An Introduction to Natural Gradient with Applications in Optimization and Neural Networks

Luigi Malagò, Principal Investigator at the Transylvanian Institute of Neuroscience, Romania.

Abstract

In Information Geometry parameterized statistical models are represented as differentiable manifolds, by using the language of affine and Riemannian geometry. Under certain regularity conditions, a parameterization for a statistical model corresponds to a set of coordinates for the manifold, a curve is a statistical model and its velocity vectors in the tangent space are centered random variables. The geometry of the manifold can be constructed starting from the Kullback-Leibler divergence between two densities, which leads to a metric tensor associated to the Fisher information matrix. In statistics and machine learning we are often interested in solving an optimization problem over the parameters of a statistical model, typical examples are the minimization of the expected value of a function with respect to a parameterized distribution in a statistical model and maximum likelihood estimation. By providing a manifold structure for the model, Information Geometry allows the use of Riemannian methods in optimization, where one of the most well known examples is probably the natural gradient, i.e., the Riemannian gradient of a function defined over a manifold, computed with respect to the Fisher information metric. In this talk we review two applications of natural gradient: the first one for black-box derivative-free optimization and the other for the training of Helmholtz Machines, generative models based on sigmoid belief networks.