Bayesian graphical models summary

Luca Bracone

September 26, 2022

1 Introduction

We are given a set of points $y_1, \dots, y_n$ and imagine they are realisations of a random distribution $p(y)$. We would like to know how $p(y)$ looks like. A commonly used method is the following: suppose that there is a set of possible random distribution functions $P(y)$ in which $p(y)$ exists. Then we assume that there exists a set Θ and a function $\Theta \to P(y)$ which we call a *parametrisation*. Usually, Θ is a simpler set to study than $P(y)$. Then, statistics is concerned with using the observed $y_1, \dots, y_n$ to draw a $\theta \in \Theta$ such that $\theta \mapsto p(y)$ is as "close" as possible to the "true" $p(y)$.

Unlike in the usual statistical context, we view the parameter θ as being itself random with some distribution $p(\theta)$. We define the joint probability of y and θ as being a function $p(y, \theta)$ such that

$$p(\theta) = \int p(y, \theta) dy$$
$$\text{and } p(y) = \int p(y, \theta) d\theta$$

Then we define the conditional distribution "of y given θ" as

$$p(y|\theta) = \frac{p(y, \theta)}{p(\theta)}.$$

Using this property twice we obtain Bayes' theorem which allows us to "invert" and obtain the distribution of θ given the observed y. Since we only observe a finite number of y this is only an approximate distribution, we will talk about uncertainty later on.

$$p(\theta|y) = \frac{p(y, \theta)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)}.$$

In practice, the factor $p(y)$ is nothing to worry about since we can obtain it by integrating away θ in the following way, $p(y) = \int p(y|\theta)p(\theta)d\theta$.

2 Graphical models

We observe a set of points $y_1, \ldots, y_n$. In particular the covariance for these points $\mathrm{cov}(y_i, y_j)$ for $i, j = 1, \ldots, n$ is unknown and not necessarily zero. We would like to obtain estimates for $p(y_i | y_1, \ldots, \cancel{y_i}, \ldots, y_n)$.