Machine Learning for Physicists - Second Assignment

Guilherme Simplício

December 2, 2022

1 Theory Part & Plots

Question 3 Explicit the posterior distribution

$$P(\vec{s} \mid Y, \rho) \propto P(\vec{s})P(Y \mid \vec{s}, \rho)$$

and show that it can be written

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

Using the Bayes Rule:

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

As per usual P(Y|X) is the likelihood, P(X) is the prior and P(Y) is the evidence, or in other words, simply a "normalization" constant, hence all the constants can be inserted into this term. Knowing this, one can safely say that:

$$P(\vec{s} \mid Y, \rho) \propto P(\vec{s})P(Y \mid \vec{s}, \rho)$$

The prior (P($\vec{s}$)) is simply the Rademacher probability:

$$P(\vec{s}) = \frac{1}{2^N} \prod_{i=1}^{n} (\delta_{s_i,1} + \delta_{s_i,-1}) = \frac{1}{2} \prod_{i=1}^{n} (\delta_{s_i,1} + \delta_{s_i,-1}) \propto \prod_{i=1}^{n} \mathbf{1}(s_i = \pm 1)$$

The likelihood($P(Y \mid \vec{s}, \rho)$) however, includes a probability ρ of the spin being "flipped", so it can be written as:

$$P(Y \mid \vec{s}, \rho) = \prod_{i<j}^{n} (1-\rho)^{\mathbf{1}(y_{i,j}=s_i s_j)} \rho^{\mathbf{1}(y_{i,j}=-s_i s_j)}$$

