# Restricted Boltzmann Machine

Ali Adibifar | Hamid Fareghbal Khamene

Department of Mathematical Sciences
Sharif University of Technology

# Table of Contents

# Introduction

# Prologue

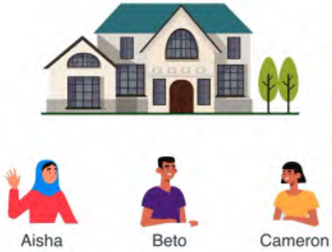
Boltzmann distribution can be used to determine the distribution of the kinetic energy of for a set of molecules

# Prologue

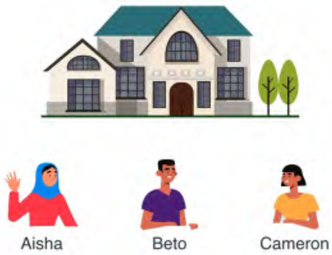
- Got popular after Netflix competition
- Is a Graphical model
- Used for unsupervised learning
- Has 1 visible and 1 hidden layer
- Involves learning a probability distribution from an original dataset and using it to make inferences about never before seen data

# Basics

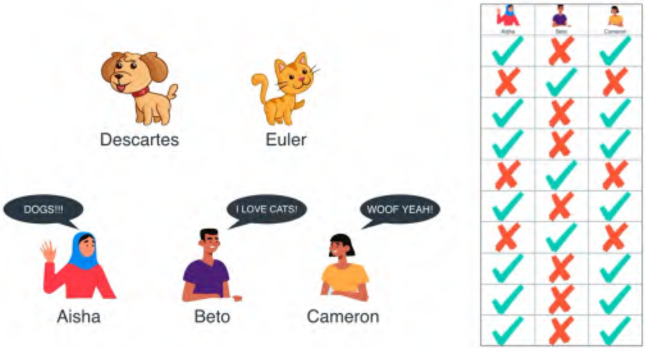
# The mystery



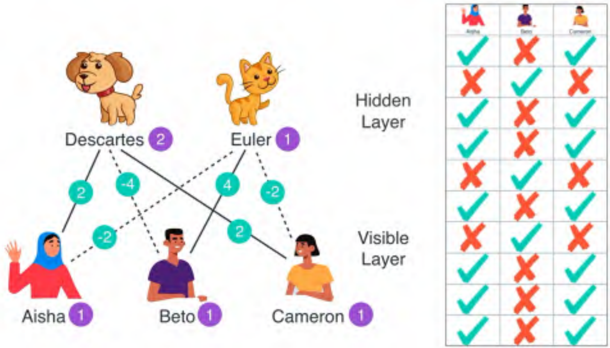
# The mystery



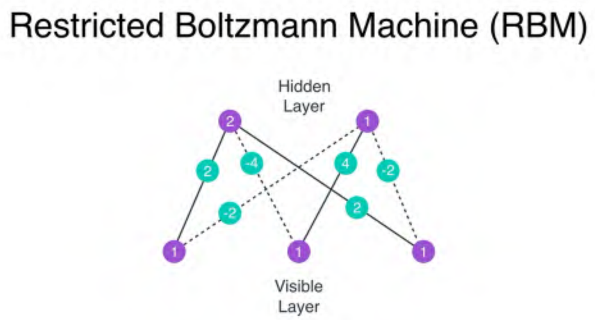

# Solution


Descartes
Euler
DOGS!!!
I LOVE CATS!
WOOF YEAH!
Aisha
Beto
Cameron

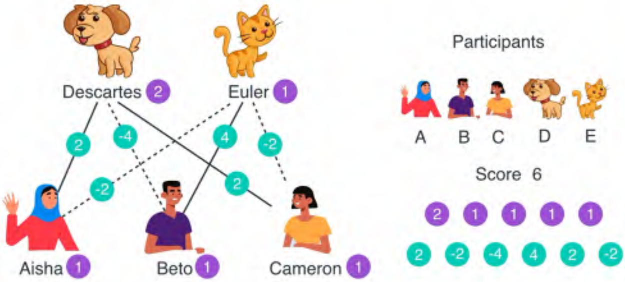
# Weights


Descartes 2
Euler 1
Hidden Layer
2
-4
4
-2
-2
2
Visible Layer
Aisha 1
Beto 1
Cameron 1

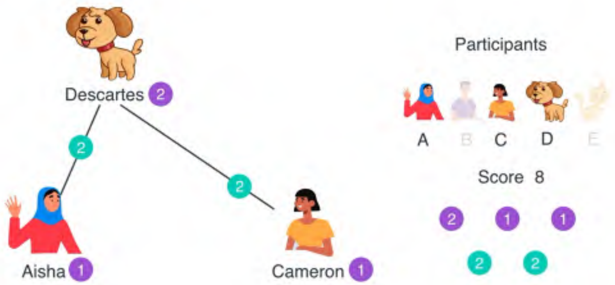
# Model


Restricted Boltzmann Machine (RBM)
Hidden
Layer
2
1
2
-4
4
-2
-2
2
1
1
1
Visible
Layer

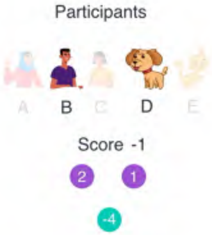
# Scores


Descartes 2
Euler 1
2
-4
4
-2
-2
2
Aisha 1
Beto 1
Cameron 1
Participants
A
B
C
D
E
Score 6
2 1 1 1 1
2 -2 -4 4 2 -2

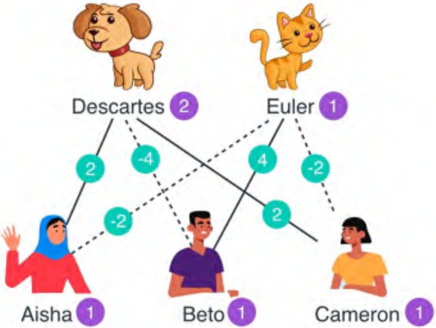
# Scores


Descartes 2
2
2
Aisha 1
Cameron 1
Participants
A
B
C
D
E
Score 8
2
1
1
2
2

# Scores

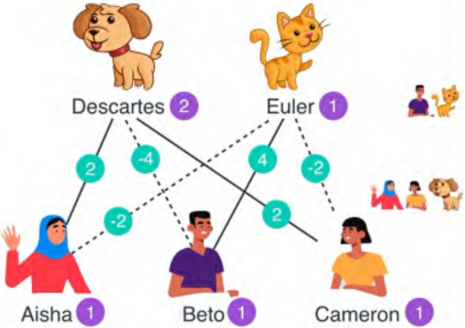



# Scores



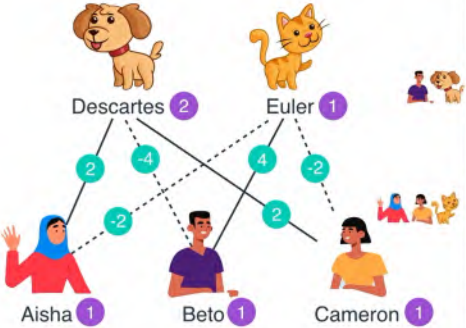

| Scenario | Score |
|---|---|
| None | 0 |
| A | 1 |
| B | 1 |
| C | 1 |
| D | 2 |
| E | 1 |
| AB | 2 |
| AC | 2 |
| AD | 5 |
| AE | 0 |
| BC | 2 |
| BD | -2 |
| BE | 7 |
| CD | 5 |
| CE | 0 |
| DE | 3 |
| ABC | 3 |
| ABD | 1 |
| ABE | 6 |
| ACD | 8 |
| ACE | -1 |
| ADE | 4 |
| BCD | 1 |
| BCE | 6 |
| BDE | 4 |
| CDE | 4 |
| ABCD | 4 |
| ABCE | 5 |
| ABDE | 5 |
| ACDE | 5 |
| BCDE | 5 |
| ABCDE | 6 |

