Approximate Inference on the Ising model

Brahim Khalil Abid, Ilyes Khemakhem, Rémi Le Priol

December 2016

1   Description of the project

We consider a graph model $G = (V, E)$, and a probability distribution from the exponential family factorising on that graph, with density $p(x) = \exp(\theta \cdot \phi(x) - A(\theta))$. In this configuration the canonical parameters θ and the sufficient statistics ϕ are known.

We would like to sample according to this distribution, infer the marginal probabilities, or the mean parameters $\mu = E_p[\phi(x)]$. This problem is easy on trees, but hard in the general case. There is an exact inference algorithm that works for all undirected graphs : the junction tree algorithm. But its computational cost grows exponentially with the size of the maximal clique in the graph.

In most cases we can only hope for approximations of these mean parameters. We will describe and implement three methods for that purpose :

— **Gibbs sampling :** Starting from a given value of the variables, we iteratively sample each node conditionally to its neighbours. This algorithm converges to sample for the mean parameters of the graph (for dense graphs).

— **Loopy belief, or sum-product algorithm :** This is a variation method on graphs with cycles. To understand how, we introduce the Bethe Variational Problem : we work on pseudomarginals instead of marginals which are positive vectors that satisfy the local marginalization constraints, and we solve an approximate variational problem on $A(\theta)$ by using the expression of $A*(\theta)$ obtained on tree-structured graphs on graphs with cycles. The sum-product algorithms provides the solution for this problem's Lagrangian. However, unlike on trees, the algorithm is not guaranteed to converge on graphs with cycles.

— **Mean Field algorithm :** This is again a variational method. We are looking for a lower bound on $A(\theta)$. We simplify our problem by removing edges in the graph, ie adding independence statement, ie removing some sufficient statistics and some parameters. Once we have a tractable graph, either a forest, either the empty graph), we look for the distribution that is the closer from the original one, in terms of KL divergence. This is a simple non-convex optimization problem, which yields an iterative procedure.

2   Proposed plan

— Derivation of all 3 algorithms on the Ising model
— Implementation on a grid-graph.
— Comparison of the 3 methods in terms of speed, complexity, and sensitivity to variables' correlation.
— Maybe an application to image segmentation or denoising.