B. Stochastic Neural Networks

(in particular, the stochastic Hopfield network)

Trapping in Local Minimum

Escape from Local Minimum

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: $x'_i = \text{sgn}(h_i)$

\[
\begin{align*}
\Pr\{x'_i = +1\} &= \Theta(h_i) \\
\Pr\{x'_i = -1\} &= 1 - \Theta(h_i)
\end{align*}
\]

Stochastic neuron:

\[
\begin{align*}
\Pr\{x'_i = +1\} &= \sigma(h_i) \\
\Pr\{x'_i = -1\} &= 1 - \sigma(h_i)
\end{align*}
\]

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

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

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

Logistic Sigmoid

\( T = 0.5 \)

Slope at origin = \( 1 / 2T \)

Logistic Sigmoid

\( T = 0.01 \)

Logistic Sigmoid

\( T = 0.1 \)

Logistic Sigmoid

\( T = 1 \)

\( T \) varying from 0.05 to \( \infty \) (\( 1/T = \beta = 0, 1, 2, \ldots, 20 \) )
Logistic Sigmoid

\[ T = 10 \]

\[ T = 100 \]

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 \( \Delta E = -\Delta s_i h_i \)

\[ = 2s_i h_i \]

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

\[ \Pr\{s_i \to -s_i\} = \frac{1}{1 + \exp(2s_i h_i / T)} \]

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

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

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

Conceptual Diagrams of Energy Landscape

Phase Diagram Detail
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

Effect of Low Temperature

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

Additional Bibliography