# Scores



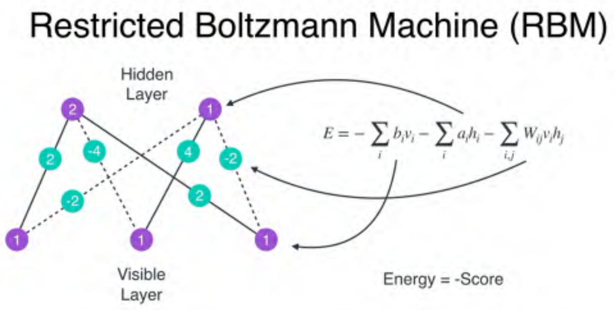

| Scenario | Score |
|---|---|
| None | 0 |
| A | 1 |
| B | 1 |
| C | 1 |
| D | 2 |
| E | 1 |
| AB | 2 |
| AC | 2 |
| AD | 5 |
| AE | 0 |
| BC | 2 |
| BD | -2 |
| BE | 7 |
| CD | 5 |
| CE | 0 |
| DE | 3 |
| ABC | 3 |
| ABD | 1 |
| ABE | 6 |
| ACD | 8 |
| ACE | -1 |
| ADE | 4 |
| BCD | 1 |
| BCE | 6 |
| BDE | 4 |
| CDE | 4 |
| ABCD | 4 |
| ABCE | 5 |
| ABDE | 5 |
| ACDE | 5 |
| BCDE | 5 |
| ABCDE | 6 |

# Scores

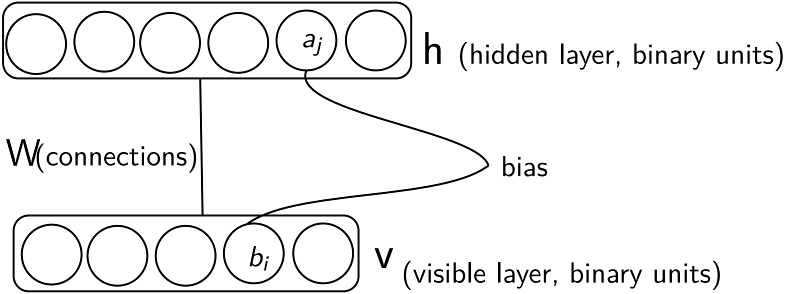



| Scenario | Score |
|---|---|
| None | 0 |
| A | 1 |
| B | 1 |
| C | 1 |
| D | 2 |
| E | 1 |
| AB | 2 |
| AC | 2 |
| AD | 5 |
| AE | 0 |
| BC | 2 |
| BD | -2 |
| BE | 7 |
| CD | 5 |
| CE | 0 |
| DE | 3 |
| ABC | 3 |
| ABD | 1 |
| ABE | 6 |
| ACD | 8 |
| ACE | -1 |
| ADE | 4 |
| BCD | 1 |
| BCE | 6 |
| BDE | 4 |
| CDE | 4 |
| ABCD | 4 |
| ABCE | 5 |
| ABDE | 5 |
| ACDE | 5 |
| BCDE | 5 |
| ABCDE | 6 |

# Model


Restricted Boltzmann Machine (RBM)
Hidden
Layer
Visible
Layer
$E = -\sum_i b_i v_i - \sum_i a_i h_i - \sum_{i,j} W_{ij} v_i h_j$
Energy = -Score

# Model

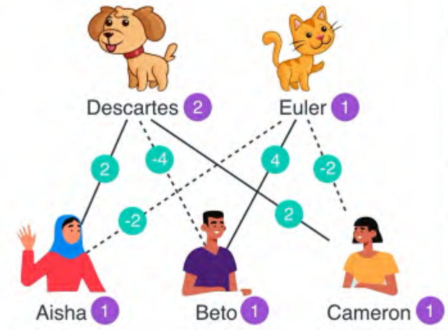



- Energy function:

$$E(v,h) = -h^T W v - b^T v - a^T h$$
$$= -\sum_j \sum_i W_{j,i} h_j v_i - \sum_i b_i v_i - \sum_j a_j h_j$$

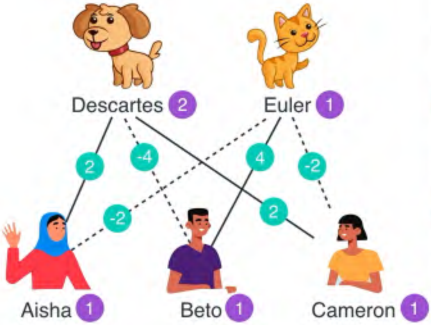
# Probability and Inference

# Probability




| Scenario | Score | $e^{Score}$ | Probability |
|---|---|---|---|
| None | 0 | 1 | 0 |
| A | 1 | 2.72 | 0 |
| B | 1 | 2.72 | 0 |
| C | 1 | 2.72 | 0 |
| D | 2 | 7.38 | 0 |
| E | 1 | 2.72 | 0 |
| AB | 2 | 7.38 | 0 |
| AC | 2 | 7.38 | 0 |
| AD | 5 | 148.41 | 0.02 |
| AE | 0 | 2.72 | 0 |
| BC | 2 | 7.38 | 0 |
| BD | -2 | 0.14 | 0 |
| BE | 7 | 1096.63 | 0.17 |
| CD | 5 | 148.41 | 0.02 |
| CE | 0 | 1 | 0 |
| DE | 3 | 20.08 | 0 |
| ABC | 3 | 20.08 | 0 |
| ABD | 1 | 2.72 | 0 |
| ABE | 6 | 403.43 | 0.06 |
| ACD | 8 | 2980.96 | 0.45 |
| ACE | -1 | 0.37 | 0 |
| ADE | 4 | 54.6 | 0 |
| BCD | 1 | 2.72 | 0 |
| BCE | 6 | 403.43 | 0.06 |
| BDE | 4 | 54.6 | 0 |
| CDE | 4 | 54.6 | 0 |
| ABCD | 4 | 54.6 | 0.02 |
| ABCE | 5 | 148.41 | 0.02 |
| ABDE | 5 | 148.41 | 0.02 |
| ACDE | 5 | 148.41 | 0.02 |
| BCDE | 5 | 148.41 | 0.02 |
| ABCDE | 6 | 403.43 | 0.06 |