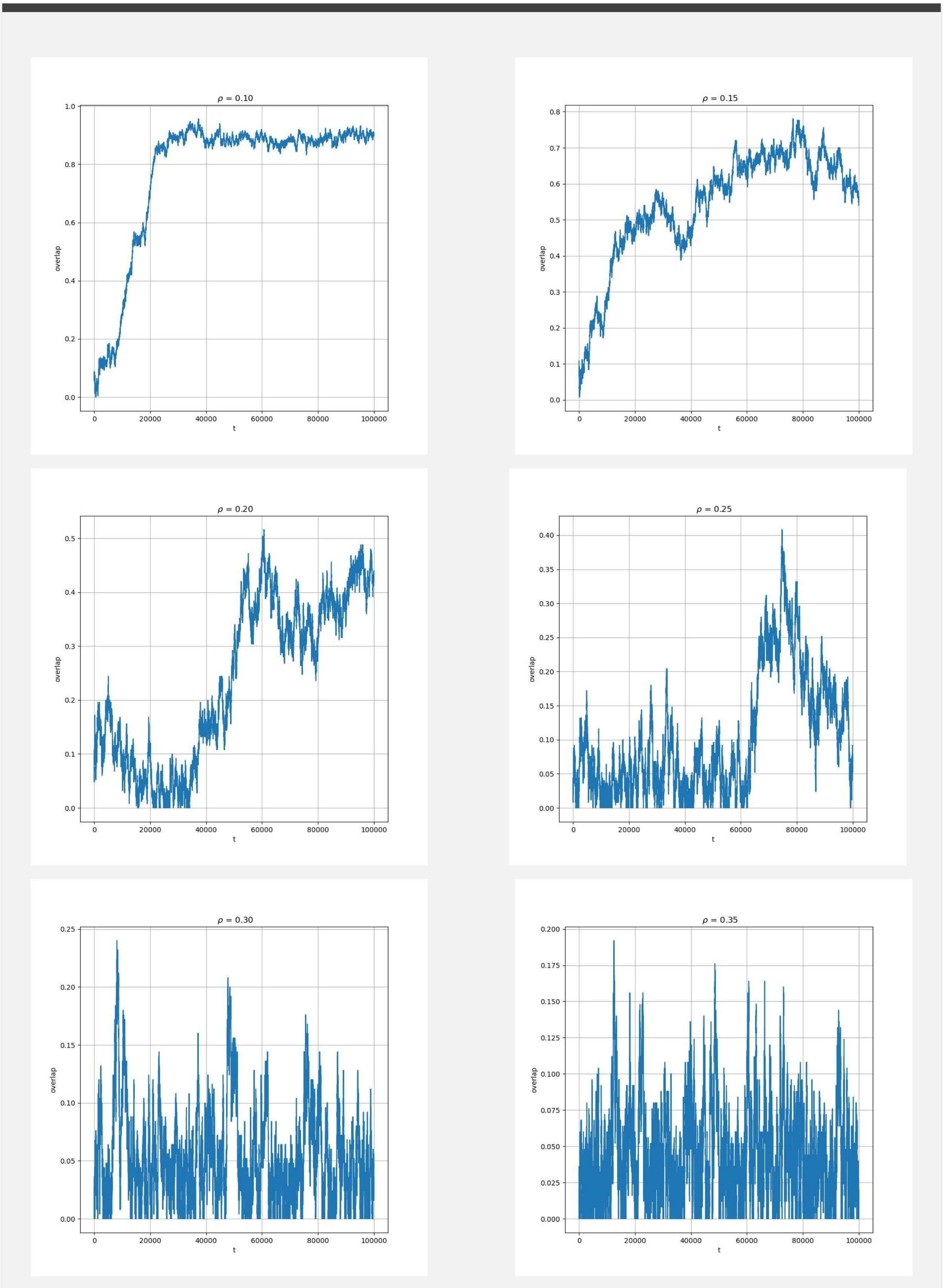
Hence,

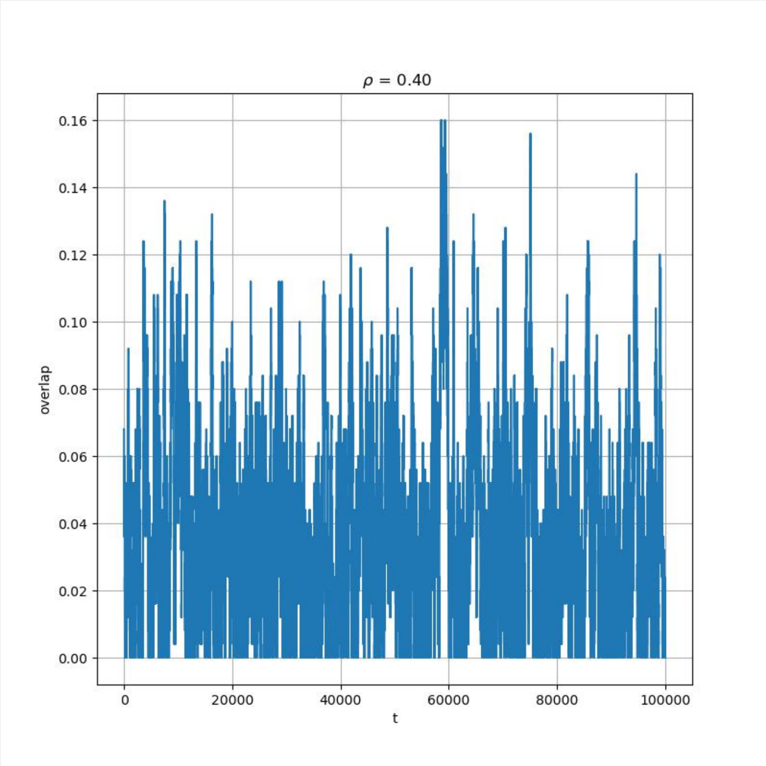
$$P(\vec{s} \mid Y, \rho) \propto \prod_{i=1}^{n} \mathbf{1}(s_i = \pm 1) \prod_{i<j}^{n} (1-\rho)^{\mathbf{1}(y_{i,j}=s_i s_j)} \rho^{\mathbf{1}(y_{i,j}=-s_i s_j)} =$$

$$= \prod_i \mathbf{1}(s_i = \pm 1)(1-\rho)^{\sum_{i<j} \mathbf{1}\left(y_{i,j}=s_i s_j\right)} \rho^{\sum_{i<j} \mathbf{1}\left(y_{i,j}=-s_i s_j\right)}$$

As it was intended to show!

