CS623: Introduction to Computing with Neural Nets *(lecture-17)*

Pushpak Bhattacharyya
Computer Science and Engineering Department
IIT Bombay

Boltzmann Machine

- Hopfield net
- Probabilistic neurons
- Energy expression = $-\Sigma_i \Sigma_{j>i} w_{ij} x_i x_j$
  where x_i = activation of i^{th} neuron
- Used for optimization
- Central concern is to ensure global minimum
- Based on simulated annealing

Comparative Remarks

Feed forward n/w with BP	**Hopfield net**	**Boltzmann m/c**
Mapping device: (i/p pattern --> o/p pattern), *i.e.* Classification	Associative Memory + Optimization device	Constraint satisfaction. (Mapping + Optimization device)
Minimizes total sum square error	Energy	Entropy (Kullback–Leibler **divergence**)

Comparative Remarks (contd.)

Feed forward n/w with BP	**Hopfield net**	**Boltzmann m/c**
Deterministic neurons	Deterministic neurons	Probabilistic neurons
Learning to associate i/p with o/p *i.e.* equivalent to a function	Pattern	Probability Distribution

Comparative Remarks (contd.)

Feed forward n/w with BP	**Hopfield net**	**Boltzmann m/c**
Can get stuck in local minimum (Greedy approach)	Local minimum possible	Can come out of local minimum
Credit/blame assignment (consistent with Hebbian rule)	Activation product (consistent with Hebbian rule)	Probability and activation product (consistent with Hebbian rule)

Theory of Boltzmann m/c

- For the m/c the computation means the following:

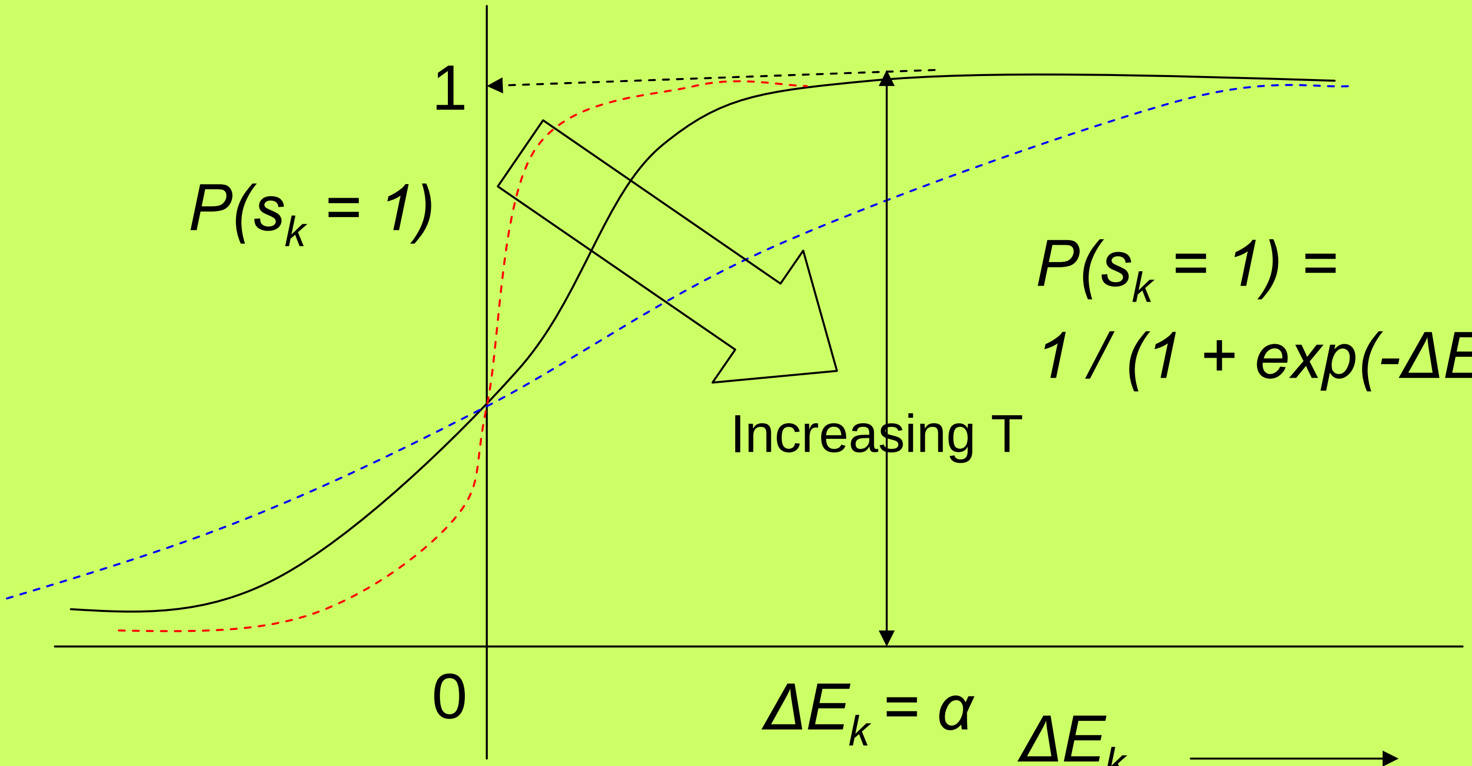
  At any time instant, make the state of the k^{th} neuron (s_k) equal to 1, with probability:

$$1 / (1 + exp(-\Delta E_k / T))$$

ΔE_k = change in energy of the m/c when the k^{th} neuron changes state

T = temperature which is a parameter of the m/c

Theory of Boltzmann m/c (contd.)



$P(s_k = 1) =$

$1 / (1 + exp(-\Delta Ek / T))$

Theory of Boltzmann m/c (contd.)