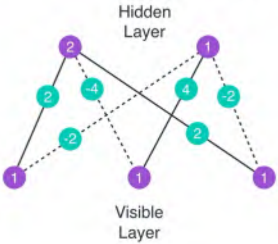
# Probability




| Scenario | Score | $e^{Score}$ | Probability |
|---|---|---|---|
| None | 0 | 1 | 0 |
| A | 1 | 2.72 | 0 |
| B | 1 | 2.72 | 0 |
| C | 1 | 2.72 | 0 |
| D | 2 | 7.38 | 0 |
| E | 1 | 2.72 | 0 |
| AB | 2 | 7.38 | 0 |
| AC | 2 | 7.38 | 0 |
| AD | 5 | 148.41 | 0.02 |
| AE | 0 | 2.72 | 0 |
| BC | 2 | 7.38 | 0 |
| BD | -2 | 0.14 | 0 |
| BE | 7 | 1096.63 | 0.17 |
| CD | 5 | 148.41 | 0.02 |
| CE | 0 | 1 | 0 |
| DE | 3 | 20.08 | 0 |
| ABC | 3 | 20.08 | 0 |
| ABD | 1 | 2.72 | 0 |
| ABE | 6 | 403.43 | 0.06 |
| ACD | 8 | 2980.96 | 0.45 |
| ACE | -1 | 0.37 | 0 |
| ADE | 4 | 54.6 | 0 |
| BCD | 1 | 2.72 | 0 |
| BCE | 6 | 403.43 | 0.06 |
| BDE | 4 | 54.6 | 0 |
| CDE | 4 | 54.6 | 0 |
| ABCD | 4 | 54.6 | 0.02 |
| ABCE | 5 | 148.41 | 0.02 |
| ABDE | 5 | 148.41 | 0.02 |
| ACDE | 5 | 148.41 | 0.02 |
| BCDE | 5 | 148.41 | 0.02 |
| ABCDE | 6 | 403.43 | 0.06 |

# Model

## Energy to probability




$$E = -\sum_i b_i v_i - \sum_i a_i h_i - \sum_{i,j} W_{ij} v_i h_j$$

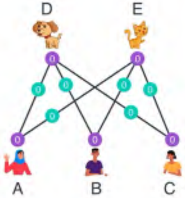
$$p(v,h) = \frac{1}{Z} e^{-E(v,h)} \qquad Z = \sum_{v,h} e^{-E(v,h)}$$

# Model

$$Z = \sum_{v,h} \exp(-E(v,h))$$

$$\text{Distribution: } p(v,h) = \exp(-E(v,h))/Z$$
$$= \exp(h^T Wv + b^T v + a^T h)/Z$$
$$= \exp(h^T Wv) \exp(b^T v) \exp(a^T h)/Z$$

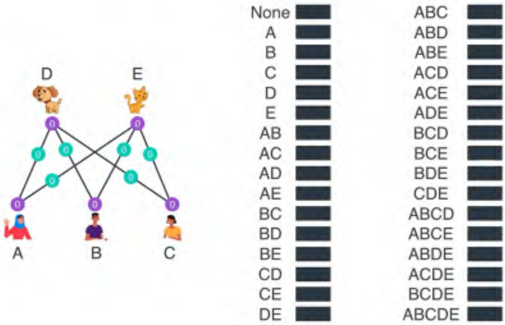
# Probability



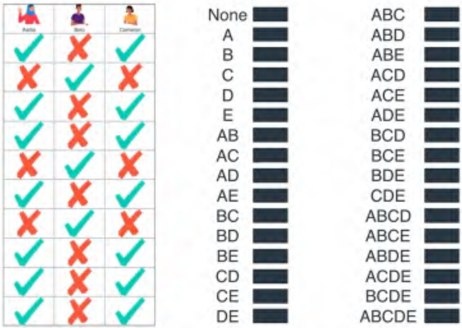

| Scenario | Score | $e^{Score}$ | Probability |
|---|---|---|---|
| None | 0 | 1 | 1/32 |
| A | 0 | 1 | 1/32 |
| B | 0 | 1 | 1/32 |
| C | 0 | 1 | 1/32 |
| D | 0 | 1 | 1/32 |
| E | 0 | 1 | 1/32 |
| AB | 0 | 1 | 1/32 |
| AC | 0 | 1 | 1/32 |
| AD | 0 | 1 | 1/32 |
| AE | 0 | 1 | 1/32 |
| BC | 0 | 1 | 1/32 |
| BD | 0 | 1 | 1/32 |
| BE | 0 | 1 | 1/32 |
| CD | 0 | 1 | 1/32 |
| CE | 0 | 1 | 1/32 |
| DE | 0 | 1 | 1/32 |

| Scenario | Score | $e^{Score}$ | Probability |
|---|---|---|---|
| ABC | 0 | 1 | 1/32 |
| ABD | 0 | 1 | 1/32 |
| ABE | 0 | 1 | 1/32 |
| ACD | 0 | 1 | 1/32 |
| ACE | 0 | 1 | 1/32 |
| ADE | 0 | 1 | 1/32 |
| BCD | 0 | 1 | 1/32 |
| BCE | 0 | 1 | 1/32 |
| BDE | 0 | 1 | 1/32 |
| CDE | 0 | 1 | 1/32 |
| ABCD | 0 | 1 | 1/32 |
| ABCE | 0 | 1 | 1/32 |
| ABDE | 0 | 1 | 1/32 |
| ACDE | 0 | 1 | 1/32 |
| BCDE | 0 | 1 | 1/32 |
| ABCDE | 0 | 1 | 1/32 |

