B. Stochastic Neural Networks

(in particular, the stochastic Hopfield network)

Motivation

- **Idea**: with low probability, go against the local field
  - move up the energy surface
  - make the “wrong” microdecision
- **Potential value for optimization**: escape from local optima
- **Potential value for associative memory**: escape from spurious states
  - because they have higher energy than imprinted states

The Stochastic Neuron

Deterministic neuron: \( x'_i = \text{sgn}(h_i) \)
- \( \Pr\{x'_i = +1\} = \Theta(h_i) \)
- \( \Pr\{x'_i = -1\} = 1 - \Theta(h_i) \)

Stochastic neuron:
- \( \Pr\{x'_i = +1\} = \alpha(h_i) \)
- \( \Pr\{x'_i = -1\} = 1 - \alpha(h_i) \)

Logistic sigmoid: \( \alpha(h) = \frac{1}{1 + \exp(-2h/T)} \)
Properties of Logistic Sigmoid

\[ \sigma(h) = \frac{1}{1 + e^{-2ht}} \]

- As \( h \rightarrow +\infty \), \( \sigma(h) \rightarrow 1 \)
- As \( h \rightarrow -\infty \), \( \sigma(h) \rightarrow 0 \)
- \( \sigma(0) = 1/2 \)

Logistic Sigmoid

\( T \) varying from 0.05 to \( \infty \) (\( 1/T = \beta = 0, 1, 2, \ldots, 20 \))
**Part 3B: Stochastic Neural Networks**

---

**Logistic Sigmoid**

\[ T = 10 \]

---

**Logistic Sigmoid**

\[ T = 100 \]

---

**Pseudo-Temperature**

- Temperature = measure of thermal energy (heat)
- Thermal energy = vibrational energy of molecules
- A source of random motion
- Pseudo-temperature = a measure of nondirected (random) change
- Logistic sigmoid gives same equilibrium probabilities as Boltzmann-Gibbs distribution

---

**Transition Probability**

Recall, change in energy \( \Delta E = -\Delta s_i h_i \)

\[
\Delta s_i h_i = 2s_i h_i
\]

\[
\Pr \{ s_i' = \pm 1 | s_i = \mp 1 \} = \sigma(s_i h_i) = \sigma(-s_i h_i)
\]

\[
\Pr \{ s_i \rightarrow -s_i \} = \frac{1}{1 + \exp(\Delta s_i h_i / T)}
\]

\[
= \frac{1}{1 + \exp(\Delta E / T)}
\]

---

**Stability**

- Are stochastic Hopfield nets stable?
- Thermal noise prevents absolute stability
- But with symmetric weights:
  - average values \( \langle s_i \rangle \) become time - invariant

---

**Does “Thermal Noise” Improve Memory Performance?**

- Experiments by Bar-Yam (pp. 316-20):
  - \( n = 100 \)
  - \( p = 8 \)
- Random initial state
- To allow convergence, after 20 cycles set \( T = 0 \)
- How often does it converge to an imprinted pattern?
Part 3B: Stochastic Neural Networks

Probability of Random State Converging on Imprinted State \((n=100, p=8)\)

\[
T = \frac{1}{\beta}
\]

(fig. from Bar-Yam)

Probability of Random State Converging on Imprinted State \((n=100, p=8)\)

Analysis of Stochastic Hopfield Network

- Complete analysis by Daniel J. Amit & colleagues in mid-80s
- The analysis is beyond the scope of this course

Phase Diagram

(A) imprinted = minima
(B) imprinted, but s.g. = min.
(C) spin-glass states
(D) all states melt

(fig. from Domany & al. 1991)

Conceptual Diagrams of Energy Landscape

(fig. from Hertz & al. Int. Theory Neur. Comp.)

Phase Diagram Detail

(fig. from Domany & al. 1991)
Simulated Annealing

(Kirkpatrick, Gelatt & Vecchi, 1983)

Dilemma

• In the early stages of search, we want a high temperature, so that we will explore the space and find the basins of the global minimum
• In the later stages we want a low temperature, so that we will relax into the global minimum and not wander away from it
• Solution: decrease the temperature gradually during search

Quenching vs. Annealing

• Quenching:
  – rapid cooling of a hot material
  – may result in defects & brittleness
  – local order but global disorder
  – locally low-energy, globally frustrated
• Annealing:
  – slow cooling (or alternate heating & cooling)
  – reaches equilibrium at each temperature
  – allows global order to emerge
  – achieves global low-energy state

Multiple Domains

Moving Domain Boundaries

Effect of Moderate Temperature
**Effect of High Temperature**

$\Delta E/T$ low

**Effect of Low Temperature**

$\Delta E/T$ high

**Annealing Schedule**

- Controlled decrease of temperature
- Should be sufficiently slow to allow equilibrium to be reached at each temperature
- With sufficiently slow annealing, the global minimum will be found with probability 1
- Design of schedules is a topic of research

**Typical Practical Annealing Schedule**

- **Initial temperature** $T_0$ sufficiently high so all transitions allowed
- **Exponential cooling**: $T_{k+1} = \alpha T_k$
  - typical $0.8 < \alpha < 0.99$
  - at least 10 accepted transitions at each temp.
- **Final temperature**: three successive temperatures without required number of accepted transitions

**Summary**

- Non-directed change (random motion) permits escape from local optima and spurious states
- Pseudo-temperature can be controlled to adjust relative degree of exploration and exploitation

**Hopfield Network for Task Assignment Problem**

- Six tasks to be done (I, II, …, VI)
- Six agents to do tasks (A, B, …, F)
- They can do tasks at various rates
  - A (10, 5, 4, 6, 5, 1)
  - B (6, 4, 9, 7, 3, 2)
  - etc
- What is the optimal assignment of tasks to agents?
NetLogo Implementation of Task Assignment Problem

Run TaskAssignment.nlogo

Additional Bibliography