Randomized Algorithms

Lecture 18: Ising Model

Raul Garcia-Patron

School of Informatics
University of Edinburgh

Ising Model

A **spin system** is a probability $p(\sigma)$ of configurations $\sigma \in \{+1, -1\}^V$, defined in a graph $G = (V, E)$.

- Interpretation as magnets: $\{+1, -1\}$ being the orientation.
- The nearest-neighbor **Ising model** is the most studied spin system. The energy of a configuration σ is defined to be:

$$H(\sigma) = \sum_{v,w \in V:(v,w) \in E} z_{v,w}\sigma(v)\sigma(w) + \sum_{z \in V} h_z \sigma(z) \tag{1}$$

- Minimization of the energy maps to NP-hard problems:
  - Ferromagnetic
  - Anti-ferromagnetic (MAX-CUT)
  - Independent sets

Ferromagnetic Ising Model

- Ferromagnetic $z_{v,w} < 0$

$$H(\sigma) = -\sum_{v,w \in V:(v,w)\in E} |z_{v,w}|\sigma(v)\sigma(w) + \sum_{z\in V} h_z\sigma(z) \qquad (2)$$

  - $z_{v,w} = -1, h_z = 0$: solution all +1/-1.
  - $z_{v,w} = -1, h_z = h$: either all +1 or -1 depending on sign(h).

Anti-Ferromagnetic Ising Model

- Anti-ferromagnetic $z_{v,w} > 0$

$$H(\sigma) = \sum_{v,w \in V:(v,w) \in E} z_{v,w}\sigma(v)\sigma(w) + \sum_{z \in V} h_\sigma(z) \tag{3}$$

- $z_{v,w} = 1, h = 0$: equivalent to MAX-CUT
  $H(\sigma) = \sum_{v,w \in V:(v,w) \in E} \sigma(v)\sigma(w)$
  - $H(\sigma) = |\text{No Cut}| - |\text{Cut}|$
  - $|\text{No Cut}| + |\text{Cut}| = |E|$
  - $H(\sigma) = |E| - 2|\text{Cut}|$
  - Minimizing $H(\sigma)$ equivalent of maximizing cut

Ising Model and independent sets

- Mapping IS to Ising model:

$$H(\sigma) = \sum_{v,w \in V:(v,w)\in E} (1-\sigma(u))(1-\sigma(v)) \tag{4}$$

  - The state of minimum energy (groundstate in physics) encodes the solution of maximum independent set.
  - The Gibbs distribution can contain configurations that are not IS !!

Gibbs distribution of Ising Model

- The energy of a configuration σ is defined to be:

$$H(\sigma) = \sum_{v,w \in V:(v,w)\in E} z_{v,w}\sigma(v)\sigma(w) + \sum_{z\in V} h_z \sigma(z) \qquad (5)$$

- Its Gibbs distribution that we will sample from:

$$\mu(\sigma) = \frac{1}{Z(\beta)} e^{-\beta H(\sigma)}$$

- It is the state of equilibrium of the system at temperature $1/\beta$
- Where $Z(\beta)$ is the **partition function** (normalization factor).

$$Z(\beta) = \sum_{\sigma\in\Omega} e^{-\beta H(\sigma)}$$

  - Variables of the system, such as the total energy, free energy, entropy, and pressure, can be expressed in terms of the partition function or its derivatives.
  - Hard to compute exactly, sometime easy to approximate with relative error.

Applications

- **Physics**: Physicist are often interested in graphs that are a lattice of a given dimension.
- **Theoretical Computer Science**: complexity phase-transition, from easy to hard problems.
  - Computational phase-transition connected to the physical phase-transition
- **Machine Learning:**
  Boltzman machines, Energy based models,...
  1. We have data that we assume comes from distribution $p(\sigma)$
  2. We learn the parameters from the data (can be hard)
  3. We generate samples form $p(\sigma)$ (can also be hard)

Simulated annealing

It is a randomized algorithm for approximating the global optimum in a large search space for an optimization problem characterized by a cost function $H(\sigma)$.

- The name of the algorithm comes from annealing in metallurgy.

Key idea:

1. Design a proper annealing (cooling) schedule (increasing β).
2. Use a sampling algorithm via Metropolis to generate samples of its Gibbs distribution, $min\{1, e^{-\beta(H(y)-H(x))}\}$.

All the art is on designing a good β schedule and how you generate samples quickly. If your Markov chain does not mix quickly you will be stuck in a local minima as you will reject all transitions. You can not circunvent NP-hardness.