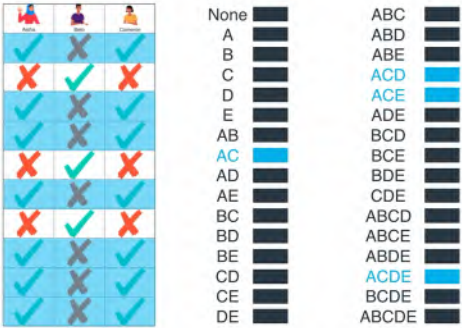
# Probability



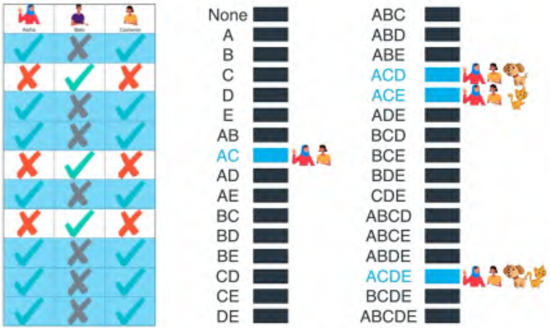
# Probability



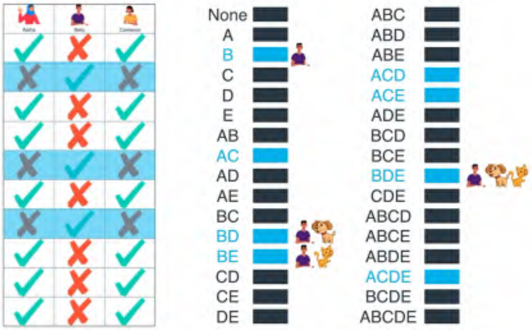
# Probability



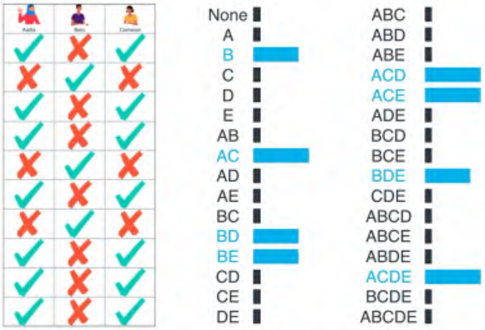
# Probability



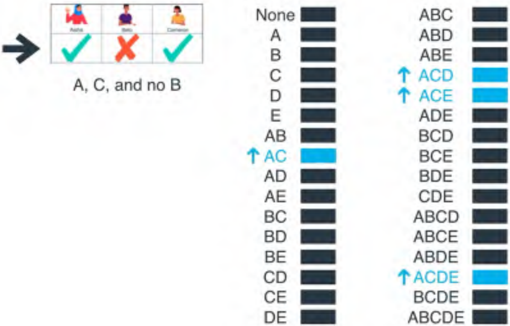
# Probability



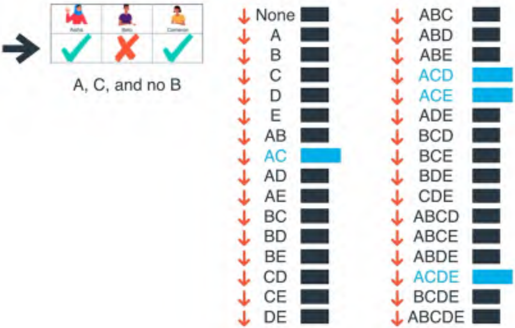
# Probability



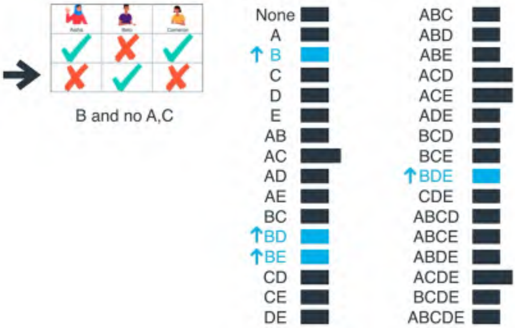
# Probability



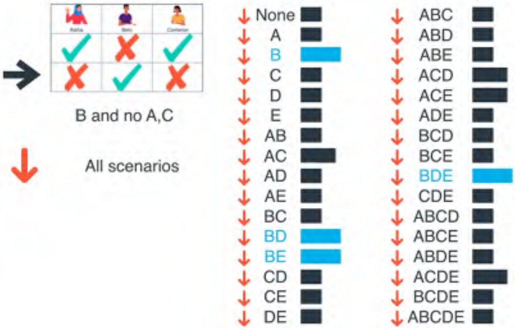
# Probability



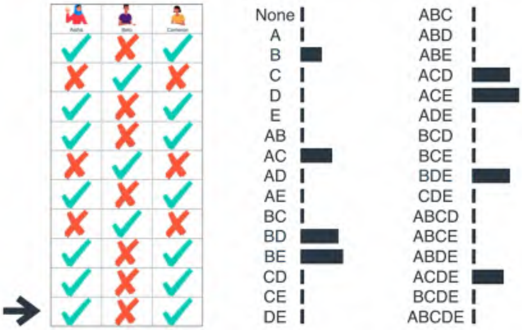
# Probability



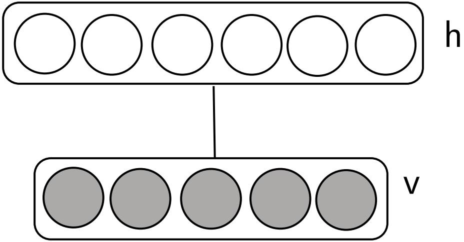
# Probability



# Probability



# Conditional distributions

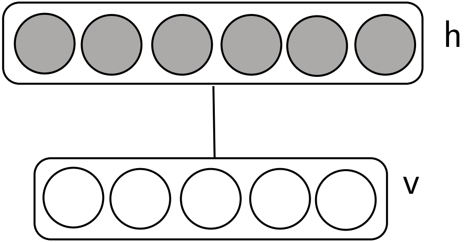



$$p(h|v) = \prod_j p(h_j|v)$$

$$p(h_j = 1|v) = sigm(a_j + W_j v)$$

- $W_j$ is the $j$th row of W

# Conditional distributions




$$p(v|h) = \prod_i p(v_i|h)$$

$$p(v_i = 1|h) = sigm(b_i + h^T W_i)$$

- $W_i$ is the $i$th column of W

# Inference

$$p(h|v) = p(v,h)/\sum_{h'} p(v,h')$$

$$= \frac{\exp(h^T W.v)\exp(b^T v)\exp(a^T h)/Z}{\sum_{h'\in\{0,1\}^H} \exp(h'^T Wv)\exp(b^T v)\exp(a^T h')/Z}$$

$$= \frac{\exp(\sum_j h_j W_j.v + a_j h_j)}{\sum_{h'_1\in\{0,1\}} \cdots \sum_{h'_H\in\{0,1\}} \exp(\sum_j h'_j W_j.v + a_j h'_j)}$$

$$= \frac{\prod_j \exp(h_j W_j.v + a_j h_j)}{\sum_{h'_1\in\{0,1\}} \cdots \sum_{h'_H\in\{0,1\}} \prod_j \exp(h'_j W_j.v + a_j h'_j)}$$

$$= \frac{\prod_j \exp(h_j W_j.v + a_j h_j)}{(\sum_{h'_1\in\{0,1\}} \exp(h'_1 W_1.v + a_1 h'_1)) \ldots (\sum_{h'_H\in\{0,1\}} \exp(h'_H W_H.v + a_H h'_H))}$$

# Inference

$$= \frac{\prod_j \exp(h_j W_j.v + a_j h_j)}{\prod_j (\sum_{h'_{j \in \{0,1\}}} \exp(h'_j W_j.v + a_j h'_j)}$$

$$= \frac{\prod_j \exp(h_j W_j.v + a_j h_j)}{\prod_j (1 + \exp(W_j.v + a_j))}$$

$$= \prod_j [\frac{\exp(h_j W_j.v + a_j h_j)}{1 + \exp(W_j.v + a_j)}]$$

$$= \prod_j p(h_j|v)$$

# Inference

$$
\begin{aligned}
p(h_j = 1|v) =& \frac{\exp(a_j + W_j.v)}{1 + exp(a_j + W_j.v)} \\
=& \frac{1}{1 + \exp(-a_j - W_j.v)} \\
=& sigm(a_j + W_j.v)
\end{aligned}
$$

# Free energy

- $p(v) = \sum_{h \in \{0,1\}^H} p(v,h) = \sum_{h \in \{0,1\}^H} \exp(-E(v,h))/Z$
- $p(v) = \exp(b^T v + \sum_{j=1}^{H} \log(1 + \exp(a_j + W_{j.} v)))/Z$
- $p(v) = \exp(-F(v))/Z$

# Inference

$$
\begin{aligned}
p(v) &= \sum_{h \in \{0,1\}^H} \exp(h^T W v + b^T v + a^T h)/Z \\
&= \exp(b^T v) \sum_{h_1 \in \{0,1\}} \cdots \sum_{h_H \in \{0,1\}} \exp(\sum_j h_j W_j v + a_j h_j)/Z
\end{aligned}
$$

# Inference

$$= \exp(b^T v)\Big(\sum_{h_1 \in \{0,1\}} \exp(h_1 W_1 v + a_1 h_1)\Big) \dots \Big(\sum_{h_H \in \{0,1\}} \exp(h_H W_H v + a_H h_H)\Big)/Z$$