Question 5 Use these plots to estimate the thermalization time T_{therm} of Metropolis-Hastings as a function of ρ.





Considering that the values of ρ = [0.10,0.15,0.20,0.25], we have T_{therm} =[20000,40000,60000,80000] and that for the other values of ρ there is no thermalization, one can just do a linear fit and reach to the conclusion that T_{term} = 400000 ρ - 20000.

Additionally, plot $Q(\vec{s}^{BO})$ as a function of $\rho \in [0.1, 0.4]$.

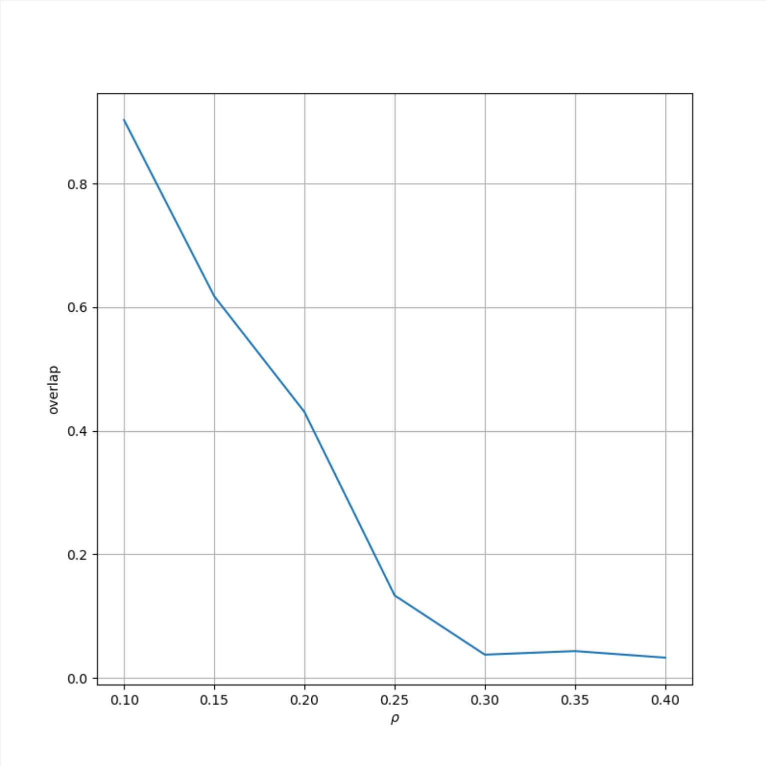


Figure 2: Overlaps vs ρ

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 ?

The phase transition seems to happen for ρ=0.30. It is noticeable in the plots of $Q(\vec{s}^{BO}))$ that for the first few values of ρ, when the noise is little, that the BO is able to recover the initial vector s with good accuracy. However, when the noise starts to get higher, at around ρ = 0.30, the function starts to be unable to recover the vector s, hence the phase transition.

Question 6 Take n $\geq$ 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 .