$\Delta E_k = E^k{}_{final} – E^k{}_{initial}$

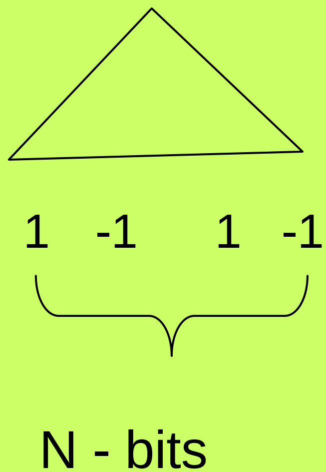
$= (s^{initial}{}_k - s^{final}{}_k) * \Sigma_{j\neq k} w_{kj} s_j$

We observe:

1. The higher the temperature, lower is $P(S_k=1)$
2. at T = *infinity*, $P(S_k=1) = P(S_k=0) = 0.5$, equal chance of being in state 0 or 1. Completely random behavior
3. If $T \rightarrow 0$, then $P(S_k=1) \rightarrow 1$
4. The derivative is proportional $P(S_k=1)*(1 - P(S_k=1))$

Consequence of the form of $P(S_k=1)$

$$P(S_\alpha)\ \alpha\ exp(\ -E(S_\alpha))\ /\ T$$

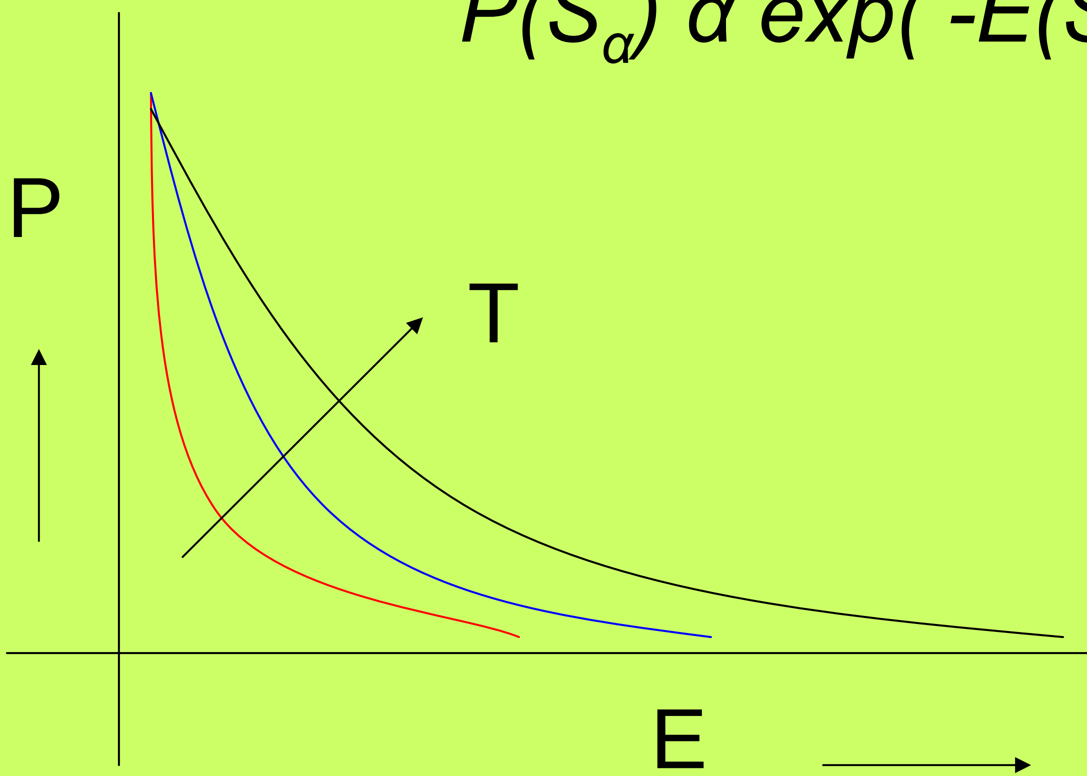


Probability Distribution called as Boltzmann Distribution

$P(S_\alpha)$ is the probability of the state S_α

Local "sigmoid" probabilistic behavior leads to global Boltzmann Distribution behaviour of the n/w

$$P(S_\alpha)\ \alpha\ exp(\ -E(S_\alpha))\ /\ T$$



Ratio of state probabilities

- Normalizing,

  $P(S_\alpha) = (exp(-E(S_\alpha)) / T) / (\sum_{\beta \in all\ states} exp(-E(S_\beta)/T)$

  $P(S_\alpha) / P(S_\beta) = exp -(E(S_\alpha) - E(S_\beta)) / T$

Learning a probability distribution

- Digression: Estimation of a probability distribution Q by another distribution P

- D = deviation = $\sum_{sample\ space} Q \ln Q/P$
- $D >= 0$, which is a required property (just like sum square error >= 0)

To prove $D >= 0$

Lemma: $ln\ (1/x) >= (1 - x)$

Let, $x = 1 / (1 - y)$

$$ln(1 - y) = -[y + y^2/2 + y^3/3 + y^4/4 + \ldots]$$

Proof (contd.)

$(1 – x) = 1 – 1/(1 - y)$

$= -y(1 - y)^{-1}$

$= -y(1 + y + y^2 + y^3 + \ldots)$

$= -[y + y^2 + y^3 + y^4 + \ldots]$

But, $y + y^2/2 + y^3/3 + \ldots <= y + y^2 + y^3 + \ldots$

Lemma proved.

Proof (contd.)

$D = \sum Q \, ln \, Q / P$

$>= \sum_{over\ the\ sample\ space} Q(1 - P/Q)$

$= \sum (Q - P)$

$= \sum Q - \sum P$

$= 1 - 1 = 0$