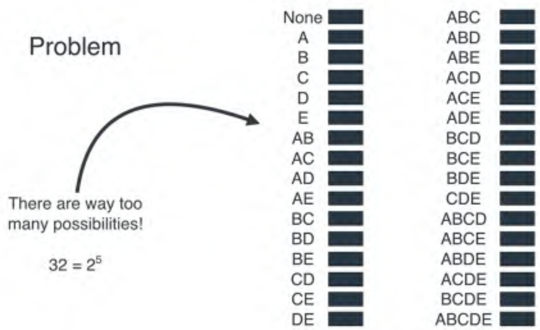
$$= \exp(b^T v)(1 + \exp(W_1 v + a_1)) \dots (1 + \exp(W_H v + a_H))/Z$$

$$= \exp(b^T v) \exp(\log(1 + \exp(W_1 v + a_1))) \dots \exp(\log(1 + \exp(W_H v + a_H)))/Z$$

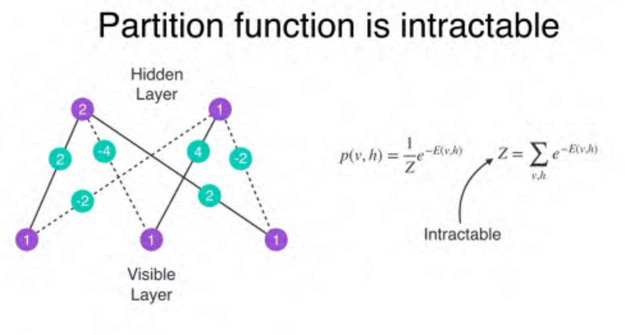
$$= \exp(b^T v + \sum_{j=1}^{H} \log(1 + \exp(W_j v + a_j)))/Z$$

# Some Details

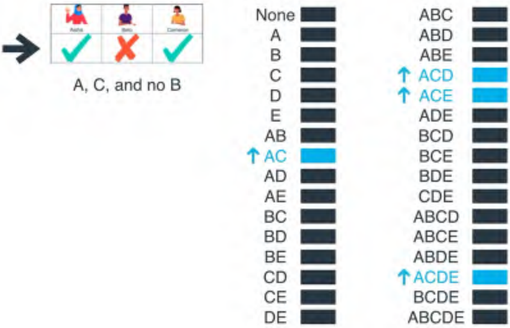
# Problem


Problem
There are way too many possibilities!
32 = 2^5
None
A
B
C
D
E
AB
AC
AD
AE
BC
BD
BE
CD
CE
DE
ABC
ABD
ABE
ACD
ACE
ADE
BCD
BCE
BDE
CDE
ABCD
ABCE
ABDE
ACDE
BCDE
ABCDE

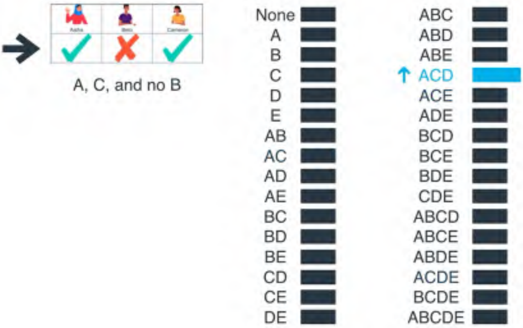
# Model


Partition function is intractable
Hidden Layer
Visible Layer
2
1
2
-4
4
-2
-2
2
1
1
1
$p(v,h) = \frac{1}{Z} e^{-E(v,h)}$
$Z = \sum_{v,h} e^{-E(v,h)}$
Intractable

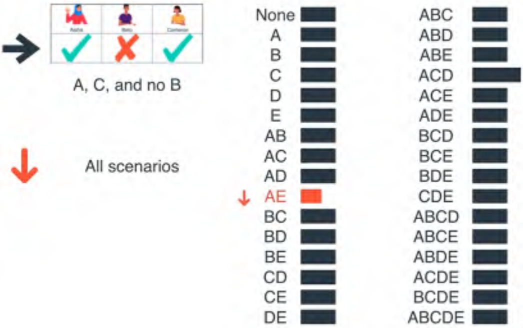
# Solution


A, C, and no B
None
A
B
C
D
E
AB
AC
AD
AE
BC
BD
BE
CD
CE
DE
ABC
ABD
ABE
ACD
ACE
ADE
BCD
BCE
BDE
CDE
ABCD
ABCE
ABDE
ACDE
BCDE
ABCDE

# Solution


A, C, and no B
None
A
B
C
D
E
AB
AC
AD
AE
BC
BD
BE
CD
CE
DE
ABC
ABD
ABE
ACD
ACE
ADE
BCD
BCE
BDE
CDE
ABCD
ABCE
ABDE
ACDE
BCDE
ABCDE

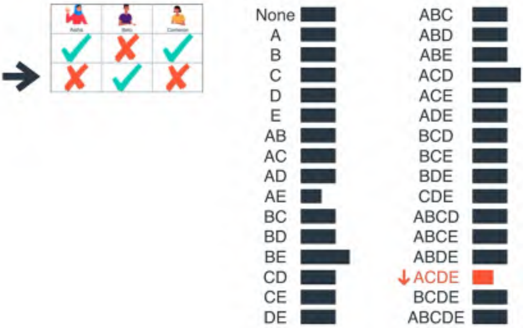
# Solution


A, C, and no B
All scenarios
None
A
B
C
D
E
AB
AC
AD
AE
BC
BD
BE
CD
CE
DE
ABC
ABD
ABE
ACD
ACE
ADE
BCD
BCE
BDE
CDE
ABCD
ABCE
ABDE
ACDE
BCDE
ABCDE