Approximating partition function

We can compute the partition function

$$Z(\beta) = \sum_{\sigma \in \Omega} e^{-\beta H(\sigma)}$$

using a *telescoping product* and constructing a cooling schedule of m steps increasing β_i, where $\beta_0 = 0$ and $\beta_m = \beta$:

$$|Z(\beta)| = \frac{|Z(\beta_m)|}{|Z(\beta_{m-1})|} \times \frac{|Z(\beta_{m-1})|}{|Z(\beta_{m-2})|} \times \frac{|Z(\beta_{m-2})|}{|Z(\beta_{m-3})|} \times \ldots \times \frac{|Z(\beta_1)|}{|Z(0)|} \times |Z(0)|.$$

- $|Z(0)| = 2^n$ as it is the sum of all configuration.
- We obtain close approximate values $\tilde{r}_i$ for each ratio $r_i = \frac{|Z(\beta_i)|}{|Z(\beta_{i-1})|}$, for $i = 1, \ldots, m$.
- Our *estimate* for the partition function $Z(\beta)$ will be:

$$2^n \prod_{i=1}^{m} \tilde{r}_i.$$

Gibbs Sampling for the Ferromagnetic Ising model

Probability $p(\sigma)$ of configurations $\sigma \in \{+1, -1\}^V$, defined in a graph $G = (V, E)$.

- The energy of a configuration σ is defined to be:

$$H(\sigma) = - \sum_{v,w \in V : (v,w) \in E} \sigma(v)\sigma(w) \tag{6}$$

- Minimization if all +1/-1. Easy solution.
- We are interested in sampling from the Gibbs distribution

$$\mu(\sigma) = \frac{1}{Z(\beta)} e^{-\beta H(\sigma)}$$

- This could have many application, among others approximating the partition function

$$Z(\sigma) = \sum_{\sigma \in \Omega} e^{-\beta H(\sigma)}$$

Glauber Dynamics recap

- Let V and S be finite sets and suppose $\Omega \subseteq \{+1, -1\}^V$.
  Ex: V vertices of a graph with a spin on each.
- Let π be a probability distribution whose support is Ω.
- The Glauber chain moves from state x as follows:
  1. An element v is chosen uniformly at random from V.
  2. A new state y is chosen s.t.: $y(w) = x(w) \,\forall\, w \neq v$.

Definition

Given $x \in \Omega, v \in V : \Omega(x, v) = \{y \in \Omega : y(w) = x(w) \,\forall\, w \neq v\}$. The chain transition reads:

$$P_{x,y} = \pi(y|\Omega(x, v)) = \begin{cases} \frac{\pi(y)}{\pi(\Omega(x,v))} & \text{if } y \in \Omega(x, v) \\ 0 & \text{if } y \notin \Omega(x, v) \end{cases} \tag{7}$$

Glauber for Ferromagnetic Ising

Consider Glauber dynamics to sample from the Gibbs distribution of the Ferromagnetic Ising model of a graph $G = (V, E)$.

- Let $\pi(x)$ the Gibbs distribution $\mu(\sigma) = \frac{1}{Z(\beta)} e^{-\beta H(\sigma)}$ where $H(\sigma) = -\sum_{v,w \in V:(v,w)\in E} \sigma(v)\sigma(w)$ and $Z(\sigma) = \sum_{\sigma \in \Omega} e^{-\beta H(\sigma)}$.

Algorithm GLAUBERFERRO$(G = (V, E))$

1. Start from a random σ.
2. **for** $i \leftarrow 0$ **to** "whenever"
3. Choose v uniformly at random from V.
4. Set $\sigma_{t+1}(x) = \sigma_t(x) \forall x \neq v$.
5. $\sigma_{t+1}(v) = +1$ with probability $(1 + \tanh(\beta) S(\sigma, v))/2$
6. $\sigma_{t+1}(v) = -1$ with probability $(1 - \tanh(\beta) S(\sigma, v))/2$

where $S(\sigma, v) = \sum_{u:u\sim v} \sigma(u)$ that depends only on the vertices adjacent to w.

