**17th Nov Machine Learning**

ML

-supervised - back propagation

- i/o HMM

-Unsupervised - HMM - i/o HMM

- Clustering

**“Boltzmann Machines”**

Comparable to a “Universal Computational Gas” that uses the laws of Thermo dynamics

Relationship proportional to the configuration of the molecules and energy

P(α) ~ e-E(α) / T

(Probability of finding specific configuration of molecules)

**JJ Hoppfield Units**

Imagine system of neurons (Units)

Zj

wij

zi

z element +/- 1

Total input to i = zj wij  (symbol is sum over j)

Units change to get to lower energy on = +1 off = -1

Both zi and zj on means a higher wij

zj = g( zj wij ) (symbol is sum over j)

g(f)

+1 Energy goes down as state of unit changes

System settles to energy minimum

-1

“Associative Memory” -> enough partial information can recollect the whole memory

Simulated Annealing – Same structure as Hoppfield Units.

(see [annealing in metallurgy](http://en.wikipedia.org/wiki/Annealing_%28metallurgy%29))

Problem – How to put transistors on VLSI chips – Best minimum?

We want to minimise the wires between the chips.

Add randomness and keep track of temperature.

As temperature goes to 0 – use the g(f) graph above.

As temperature is finite and > 0 – add randomness to crawl out of one minimum of the system to another

Taking Hoppfield + Simulated Annealing together gives rise to the

Boltzmann Machine (- learning)

Boltzmann Distribution

P(α) = e-E(α) / T

P(zi = +1) = 1 / (1 + e-(1/T)( zi wij) ) (symbol is sum over j)

Which side of this cube is closer? Can be interpreted 2 ways

Use Boltzmann Machine with different weights on each side

Let’s look at an example.

hidden

P(α β) proportional eE(α β) / T

β

α

Visible

Let P+(α) - be a liked distribution (frames of an image on the t.v.)

Given half of α, can you get all of the frames?

Clamp down known units in place and run the distribution (conditional)

I.e. given P+(α) examples -> find the weights of the units

This is the learning aspect.

G = KL (P+(α) || P-(α)) - Kullback Leibler divergence

P- = exhibited by the machine

This is how many bits of info do you learn if you thought P+ but were given P-

Delta wij  = - error ( dG/ d wij)

Relationship = -log P(α) ~ (1/T) (E(α))

G = [ P+ (α) log(P+ (α) / P- (α)) ] (symbol is sum over alpha)

P+ (α) = P+ (α β)

dG/ d wij = (1/T) (p+ij – p-ij) closer they are-> Higher the weight

pij = P (zi = zj )

Given a boltzman Machine

Visible Hidden

Let the contained weights be the same

t This Is a HMM – which is a special case of

Boltzmann Machine