# Solution


None
A
B
C
D
E
AB
AC
AD
AE
BC
BD
BE
CD
CE
DE
ABC
ABD
ABE
ACD
ACE
ADE
BCD
BCE
BDE
CDE
ABCD
ABCE
ABDE
ACDE
BCDE
ABCDE

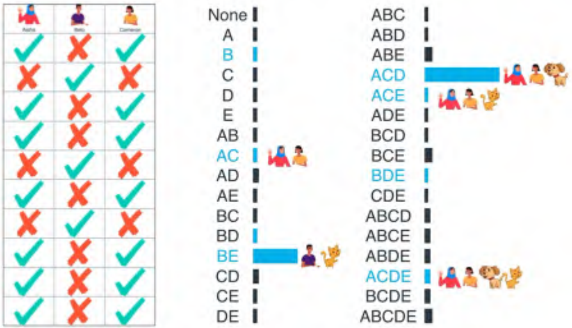
# Solution


None
A
B
C
D
E
AB
AC
AD
AE
BC
BD
BE
CD
CE
DE
ABC
ABD
ABE
ACD
ACE
ADE
BCD
BCE
BDE
CDE
ABCD
ABCE
ABDE
ACDE
BCDE
ABCDE

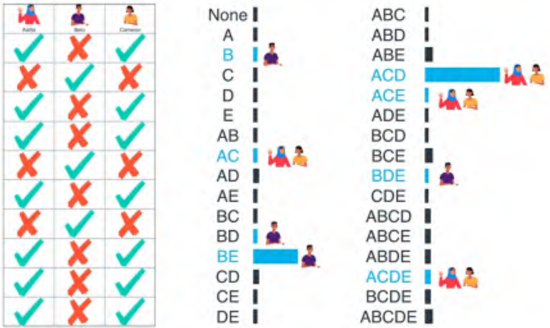
# Solution


None
A
B
C
D
E
AB
AC
AD
AE
BC
BD
BE
CD
CE
DE
ABC
ABD
ABE
ACD
ACE
ADE
BCD
BCE
BDE
CDE
ABCD
ABCE
ABDE
ACDE
BCDE
ABCDE

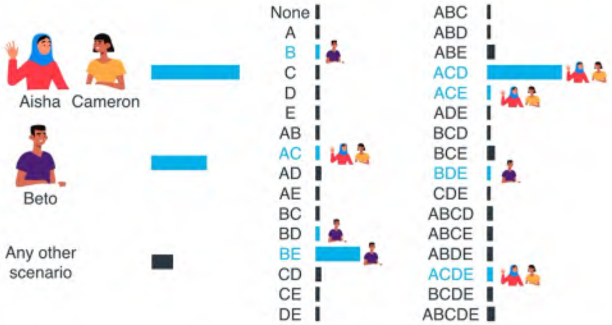
# Solution


None
A
B
C
D
E
AB
AC
AD
AE
BC
BD
BE
CD
CE
DE
ABC
ABD
ABE
ACD
ACE
ADE
BCD
BCE
BDE
CDE
ABCD
ABCE
ABDE
ACDE
BCDE
ABCDE

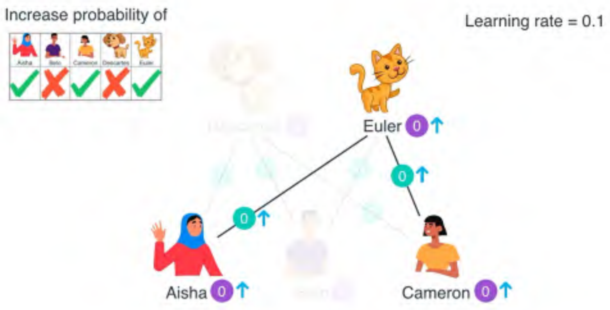
# Solution


None
A
B
C
D
E
AB
AC
AD
AE
BC
BD
BE
CD
CE
DE
ABC
ABD
ABE
ACD
ACE
ADE
BCD
BCE
BDE
CDE
ABCD
ABCE
ABDE
ACDE
BCDE
ABCDE

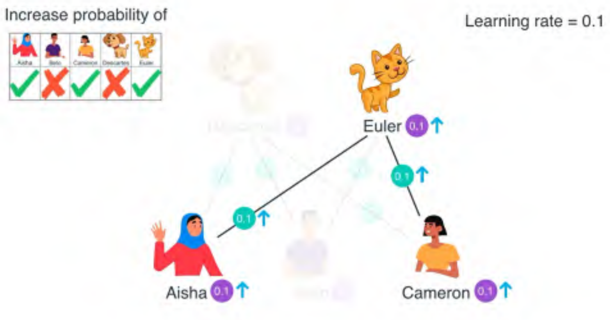
# Solution


Aisha
Cameron
Beto
Any other scenario
None
A
B
C
D
E
AB
AC
AD
AE
BC
BD
BE
CD
CE
DE
ABC
ABD
ABE
ACD
ACE
ADE
BCD
BCE
BDE
CDE
ABCD
ABCE
ABDE
ACDE
BCDE
ABCDE

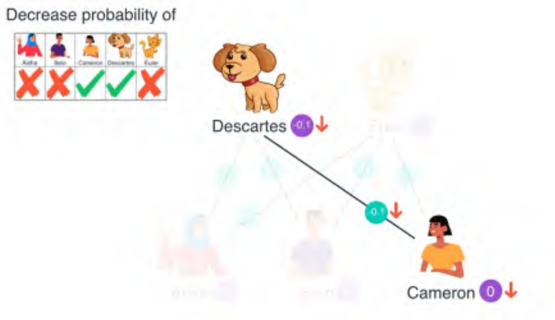
# How to change Probability



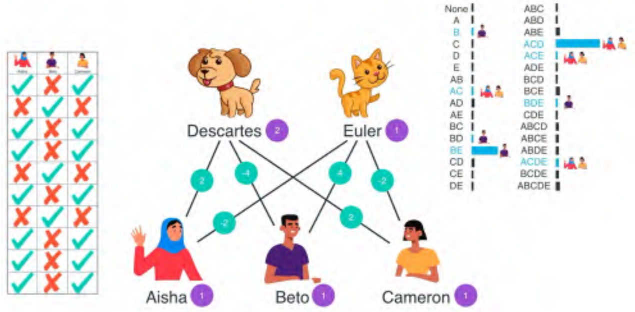
# How to change Probability



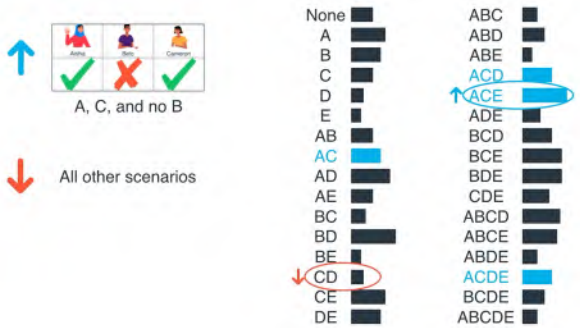
# How to change Probability



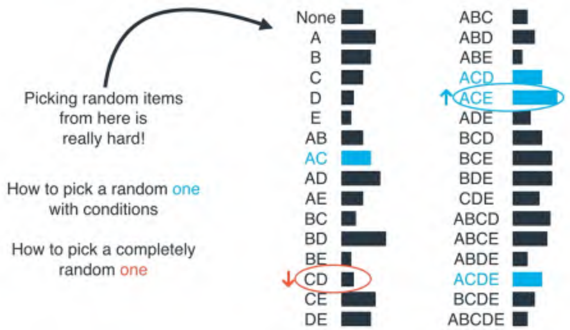
# How to change Probability



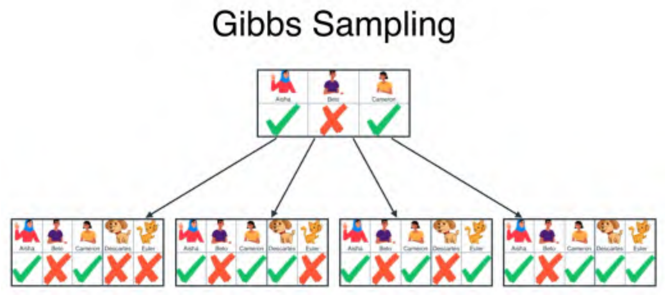
# How to Pick


A, C, and no B
All other scenarios
None
A
B
C
D
E
AB
AC
AD
AE
BC
BD
BE
CD
CE
DE
ABC
ABD
ABE
ACD
ACE
ADE
BCD
BCE
BDE
CDE
ABCD
ABCE
ABDE
ACDE
BCDE
ABCDE

# How to Pick


Picking random items
from here is
really hard!
How to pick a random one
with conditions
How to pick a completely
random one
None
A
B
C
D
E
AB
AC
AD
AE
BC
BD
BE
CD
CE
DE
ABC
ABD
ABE
ACD
ACE
ADE
BCD
BCE
BDE
CDE
ABCD
ABCE
ABDE
ACDE
BCDE
ABCDE

