From Maxent to Machine Learning and Back

T. Sears
College of Engineering and Computer Science,
The Australian National University
Canberra, Australia
(tim.sears@anu.edu.au)

Abstract

To Jaynes, in his original paper [1], maxent is “a method of reasoning which ensures that no unconscious arbitrary assumptions have been introduced”, while, fifty years later, the MAXENT conference home page suggests that the method “is not yet fully available to the statistics community at large.” In fact, it is possible to see that generalized maxent problems, often in disguise, do play a significant role in machine learning and statistics. Deviations from the classic form of the problem are typically used to incorporate some form of prior knowledge. Sometimes that knowledge would be difficult or impossible to represent with only linear constraints or an initial guess for the density.

To clarify these connections, we begin with the classic maxent problem studied by Jaynes (1957), and then generalize the problem until it encompasses a large class of problems studied by the machine learning community. Relaxed constraints, generalizations of Shannon entropy and a few tools from convex analysis make the task relatively straightforward. In many examples discussed, the original maxent problem remains embedded as a special case. By providing a trail back to the original maxent problem we highlight the potential for cross-fertilization between the two fields. For example, machine learners can gain the benefits of entropic regularization and maxent researchers might find uses for the SVM-loss function as a form of constraint relaxation. In addition, the use of Fenchel duality in describing the connection, instead of the usual Lagrangian presentation, leads naturally to component-based model design. The potential for estimating all these models through a common framework is also discussed.

References:


Key Words: Generalized Maxent, Optimization, Convex Analysis