NOTE BOOK

Rokshana Ahmed

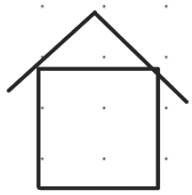
DEEP LEARNING

Date: 28 / 11 / 2023

Today RESTRICTED BOLTZMAN MACHINES (RBMs)

- Unsupervised Learning (Unlabelled data)
- Generative Model (Example: movie recommend.)
- In 2009, performed well during the NETFLIX CHALLENGE ($1 mil prize)

A MYSTERY



Q: Can you explain the data?

A: Alice & Carlos love dogs (hate cats)
Bob loves cats (hates dogs)

ALICE	BOB	CARLOS
✓	×	✓
✓	×	✓
×	✓	×
✓	×	✓
×	✓	×
×	✓	×
✓	×	✓

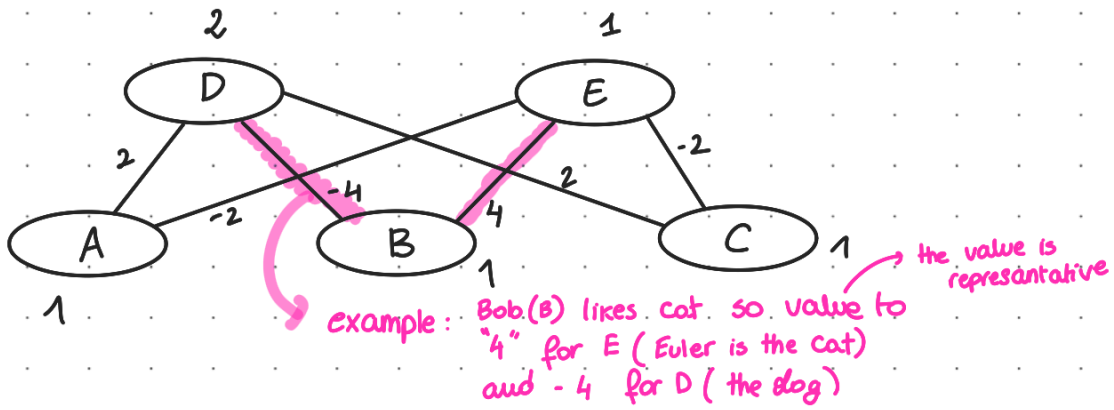
KEY IDEA

There might be HIDDEN PARAMETERS that explains the data

HIDDEN DATA

Neighbourhood dog = Descartes
" cat = Euler

names of the pets



DEEP LEARNING

Date: 28 / 11 / 2023

A scenario is a subset of the nodes (visible & hidden)

EX: ABE, it could not respect the rules of the data

Q: How many scenarios are there? $2^5 = 32$ ($2^{\#\text{nodes}}$ in general)

SCORE of a scenario = sum of weights of nodes in the scenario + " " " of edges with both ends in the scenario.

SCENARIO = ABCDE
SCORE = $2+1+1+1+1-2+2-4+4+2-2 = 6$

SCENARIO = ACD
SCORE = $1+1+2+2+2 = 8$

SCENARIO = BD
SCORE = $2+1-4 = -1$

SCENARIO	SCORE	PROBABILITIES
$\emptyset$	0	0
⋮	⋮	
BD	-1	≈ 0
⋮	⋮	
BE	7	≈ 0.17
⋮	⋮	
ACD	8	≈ 0.45
⋮	⋮	⋮

take SOFTMAX

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

DEFINITION

Each data point is a subset of the visible notes = $\{A, B, C\}$
A scenario extends a data point $X \subseteq \{A, B, C\}$ if $S \cap \{A, B, C\} = X$

Data point AC.
Scenario that extend AC are AC, ACD, ACE, ACDE = $4 = 2^2$ (the juiceee)
(In general = $2^{\#\text{hidden masks}}$)

DEEP LEARNING

Date: 28 / 11 / 2023

TRAINING

For each data point X, increase the probabilities of all scenarios that extend X, and decrease the probabilities of all scenarios that do not extend X.

Hidden layer = 100 nodes Visible layer = 200 nodes

A scenario that extend

X 2^{100} (more than # of elementary particles in the universe)

⋮

$2^{300} - 2^{100}$

HOW TO INCREASE THE PROBABILITY OF ONE SCENARIO?

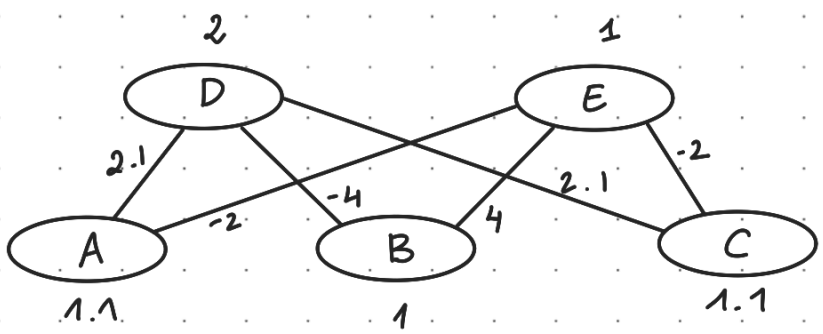
SCENARIO : ACD

learning rate = ℓ = 0.1

Increase node weights of A, C, and D by 0.1

Increase edge " of AD and CD by 0.1

Similarly, if we want to decrease the probability of ACD



DEEP LEARNING

Date: 28 / 11 / 2023

TRAINING

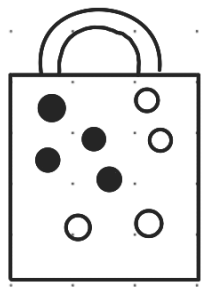
Given a data point X, pick a random scenario S that extends X and update weights to increase the probability of S.
Pick a random scenario T; that does not extend X and update weights to decrease the probability of T.

