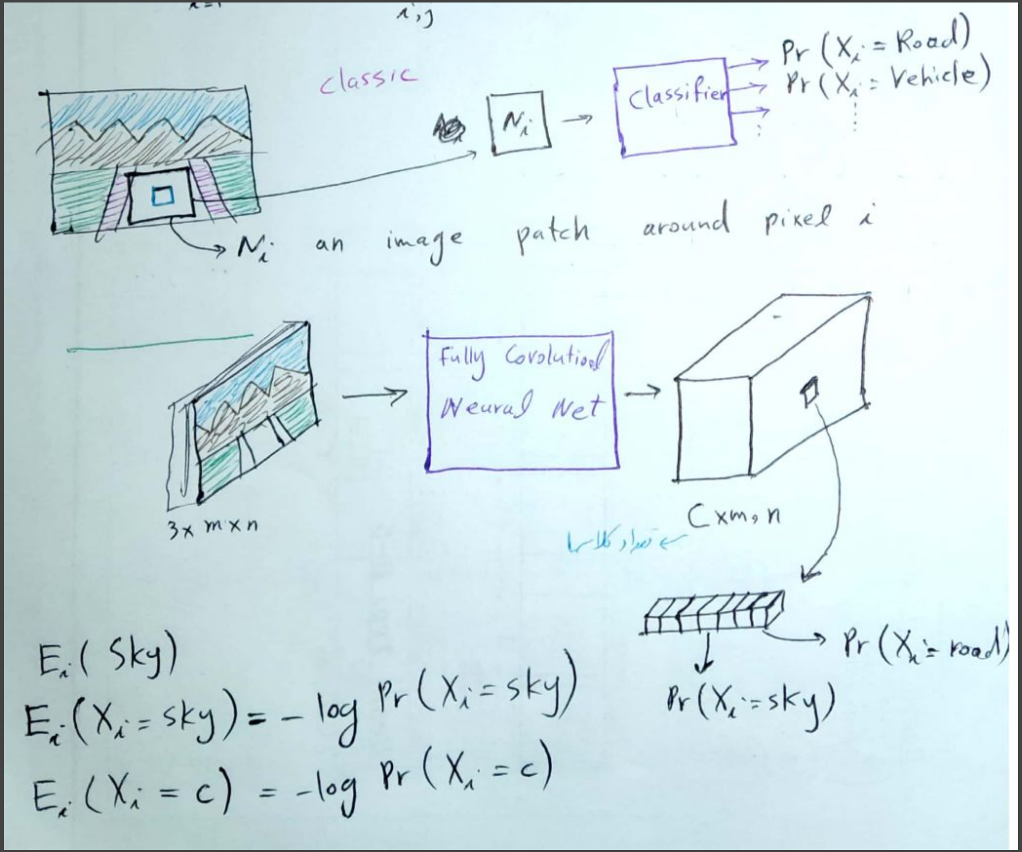


Segmentation: How to represent unary energy terms $E_i(X_i)$?



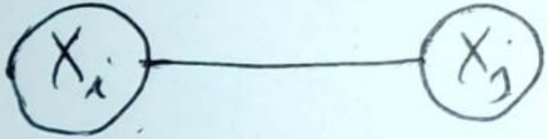
Segmentation: How to represent unary energy terms $E_i(X_i)$?

K. N. Toosi
University of Technology



How to represent binary energy terms $E_{ij}(X_i, X_j)$?

$$E(X) = \overset{\checkmark}{\sum_{i=1}^{n} E(X_i)} + \overset{?}{\sum_{(i,j)\in \Sigma} E_{ij}(X_i, X_j)}$$



Neighbouring pixels tend to ~~have~~ ~~effect~~ belong to the same class

How to represent binary energy terms $E_{ij}(X_i, X_j)$?

K. N. Toosi
University of Technology

Neighbours pixels tend to ~~have~~ belong to the same class

$$E_{ij}(X_i, X_j) = \underbrace{\mathbb{1}(X_i \neq X_j)}_{\text{Potts model}}$$

More Recent Models

$$E_{ij}(X_i, X_j) = \sum_{k=1}^{C} \sum_{l=1}^{C} w_{kl} \, \mathbb{1}(X_i = k, X_j = l)$$

w_{kl}: learnable parameters

k: vehicle

l: road