Machine Intelligence II — **Exercise Sheet 9**

SS2012, Prof. Obermayer — due: 28.06.2012

Stochastic Optimization

9.1 Simulated Annealing (5 points)

Simulated annealing can be used to optimize a *cost function* $E : \boldsymbol{S} \rightarrow \mathbb{R}$ where the state $\boldsymbol{S}$ is a set of discrete state variables $s_i \in \{-1, +1\},\ i = 1, \ldots, N$. For a fully connected "network" with $N = 6$ binary nodes, this means that $\boldsymbol{S} \in \{-1, +1\}^6$, and we will use the cost ("energy")

$$E(\boldsymbol{S}) = -\frac{1}{2} \sum_{i,j=1}^{N} w_{ij} s_i s_j,$$

where $w_{ij} = w_{ji} \in \mathbb{R}$, and $w_{ii} = 0$.

The probability that the network is in a state $\boldsymbol{S}$ with energy $E(\boldsymbol{S})$ is given by

$$P(\boldsymbol{S}) = \frac{1}{Z} \exp\big(-\beta E(\boldsymbol{S})\big),$$

where the *partition function* Z guarantees $P(S)$ to be a valid probability mass function and is given as the sum over all possible configurations, i.e.

$$Z = \sum_{\boldsymbol{S}} \exp\big(-\beta E(\boldsymbol{S})\big).$$

Write a program that finds the optimal configuration $\boldsymbol{S}$ for a given set of weights $\boldsymbol{W}$. It should execute the following steps:

Initialization:

- β_0 small enough; $\tau > 1$; set t_{max}
- $\boldsymbol{S}_0$ randomly; $\boldsymbol{W}_0$ randomly, but symmetrically and with zero diagonal

Optimization: for each iteration $t = 0, ..., t_{max}$

- Select node i with state s_i randomly.
- Compute local energies and their difference

  $$E_{s_i} = -\frac{1}{2} \sum_{j \in \mathcal{N}_i} w_{ij} s_i s_j \qquad \text{and} \qquad E_{-s_i} = -E_{s_i} \qquad \rightarrow \qquad \Delta E = E_{-s_i} - E_{s_i}$$

  where $\mathcal{N}_i$ is the set of neighbors of node i, i.e. here the set of all other nodes.
- Flip state s_i with probability $P(s_i \rightarrow -s_i) = (1 + e^{\beta_t \Delta E})^{-1}$.
- Increase β using $\beta_{t+1} = \tau \beta_t$.

Plotting:

- Plot the temperature $T_t = \frac{1}{\beta_t}$ and the energy E_t over the iterations $t = 0, \ldots, t_{max}$.
- Plot the energy $E(\boldsymbol{S})$ for all possible 2^6 states as bar plot. The sequence of the states is not relevant. Additionally, plot the probabilities $P(\boldsymbol{S})$ for different βs as a bar plot. Choose the βs in a way, that the probability distributions differ discernibly.

9.2 Mean-Field Annealing (5 points)

Mean-field annealing is a deterministic approximation of simulated annealing. During optimization the nodes have continuous instead of binary states. These states represent the mean with respect to the factorized distribution $Q(\boldsymbol{S}) \approx P(\boldsymbol{S})$.

Consider again a fully connected network with $N = 6$ nodes, but now with state space $\boldsymbol{S} \in [-1, +1]^6$. The cost (energy) function remains the same:

$$E(\boldsymbol{S}) = -\frac{1}{2} \sum_{i,j=1}^{N} w_{ij} s_i s_j,$$

where the $w_{ij} \in \mathbb{R}$ are symmetric, and $w_{ii} = 0$. The approximated probability of a state $\boldsymbol{S}$ is now given by

$$Q(\boldsymbol{S}) = \frac{1}{Z_Q} \exp\Big(-\beta \sum_j e_j s_j \Big).$$

Write a program that finds the optimal configuration $\boldsymbol{S}$ of the network for given weights $\boldsymbol{W}$. It should execute the following steps:

Initialization:

- β_0 small enough, $\tau > 1$, set t_{max}
- $\boldsymbol{S}_0$ randomly, $\boldsymbol{W}_0$ randomly, but symmetrically and with zero diagonal

Optimization: for each iteration $t = 0, ..., t_{max}$

- Select node i with state s_i randomly.
- Compute mean-fields

$$e_i = \sum_{j \in \mathcal{N}_i} w_{ij} s_j$$

  where $\mathcal{N}_i$ is the set of neighbors of node i, i.e. here the set of all other nodes.
- Update the state using $s_i = \tanh(\beta e_i)$
- Increase β using $\beta_{t+1} = \tau \beta_t$

Plotting:

- Plot the temperature $T_t = \frac{1}{\beta_t}$ and the energy E_t over the iterations $t = 0, \ldots, t_{max}$.

Compare the number of iterations until convergence for simulated and mean-field annealing.

Total points: 10