

---

## Stochastic Optimization

### Exercise T7.1: Simulated Annealing and Mean-Field Annealing (tutorial)

- (a) What is the Exploration-Exploitation tradeoff?
- (b) How does simulated (stochastic) annealing optimize discrete variables?
- (c) How do the stationary distribution, the transition probability, the cost function change with the inverse temperature?
- (d) How does Mean-field (deterministic) annealing optimize discrete variables?

### Exercise H7.1: Simulated Annealing (homework, 5 points)

Simulated annealing can be used to optimize a cost function  $E : \underline{s} \rightarrow \mathbb{R}$  where the state  $\underline{s}$  is a set of discrete state variables  $s_i \in \{-1, +1\}$ ,  $i = 1, \dots, N$ . For an all-to-all connected “network” with  $N = 6$  binary nodes, this means that  $\underline{s} \in \{-1, +1\}^6$ . Our cost (“energy”) is defined as follows:

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

where  $w_{ij} = w_{ji} \in \mathbb{R}$ , and  $w_{ii} = 0$ .  
The probability that the network is in a specific state  $\underline{s}$  with energy  $E(\underline{s})$  is given by

$$P(\underline{s}) = \frac{1}{Z} \exp(-\beta E(\underline{s})),$$

where the *partition function*  $Z$  guarantees  $P(\underline{s})$  to be a valid probability mass function and is given as the sum over all possible configurations, i.e.  $Z = \sum_{\underline{s}} \exp(-\beta E(\underline{s}))$ .

Write a program that finds the optimal configuration  $\underline{s}^*$  for a given set of weights  $\underline{\mathbf{W}}$  as follows

#### A: Initialization:

1. Set both  $\beta_0, \tau > 1$  (and that both are close to 1)<sup>1</sup>, and  $t_{max} > 150$ . Try both  $M = 1$  and  $M = 20$ .
2. Set the initial state  $\underline{s}$  randomly;
3. Set  $\underline{\mathbf{W}}$  arbitrarily, e.g.  $w_{ij} \in [-1, 1]$  uniformly random, but force (also during optimization)  $\underline{\mathbf{W}}$  to be symmetric and have a zero diagonal.

#### B: Optimization: for each iteration $t = 0, \dots, t_{max}$

1. Repeat the following  $M$  times (state update loop):
  - (i) Select node  $i$  randomly

---

<sup>1</sup>By setting  $\beta > 1$  we will have actually already skipped the exploration phase. We recommend setting it slightly above 1, e.g.  $\beta_0, \tau = 1.01$ .

- (ii) Determine the energies for the two options for flipping  $s_i$  and compute their difference

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

where  $\mathcal{N}_i$  is the set of neighbors of node  $i$ . Since we have all-to-all connections the set of neighbors is effectively **all other** nodes.

- (iii) flip state  $s_i$  with probability  $P(s_i \rightarrow -s_i) = \frac{1}{(1 + e^{\beta_t \Delta E})}$

2. Increase  $\beta$  using  $\beta_{t+1} = \tau \beta_t$  before iterating over the state update loop another  $M$  times.

**C: Visualization:** For both  $M = 1$  and  $M = 20$  in the same figure.

1. Plot the temperature  $T_t = \frac{1}{\beta_t}$  and the energy  $E(\underline{s})$  over the iterations  $t = 0, \dots, t_{max}$ .
2. Show the energy  $E(\underline{s})$  for all possible  $2^6$  states using a bar plot. The order of the states is not relevant. Additionally, plot the probabilities  $P(\underline{s})$  for different  $\beta$  as a bar plot. There is no need to do this for all values of  $\beta$ , simply choose  $\{\beta\}$  in a way, that the probability distributions differ sufficiently to illustrate the effect of  $\beta$  on  $P(\underline{s})$  (e.g.  $\beta$  changes by one order of magnitude).

### Exercise H7.2: Mean-Field Annealing

(homework, 5 points)

Mean-field annealing is a deterministic approximation of simulated annealing. During optimization each node  $s_i$  corresponds to the first moment of the original  $i$ -th binary variable and thus takes continuous values. The states  $s_i$  represent the mean with respect to the factorized distribution  $Q(\underline{s}) \approx P(\underline{s})$ .

Consider again an all-to-all connected network with  $N = 6$  nodes. The cost (energy) function remains the same:

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

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

$$Q(\underline{s}) = \frac{1}{Z_Q} \exp\left(-\beta \sum_j e_j s_j\right).$$

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

#### A: Initialization:

1.  $\beta_0$  small enough but close to 1,  $\tau > 1$  and close to 1,  $\varepsilon$  small enough, set  $t_{max}$ ,
2. Set the initial values of the first moments  $\langle \underline{s} \rangle_0$  with non-zero values chosen randomly.
3. Choose the same  $\underline{W}$  used in the exercise above.

#### B: Optimization:

for each iteration  $t = 0, \dots, t_{max}$

1. repeat the following until convergence, i.e.,  $|e_k^{\text{new}} - e_k^{\text{old}}| < \varepsilon$ ,  $k = 1, \dots, N$ 
  - for  $k = 1, \dots, N$ :
    - a) compute mean-field

$$e_k = - \sum_{i \in \mathcal{N}_k} w_{ik} s_i = - \sum_{\substack{i=1 \\ i \neq k}} w_{ik} \langle s_i \rangle_Q$$

where  $\mathcal{N}_k$  is the set of neighbors of node  $k$ . In the case of all-to-all connections it is effectively the set of **all other** nodes.

- b) update the first moment using  $\langle s_k \rangle_Q = \tanh(-\beta e_k)$
2. Increase  $\beta$  using  $\beta_{t+1} = \tau \beta_t$ .

**C: Visualization:**

1. Plot the temperature  $T_t = \frac{1}{\beta_t}$  and the energy  $E(\underline{s})$  over the iterations  $t = 0, \dots, t_{max}$ .

**Total 10 points.**