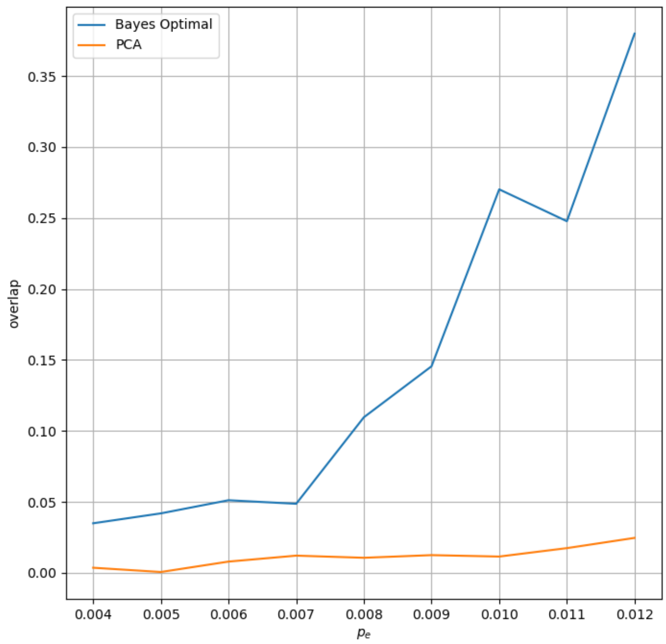
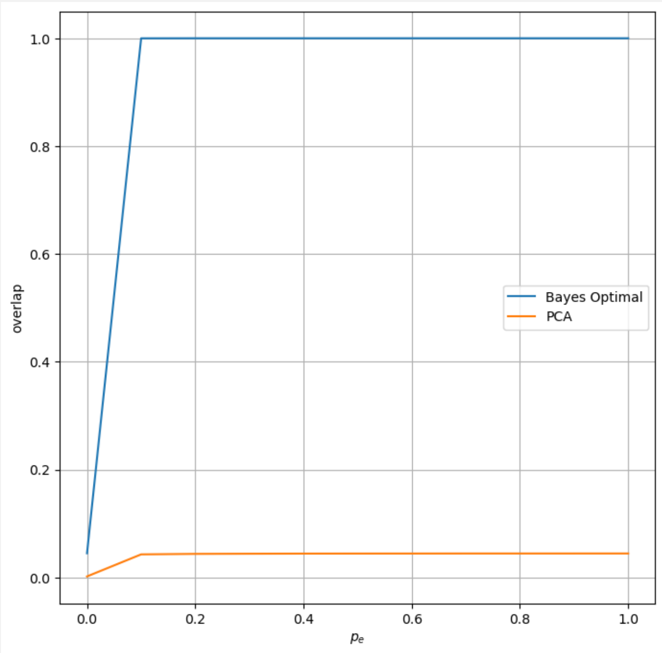


Figure 3: $Q(\vec{s}^{BO})$(Bayes Optimal) and $Q(\vec{s}^{PCA})$) as a function of p_e

Compare the performance of the two estimators.

In short, the Principal Component Analysis (PCA) is a really quick method, computationally fairly cheap, that consists in increasing the interpretability of data while preserving the maximum amount of information, and enabling the visualization of multidimensional data. The Bayes-optimal estimator is computationally much more demanding. However, it is clear that the overlaps are way higher for this method, thus giving us better results than the PCA! The plot for a full range of p_e is:



Interestingly enough, it seems that both methods are maximized for the same value p_e (≈ 0.1)

2 Code explanation

In **exercise 4**, to simplify the coding of the function *log_ratio_posterior* we did quite a bit of Algebra. First of all we used the property that states:

$$log\left(\frac{A}{B}\right) = log(A) - log(B) \tag{2.1}$$

Also, trivially,

$$P(\tilde{\vec{s}} \mid Y, \rho) = a' \cdot log(1-\rho) + b' \cdot log(\rho) \tag{2.2}$$

and

$$P(\vec{s} \mid Y, \rho) = a \cdot log(1-\rho) + b \cdot log(\rho) \tag{2.3}$$

Hence,

$$log(\eta) = log\left(\frac{P(\tilde{\vec{s}} \mid Y, \rho)}{P(\vec{s} \mid Y, \rho)}\right) = (a' - a) \cdot log(1-\rho) + (b' - b) \cdot log(\rho) \tag{2.4}$$

Finally noticing that $(a' - a) + (b' - b) = 0$ one can write 2.4 as:

$$log(\eta) = (b' - b) \cdot [-log(1 - \rho) + log(\rho)] \tag{2.5}$$

It is also important to explain why b'-b, can be done simply by a dot product. b'-b is the number of flips - number of edges. Knowing the row i where the flip happens, one can simply check the differences between the ith row of Y and s_is. Being even more accurate, one doesn't even need to check the full row, as the matrix Y is symmetric, one can first "break" Y into an upper triangular matrix and then checking only the row i.

3 Acknowledgments

I would like to thank Albert Riber, Arianna Alonso Bizzi,Francisco Simões and Tomás Feith(not in the course), for having very useful debates with me regarding this assignment and the course, in general.