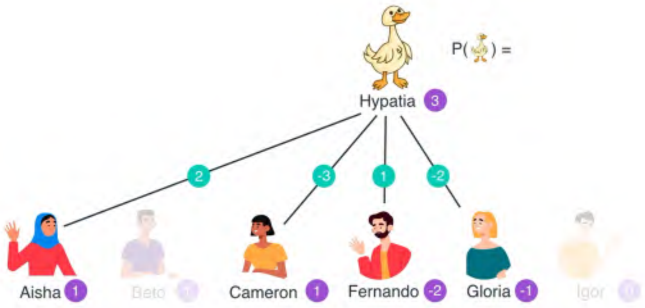
# How to Pick



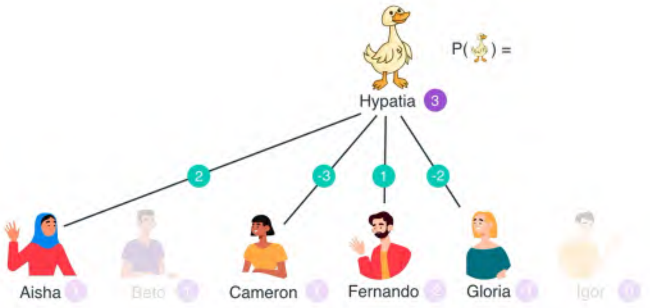
# How to Pick


Descartes 2
Euler 1
Hypatia 3
Aisha 1
Beto 1
Cameron 1
Fernando -2
Gloria -1
Igor 0

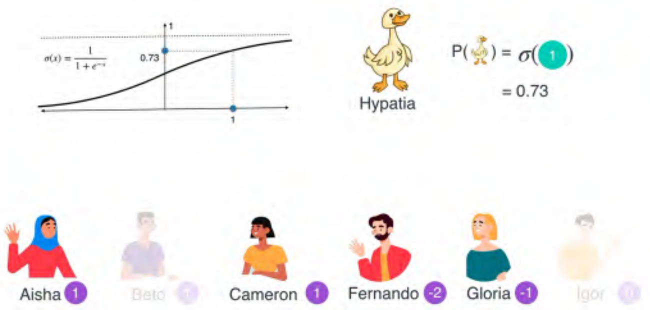
# How to Pick


P( ) =
Hypatia 3
2
-3
1
-2
Aisha 1
Beto 1
Cameron 1
Fernando -2
Gloria -1
Igor 0

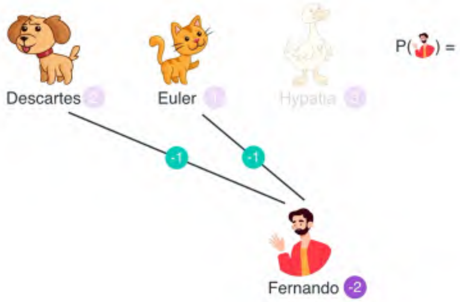
# How to Pick



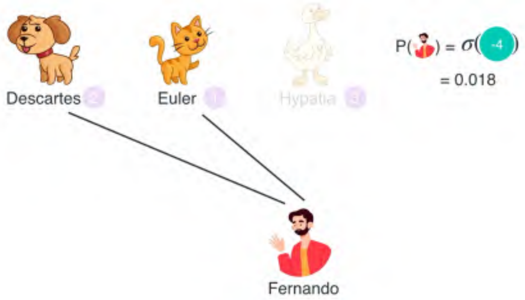
# How to Pick


$\sigma(x) = \frac{1}{1+e^{-x}}$
0.73
1
1
Hypatia
P( ) = σ( 1 )
= 0.73
Aisha 1
Beto 1
Cameron 1
Fernando -2
Gloria -1
Igor 0

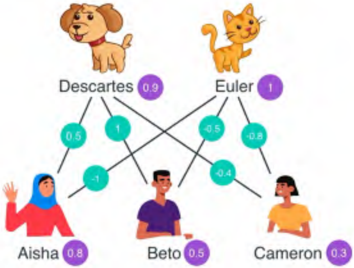
# How to Pick



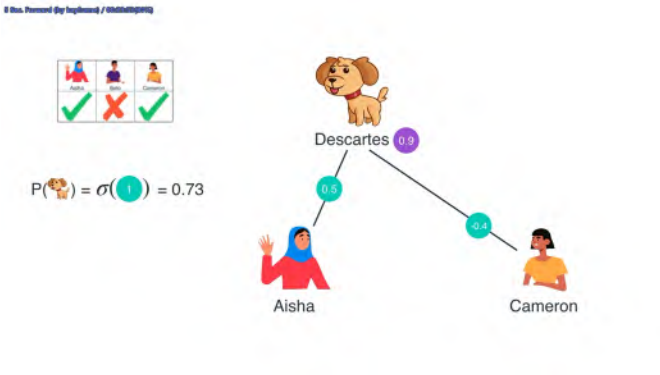
# How to Pick



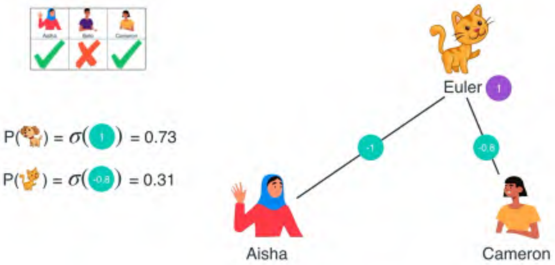
# How to Pick



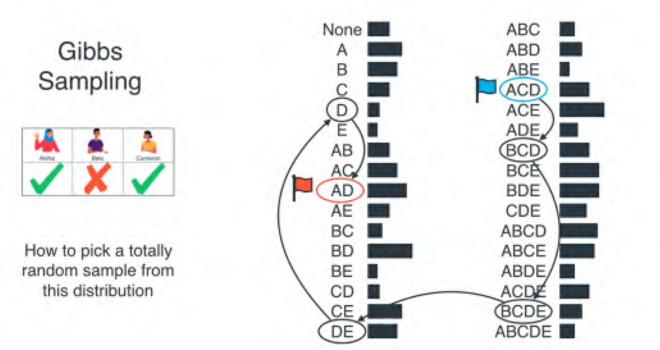
# How to Pick



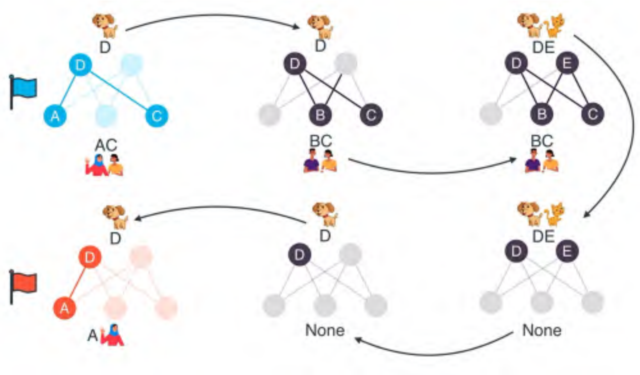
# How to Pick



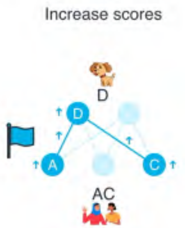
# How to Pick

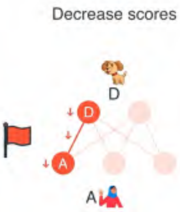

Gibbs
Sampling
How to pick a totally
random sample from
this distribution
None
A
B
C
D
E
AB
AC
AD
AE
BC
BD
BE
CD
CE
DE
ABC
ABD
ABE
ACD
ACE
ADE
BCD
BCE
BDE
CDE
ABCD
ABCE
ABDE
ACDE
BCDE
ABCDE

