PHYS-467 Assignment 2
Censored spin-glass

November 18, 2022

Instructions

You are asked to submit a file `code_assignment_2_your_name.py` (for example `code_assignment_2_lucas_clarte.py`), where the different functions are duly implemented. For the questions 3, 5 and 6, you are asked to submit a pdf file- `answers_assignment_2_your_name.pdf` containing your answers and the different plots.

In this exercise, we consider a variation of the card game seen in class. We consider a vector $\vec{s} \in \{-1, +1\}^n$ where the elements s_i are i.i.d sampled uniformly on $\{-1, +1\}$. We now consider a graph $G = (V, E)$ where the nodes V are $V = \{1, ..., n\}$ and for $1 \leqslant i < j \leqslant n$, the edge (i, j) is in E with probability $p_e > 0$.

We now define the observations matrix matrix Y such that $Y_{i,i} = 0$ and for $i \neq j$, $Y_{i,j} = 0$ if $(i, j) \notin E$. For $(i, j) \in E$, we take

$$\begin{cases} Y_{i,j} & = s_i s_j \text{ with probability } 1 - \rho \\ Y_{i,j} & = -s_i s_j \text{ with probability } \rho \end{cases} \tag{1}$$

For some $1 \geqslant \rho \geqslant 0$. The edges E will be stored as a $n \times n$ matrix such that $E_{ij} = 1$ if and only if the edge (i, j) is present in the graph. Note that for all i, $(i, i) \notin E$. The goal of this exercise is to compare the performance of the Bayes-optimal estimator and the estimator returned by PCA.

Question 1 (1 pt) Implement the function `generate_data` that takes as argument the parameters n, ρ and p_e and returns an instance $\vec{s}, Y, E$ of the problem.

Question 2 (1 pt) Implement the function `run_pca` that returns the eigenvector $\vec{s}^{PCA}$ of Y that corresponds to its highest-eigenvalue.

Question 3 (2 pt) Explicit the posterior distribution

$$P(\vec{s}|Y, \rho) \propto P(\vec{s})P(Y|\vec{s}, \rho) \tag{2}$$

and show that it can be written

$$P(\vec{s}|Y, \rho) \propto \prod_i \mathbf{1}(s_i = \pm 1)\, (1 - \rho)^{\sum_{i<j} \mathbf{1}(y_{i,j} = s_i s_j)}\, \rho^{\sum_{i<j} \mathbf{1}(y_{i,j} = -s_i s_j)} \tag{3}$$

Question 4 (4 pt) We consider the following Metropolis-Hastings scheme that runs for T iterations:

1. At $t = 1$, sample $\vec{s}^1$ from $P(\vec{s})$.
2. At each iteration t, pick an index i uniformly in $[1, n]$
3. Define $\tilde{\vec{s}} = (s_1^t, ..., s_{i-1}^t, -s_i^t, s_{i+1}^t, s_n)$ and compute the ratio $\eta = \frac{P(\tilde{\vec{s}}|Y)}{P(\vec{s}^t|Y)}$. With probability $\min(1, \eta)$, define $\vec{s}^{t+1} = \tilde{\vec{s}}$, otherwise define $\vec{s}^{t+1} = \vec{s}^t$

Implement the function `log_ratio_posterior` that returns $\log\left(\frac{P(\tilde{\vec{s}}|Y)}{P(\vec{s}|Y)}\right)$ given $\vec{s}$ and with $\tilde{\vec{s}}$ defined as above. Use it to implement the function `run_mcmc` that runs Metropolis-Hastings for T iterations and returns a list $(\vec{s}^t)_{t=1}^T$.

Question 5 (3 pt) Once we have the list of iterations $(\vec{s}^t)_{t=1}^T$ we can build an approximation $\vec{s}^{BO}$ of the Bayes-optimal estimator. Recall that the Bayes-optimal estimator that maximizes the overlap

$$Q(\hat{\vec{s}}) = |\frac{1}{n}\vec{s} \cdot \hat{\vec{s}}| \tag{4}$$

is equal to

$$\vec{s}_i^{BO} = \arg\max_{s=\pm 1} P(s_i = s|Y, \rho) \tag{5}$$

Take $n \geqslant 500, p_e = 4/n$. For $\rho \in [0.1, 0.4]$, run Metropolis-Hastings and plot $Q(\vec{s}^t)$ as a function of time. Use these plots to estimate the thermalization time T_{therm} of Metropolis-Hastings as a function of ρ. Additionally, plot $Q(\vec{s}^{BO})$ as a function of $\rho \in [0.1, 0.4]$.

Relate the behaviour of $Q(\vec{s}^{BO})$ with the behaviour of T_{therm}. In particular, do you notice a phase transition for $Q(\vec{s}^{BO})$ at some ρ ?

Question 6 (3 pt) Take $n \geqslant 500, \rho = 0.25$. For $p_e \in [2/n, 6/n]$, compute and plot the overlaps $Q(\vec{s}^{BO})$ and $Q(\vec{s}^{PCA})$ as a function of p_e. Compare the performance of the two estimators.