CS 365, Lecture 8
Foundations of Data Science
Boston University

Charalampos E. Tsourakakis

February 14th, 2023

Bayesian inference in the real-world

Reminder: Bayes' theorem

Bayes' theorem (aka Bayes' Law and Bayes' rule) is a direct application of conditional probabilities.

$$\mathbf{Pr}[H|D] = \frac{\mathbf{Pr}[D|H]\mathbf{Pr}[H]}{\mathbf{Pr}[D]}, \text{and } \mathbf{Pr}[D] > 0, \text{or } \ldots$$

posterior $\propto$ likelihood $\times$ prior.

Exact MAP Estimation for Binary Images

Problem: Given (b) can we infer (a)? In other words, can we **restore** the image from its corrupted-by-noise version?

GREIG, PORTEOUS AND SEHEULT

(a) (b)

How to formulate the problem? Any ideas?

Exact MAP Estimation for Binary Images

Let's be Bayesian!

$x = (x_1, \ldots, x_n)$ the original image (shown in (a))

$y = (y_1, \ldots, y_n)$ the observed corrupted image (e.g., the one shown in (b))

Assumption: The records $y_1, \ldots, y_n$ are conditionally independent given x, and each has known conditional density $f(y_i|x_i)$ that depends only on x_i.

By Bayes' theorem:

$$p(x|y) \propto \underbrace{p(y|x)}_{\text{likelihood:how do we compute it?}} \times \underbrace{p(x)}_{\text{prior:what is a good prior?}}$$

Goal: output

$$\boxed{x^* = \arg\max p(x|y)}$$

Likelihood and Prior

Given our assumption, the likelihood function is

$$p(y|x) = \prod_{i=1}^{n} f(y_i|1)^{x_i} f(y_i|0)^{1-x_i}.$$

- What kind of patterns would we like the prior to enforce?

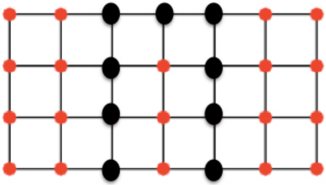
- Let's imagine how these characters would look on a binary image:

  a,b,c,f,y,x,z,1,&,$,@

  Homogeneous patches that occasionally change discontinuously

Prior $p(x)$

$$p(x) \propto \exp\left\{\frac{1}{2}\sum_{i\neq j}\beta_{ij}\big(x_i x_j + (1-x_i)(1-x_j)\big)\right\}$$



- For an edge (u, v) where $val(u) = val(v)$

$$x_u x_v + (1-x_v)(1-x_u) = x_u^2 + (1-x_u)^2 = 1.$$

- On the contrary for an edge where $val(u) \neq val(v)$

$$x_u(1-x_u) + (1-x_u)x_u = 0.$$

Exact MAP Estimation for Binary Images

Our MAP inference becomes equivalent to minimizing (details on whiteboard)

$$\sum_{i=1}^{n} x_i \max(0, -\lambda_i) + \sum_{i=1}^{n} \max(0, \lambda_i)(1 - x_i) + \frac{1}{2} \sum_{i \sim j} \beta_{ij}(x_i - x_j)^2,$$

where $\lambda_i = \frac{f(y_i|1)}{f(y_i|0)}$.

Let's rephrase this problem. Suppose $b_{ij} = b$ for all neighboring nodes for simplicity.

Exact MAP Estimation for Binary Images

- We have a $n \times m$ binary matrix
- We impose a grid structure
- We call two neighboring nodes bad if they have different values. We pay K units for each such pair.
- We are allowed to flip the value of any node, but we have to pay R units.
- The total cost is the sum of these two terms. How do we find the best assignment of values to nodes?

Any ideas? Is it NP-hard, poly-time solvable?

Exact MAP Estimation for Binary Images

Max flow problem!

- Source s, sink t
- Arc of capacity R from s to each node u with value 0.
- Arc of capacity R from each u node with value 1 to sink t.
- Directed arcs from each node u to its neighbors with capacity K.

Details on whiteboard.

Assigned reading: Exact maximum a posteriori estimation for binary images Greig-Porteous-Seheult paper.