

## Abstract

Information Geometry is the differential geometric study of the manifold of probability models, and promises to be a unifying geometric framework for investigating statistical inference, information theory, machine learning, etc. Instead of using metric for measuring distances on such manifolds, these applications often use “divergence functions” for measuring proximity of two points (that do not impose symmetry and triangular inequality). Divergence functions are tied to generalized entropy and cross-entropy functions widely used in machine learning and information sciences. It turns out that divergence functions enjoy pleasant geometric properties – they induce what is called “statistical structure” on a manifold  $M$ : a Riemannian metric  $g$  together with a pair of torsion-free affine connections  $D, D^*$ , such that  $D$  and  $D^*$  are both Codazzi coupled to  $g$  while being conjugate to each other. We recently characterize holomorphicity of  $D, D^*$  in the (para-)Hermitian setting, and show that statistical structures (with torsion-free  $D, D^*$ ) can be enhanced to Kahler or para-Kahler manifolds. The surprisingly rich geometric structures and properties of a statistical manifold open up the intriguing possibility of geometrizing statistical inference, information, and machine learning in string-theoretic languages.