# How to Pick



# How to Pick

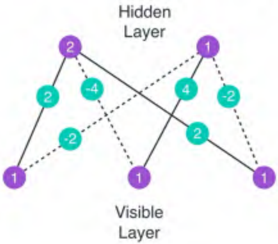


# Contrastive Divergence

# Model

## Maximizing the probability of the data




Find $\arg\max_{W} \prod_{v \in V} P(v)$

Maximize $\arg\max_{W} \mathbb{E}[\log P(v)]$

Derivative: $\frac{\partial}{\partial W} \log P(v_n)$

$$= \mathbb{E}\left[\frac{\partial}{\partial W} - E(v,h) \,|\, v = v_n\right] - \mathbb{E}\left[\frac{\partial}{\partial W} - E(v,h)\right]$$

# Traning objective

- Minimize the average negative log-likelihood (NLL)

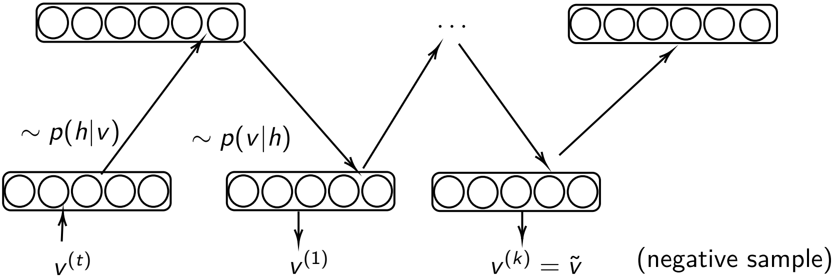
$$\frac{1}{T}\sum_t -\log p(v^{(t)})$$

- We'd like to proceed by stochastic gradient descent

$$\frac{\partial - \log p(v^{(t)})}{\partial\theta} = \mathbb{E}_h[\frac{\partial}{\partial\theta} - \mathbb{E}(v^{(t)}, h)|v^{(t)}] - \mathbb{E}_{v,h}[\frac{\partial}{\partial\theta}\mathbb{E}(v, h)]$$

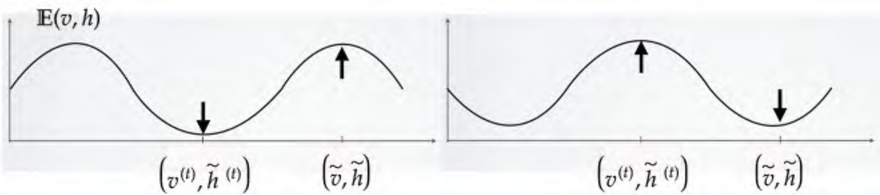
# Idea

- replace the expectation by a point estimate at $\tilde{v}$
- obtain the point $\tilde{v}$ by Gibbs sampling
- start sampling chain at $v^{(t)}$



# Approximation

- $\mathbb{E}_h[\frac{\partial}{\partial\theta}\mathbb{E}(v^{(t)}, h)|v^{(t)}] = \frac{\partial}{\partial\theta}\mathbb{E}(v^{(t)}, \tilde{h}^{(t)})$
- $\mathbb{E}_{v,h}[\frac{\partial}{\partial\theta}\mathbb{E}(v, h)] = \frac{\partial}{\partial\theta}\mathbb{E}(\tilde{v}, \tilde{h})$



# Inference

- $\frac{\partial}{\partial\theta}\mathbb{E}(v,h)$ for $\theta = W_{ji}$
- $\frac{\partial}{\partial W_{ji}}\mathbb{E}(v,h) = -h_j v_i$
- $\nabla_W \mathbb{E}(v,h) = -hv^T$
- $\mathbb{E}_h[\frac{\partial}{\partial\theta}\mathbb{E}(v,h)|v]$ for $\theta = W_{ji}$
- $\mathbb{E}_h[\frac{\partial}{\partial\theta}\mathbb{E}(v,h)|v] = \mathbb{E}_h[-h_j v_i] = -v_i p(h_j = 1|v)$
- $\mathbb{E}_h[\nabla_W \mathbb{E}_h(v,h)|v] = -h(v)v^T$

# Inference

Given $v^{(t)}$ and $\tilde{v}$ the learning rule for $\theta = W$ becomes

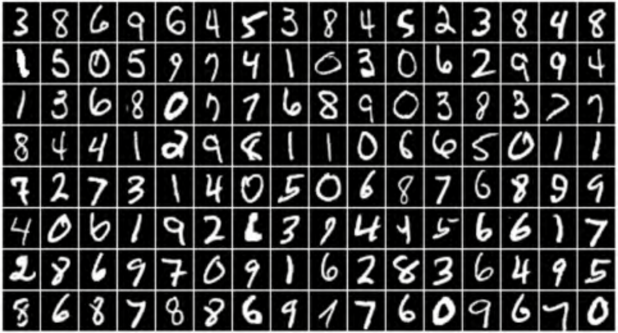
- $W_{new} = W - \alpha(\nabla_W - \log p(v^{(t)}))$
- $W_{new} = W - \alpha(\mathbb{E}_h[\nabla_W \mathbb{E}(v^{(t)}, h)|v^{(t)}] - \mathbb{E}_{v,h}[\nabla_W \mathbb{E}_(v, h)])$
- $W_{new} = W - \alpha(\mathbb{E}_h[\nabla_W \mathbb{E}(v^{(t)}, h)|v^{(t)}] - \mathbb{E}_v[\nabla_W \mathbb{E}(\tilde{v}, h)|\tilde{v}])$
- $W_{new} = W + \alpha(h(v^{(t)}){v^{(t)}}^T - h(\tilde{v})\tilde{v}^T)$

# Inference

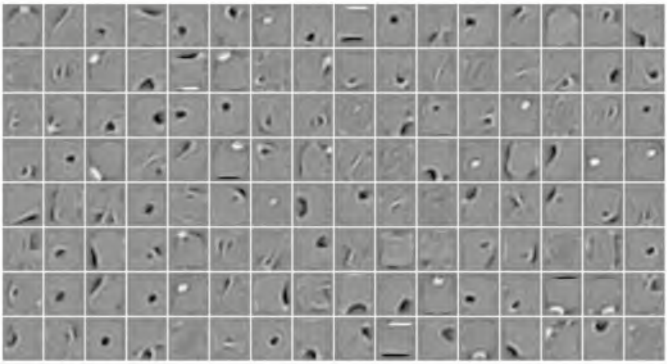
- $W_{new} = W + \alpha(h(v^{(t)})v^{(t)^T} - h(\tilde{v})\tilde{v}^T)$
- $a_{new} = a + \alpha(h(v^{(t)}) - h(\tilde{v}))$
- $b_{new} = b + \alpha(v^{(t)} - \tilde{v})$

# Application

# MNIST



# Filter



# Extensions

- k=1 Gibbs sampling
- Persistent CD
- Debugging
- RBM for unbounded reals
- Boltzmann machine & semi-restricted Boltzmann machine

# Thank you for your time

# Refrences

- A Practical Guide to Training Restricted Boltzmann Machines: +
- Examples are from Luis Serrano: +
- Neural networks class - Université de Sherbrooke +