STCS 6702: Foundations of Graphical Models: Reading 2

Maxime TCHIBOZO (MT3390)

September 2020

1 Bishop (2013) – Model-Based Machine Learning

In this paper, Bishop establishes a distinction between the traditional method-based approach to Machine Learning (problem is re-interpreted so as to be solvable by an existing and general-purpose algorithm) and a model-based approach to Machine Learning (method is adapted to the problem, with each problem requiring a new model).

The author mentions that the traditional approach is not appropriate in problems which require modifications or even novel implementations of off-the-shelf algorithms. Bishop believes a Bayesian approach using probabilistic graphical models and deterministic approximate inference algorithms offers a model-based solution to these limitations.

The benefits of the model-based approach are that they offer flexibility and adaptability, can combine supervised and unsupervised modules, automatically adapt to changes in the data or model, are written with compact and readable code, and use a single easy to learn framework. Additionally, Bayesian methods work well with limited amounts of data and high uncertainty, whereas traditional methods overfit.

The Bayesian framework is founded upon the update of prior distributions for new observations using posterior distribution derived from past observations. This framework is well suited to online learning and can offer hierarchical results tailored to the granularity of the data (possibility of using either individual-level priors or population-level hyper-priors). For the Bayesian framework to be successful in solving today's Big Data problems, model-based inference methods must be efficient and scalable.

Probabilistic graphical models offer a pictorial representation of Bayesian models (i.e joint distributions over random observed and latent variables) while respecting conditional probability properties. We use a graph structure to translate our assumptions about the distributions at play in the Data Generating Process. We then imagine that the observed data is generated by sampling through the graph nodes.

Many standard machine learning or signal processing techniques can be expressed as special cases of graphical models (PCA, factor analysis, logistic regression, Gaussian mixtures). This viewpoint allows us to extend these methods by modifying their graphs to include additional assumptions: Bishop uses the example of Auto-Regressive HMM and factorial HMM as extensions of the HMM.

Factorizing the joint distribution of the model using the graphs usually yields a significant dimensionality reduction of the problem. A wise ordering of the factorized distribution can then make integration more efficient (computations scale linearly in size of graph in some cases).

Exact evaluations of marginal posterior distributions are sometimes infeasible, and we use efficient algorithms for approximation. Markov chain Monte Carlo techniques such as Gibbs sampling approximate a broad range of distributions with asymptotic convergence but are not scalable. In practice, we resort to methods such as Expectation propagation (Minka 2001), variational message passing (Winn and Bishop 2005) and fractional belief propagation (Wiegerinck and Heskes 2003).

Bishop gives the example of TrueSkill (Herbrich et al. 2007), a Bayesian model-based approach to Xbox Live player ranking.

2 Freedman (1994) – Some issues in the foundation of statistics

This paper highlights the difference between objectivist (frequentist) and subjectivist (Bayesian) statisticians and provides a stern warning to those seeking to apply statistical methods to real life problems.

Objectivists (Frequentists) deal with empirical percentage estimators and study the asymptotic properties of these estimators. To objectivists, intrinsic probabilities exist separately of observed data, and statisticians only use estimates or hypothesis tests to gain understanding of the properties of the probabilities. The challenge of this approach is that we deal with approximations of a probability event and our properties hold true only in convergence.

Subjectivists (Bayesians) deal with probability density functions, and assign prior distributions to objective parameters. For objectivists, we could know the true value of an estimator if we had an infinite amount of data. To subjectivists, parameters are distributions, and we can only gain confidence in the assumptions we make on the prior distribution of parameters. Bayesian frameworks imply subjectivity of the statistician, which implies that different statisticians are likely to disagree on their conclusions.

In fairness, this philosophical debate between both views is somewhat nonsensical for applied mathematicians, who rarely focus on identifying the true prior, and instead seek to make up mathematically convenient priors. Drawing similarities Bayesian viewpoint in Mathematics with Rationalism in Philosophy is misguided at best, and these comparisons usually do not hold upon further examination. Prior probabilities are meant to capture ether gross intuitions, or convenient assumptions. They are crude approximations, and in real-life problems, there is usually very little proof to support these assumptions. All inferences that are derived from these priors hold for the priors themselves, not the process a statistician was aiming to fundamentally understand.

The conceptualization of a prior is a factor of the Bayesian statistician's opinion, and does not hold any objective truth. This is not to say Bayesian statistics are useless, quite the opposite, they have provided brilliant advances in many disciplines. Like the rest of Applied Mathematics, they are tools which describe the logical consequences of certain subjective assumptions. For every model we create, we should not forget these assumptions, and the consistency of these assumptions should be checked after the experiment. For example, the use of linear regression in social sciences (usually in a causal inference framework) is not valid from a mathematical point of view unless the error assumptions and statistical significance of parameters are proven to hold. Unfortunately, in these fields, these assumptions can rarely be proven rigorously - such models can therefore not be validated, and should not be used as or compared to mathematical demonstrations.

This paper was a great read and really put my assumptions about Bayesian Inference in check. Freedman justly points out that model validation is underappreciated if not voluntarily omitted in most cases. Machines don't learn, they just translate our assumptions into results. These models are can be wrong, but most of the time we will never know with measurable certainty. This is not enough to claim that they are true.

The following two sentences summarize this paper better than I could in one page:

– "Developing models, and testing their connection to the phenomena, is a serious problem in statistics."

– “an enormous amount of fiction has been produced, masquerading as rigorous science.”