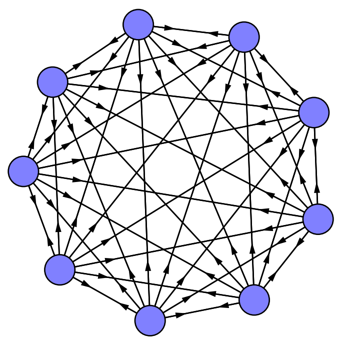
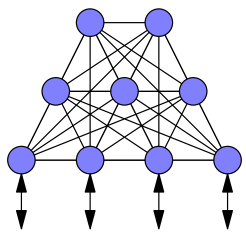
Boltzmann Machines

- Recurrent network
- Dynamical system

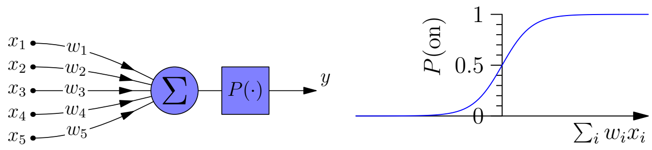


- Visible och Hidden Units
- Clamping



Boltzmann Machine

- Stochastic neural network
- Ideas from statistical mechanics



$$P(\text{on}) = \frac{1}{1 + e^{-\frac{1}{T}\sum}}$$

The parameter T controls the level of randomness

Network activity Consensus (samstämmighet)

$$C = \sum_{i,j} x_i x_j w_{ji}$$

The net tends to select patterns with high consensus

- $w_{ji} > 0 \rightarrow$ high consensus when x_i and x_j are similar
- $w_{ji} < 0 \rightarrow$ high consensus when x_i and x_j are different

Comparison with statistical mechanics

- $-C$ corresponds to free energy
- T corresponds to temperature
- Probability of being in a specific state depends on the states energy (consensus) and T

  Gibbs distribution: $P(x) = \frac{e^{\frac{E_x}{T}}}{Z}$
- Low energy $\rightarrow$ high probability
- Lowering $T \rightarrow$ concentration to low-energy states
- Slow reduction of $T \rightarrow$ concentration at the state with lowest possible energy

Simulated Annealing — Simulated cooling
Gradual reduction of the temperature

If the reduction is slow enough we have a high probability of ending up in the state with maximal consensus

Useful for optimization with constraints

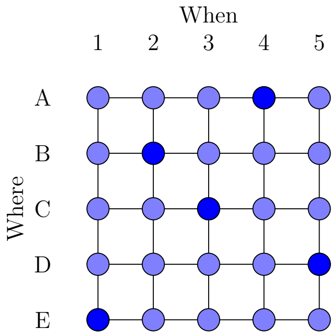
Optimization with Constraints

1. Choose a suitable representation
2. Formulate the constraints and goal function as weights
3. Use simulated annealing to find an optimal solution

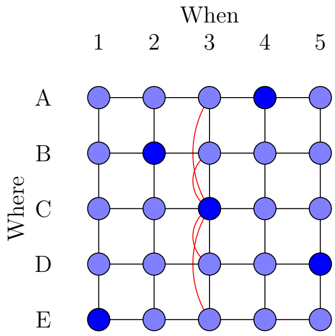
Example: Travelling Sales-Person (TSP)

Find the shortest path which passes all cities

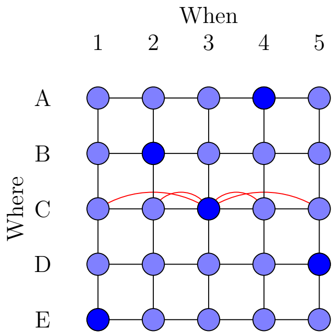
Redundant representation



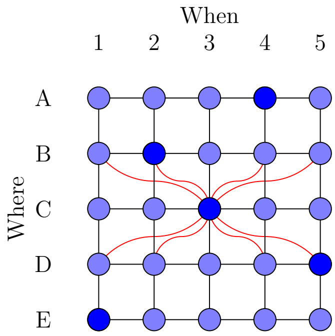
Negative connections to prevent being in several places simultaneously



Negative connections to prevent visiting the same place twice



Weaker negative weights to punish long routes



Two kinds of constraints

- Hard constraints (conditions)
  Strong weights
- Soft constraints (wishes)
  Weaker weights

Another example — Sudoku

- Optimization problem?
- Only hard constraints
- Representation:
  One unit per possible digit and place ($9 \times 9 \times 9$)
- Negative connections between different digits on the same place
- Negative connections between the same digit within the same field

Learning for Boltzmann Machines

Learning Principle

Adjust the weights so that the internally produced activity distribution resembles the external one

1. Measure pairwise correlation $\rho_{ij}^{+} = \langle x_i, x_j \rangle$ with input clamped
2. Measure pairwise correlation $\rho_{ij}^{-} = \langle x_i, x_j \rangle$ without input
3. Update weights
$$\Delta w_{ij} = \eta(\rho_{ij}^{+} - \rho_{ij}^{-})$$

Simplification of the model

- Mean-Field approximation
- All stochastic signals are replaced by their mean value
- Graded signals
- Fast convergence

Disadvantage:
Can not capture higher order correlations