How do we do the random sampling ?

GIBBS sampling

RULES

1. Infinite number of black balls outside the bag
2. Cannot look in the bag
3. n° of balls in the bag must always be the same



WANT : Almost all the balls in the bag are black

SOLUTION

1. Take a random ball out of the bag, replace it with a black ball

---

(1) How to pick a random scenario that extends a data point x.

INDEPENDENT SAMPLING

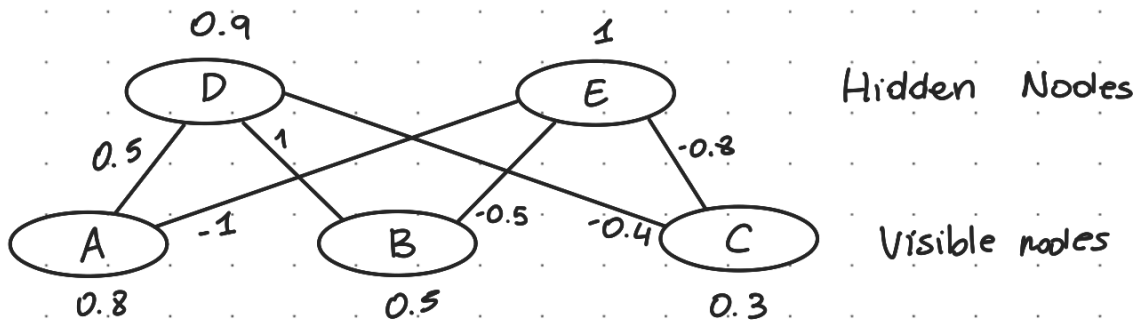
Suppose $X = \{A, C\}$

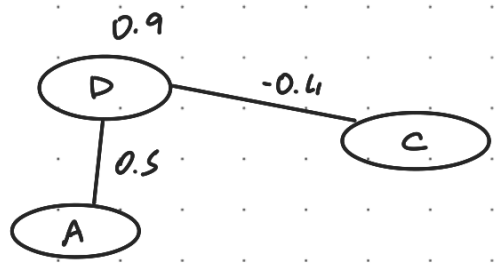
Q : What is the probability of D given that AC occured ?

A : $\sigma(0.9 + 0.5 - 0.4) = \sigma(1) = 0.73$

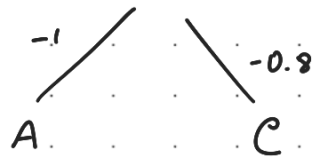
DEEP LEARNING

Date: 28 / 11 / 2023





$$\text{Prob}(E \mid AC) = \sigma(1 + (-1) + (-0.8))$$



$$= \sigma(-0.8) = 0.31$$

Random scenario extending AC is given by flipping 2 unfair coins.

Let H be a hidden node, and X be a data point.

$$\text{Prob}(H \mid X) = \sigma\Big(\text{weight}(H) + \sum_{x \in X} \text{weight}(xH)\Big)$$

Use this formula to compute a separate probability for each hidden node H.

Flip m coins (where m is the number of hidden nodes) to decide which hidden nodes to put into the scenario.

---

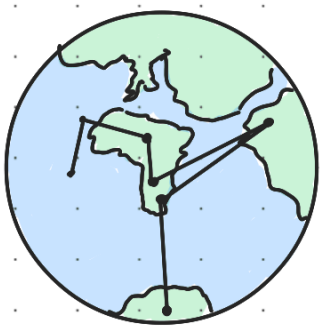
(2) How to sample a completely random scenario?

THIS IS HARD

So we will sample approximately

DEEP LEARNING

Date: 28 / 11 / 2023

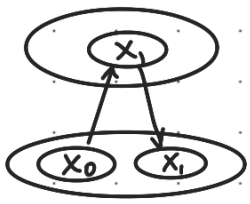


(pardon me for the weird world ._.)

Instead, we take a random walk.

The length of the random walk is a hyperparameter

Start with random $X \subseteq \{\text{visible nodes}\}$



HIDDEN NODES

VISIBLE NODES

(1) Generate a random subset X_1 of hidden nodes using the formula.

(2) Generate a random subset X_2 of visible nodes using the formula.

Repeat to obtain X_0, X_1, X_2, ..., X_p p = hyperparameter.

$X_0, X_2, X_4, \ldots$ are subsets of visible nodes
$X_1, X_3, X_5, \ldots$ " " " hidden nodes

OUTPUT $\boxed{X_{p-1} \cup X_p}$

e.g. hidden nodes = feature
" nodes = movies