Glauber for Ferromagnetic Ising - Transition rule

- Following the Glauber update rules there is only two possibilities, either $\sigma_{t+1}(v) = +1$ or $\sigma_{t+1}(v) = -1$, let's call their corresponding configurations σ_{t+1}^{+1} and σ_{t+1}^{-1} respectively.
- The Glauber transition probability should read:

$$P_{\sigma_t, \sigma_{t+1}^{\pm 1}} = \frac{\mu(\sigma_{t+1}^{\pm 1})}{\mu(\sigma_{t+1}^{+1}) + \mu(\sigma_{t+1}^{-1})} \text{ with } \mu(\sigma) = \frac{1}{Z(\beta)} e^{-\beta H(\sigma)}. \quad (8)$$

- Because both $\sigma_{t+1}^{\pm 1}$ only differ on vertex v we can write:

$$H(\sigma_{t+1}) = H_{V \setminus v} - \sigma_{t+1}(v) \sum_{x \in V : x \sim v} \sigma_t(u) = H_{V \setminus v} - \sigma_{t+1}(v) S(\sigma_{t+1}, v). \quad (9)$$

where $H_{V \setminus v}$ is the energy independent of vertex v.

$$\begin{aligned} P_{\sigma_t, \sigma_{t+1}^{\pm 1}} &= \frac{e^{-\beta H_{V \setminus v}} e^{\beta \sigma(v) S(\sigma, v)}}{e^{-\beta H_{V \setminus v}} e^{\beta S(\sigma, v)} + e^{-\beta H_{V \setminus v}} e^{-\beta S(\sigma, v)}} \\ &= \frac{e^{\beta \sigma(v) S(\sigma, v)}}{e^{\beta S(\sigma, v)} + e^{-\beta S(\sigma, v)}} = (1 + \sigma(v) \tanh(\beta S(\sigma, v)))/2. \end{aligned}$$

Path Coupling in a nutshell

- We define a distance $d(X, Y)$. Neighbors if $d(X, Y) = 1$.
- Our goal is to prove concentration of expect. of distance: $\mathbf{E}[d_{t+1}] =\leq \beta\mathbf{E}[d_t]$, with $\beta < 1$.
- Path: $X_t = Z_0, Z_1, ..., Z_{d_t} = Y_t$ where $d(Z_{i+1}, Z_i) = 1$
  - $d_t = \sum_{i=1}^{d_t} d(Z_{i+1}, Z_i)$ (by construction)
- Updated path: $X_{t+1} = Z'_0, Z'_1, ..., Z'_{d_t} = Y_{t+1}$.
  - $d_{t+1} \leq \sum_{i=1}^{d_t} d(Z'_{i+1}, Z'_i)$ (by triangle inequality)

1. Prove $\mathbf{E}[d(Z'_{i+1}, Z'_i)] \leq \beta d(Z_{i+1}, Z_i) = \beta$.
2. Leads to $\mathbf{E}[d_{t+1}|d_t] \leq \sum_{i=1}^{d_t} \mathbf{E}[d(Z'_{i+1}, Z'_i)] \leq \beta d_t$.
3. Then $\mathbf{E}[d_{t+1}] = \mathbf{E}[\mathbf{E}[d_{t+1}|d_t]] =\leq \beta\mathbf{E}[d_t]$.

Left to do: prove $\mathbf{E}[d(Z'_{i+1}, Z'_i)] \leq \beta$ for our problem of interest.

Path Coupling in a nutshell - Analysis

We will now prove $\mathbf{E}[d_{t+1}|d_t = 1] \leq 1 - \frac{c(\beta,\Delta)}{n}$.

- We define the distance $\rho(\sigma,\tau) = \frac{1}{2}\sum_{u\in V}|\sigma(u) - \tau(u)|$
- We have $\sigma_t(x) = \tau_t(x) \quad \forall u \neq x$ as $\rho(\sigma,\tau) = 1$.
- We do not care about how X_t or Y_t change but on when they have different updates that lead to $d_{t+1} \neq d_t$.
- The discussion depends on v (Glauber update spin) vs x (spin of difference) and its neighbors:
  1. Case I: (when $v = x$) $\mathbf{E}[d_{t+1} - d_t|d_t = 1] = -1/n$.
  2. Case II ($v \notin N(x) \cup \{x\}$: $\mathbf{E}[d_{t+1} - d_t|d_t = 1] = 0$.
  3. Case III ($v \in N(x)$): $\mathbf{E}[d_{t+1} - d_t|d_t = 1] \leq \frac{\Delta}{n}\tanh\beta$.

We obtain:

$$\mathbf{E}[d_{t+1}|d_t = 1] \leq 1 - \frac{1}{n} + \frac{\Delta}{n}\tanh\beta = 1 - \frac{1 - \Delta\tanh(\beta)}{n} \qquad (10)$$

If $\Delta\tanh\beta < 1$ we obtain: $\mathbf{E}[d_{t+1}|d_t = 1] \leq 1 - \frac{c(\beta,\Delta)}{n} \leq e^{-c(\beta,\Delta)/n}$
with $c(\beta,\Delta) = 1 - \Delta\tanh\beta > 0$.
Finally: $\tau(\epsilon) = \frac{n}{c(\beta)}(\log n + \log(1/\epsilon))$.