Information Geometry: Geometric Structures for Information Sciences

Hadi Hadavi, Sepehr Heidari, Borna Khodabandeh, Danial zinati

Introduction

Information geometry provides a powerful geometric framework for studying and analyzing problems in various information sciences, including statistics, machine learning, and signal processing. By representing families of probability distributions or models as manifolds equipped with geometric structures, information geometry offers principled and unified approaches to tackling tasks such as inference, learning, and decision-making.

Background and Fundamentals

In this project, We have studied the foundational concepts of information geometry, as outlined in the introductory survey by Frank Nielsen.[1] The key structures explored include:

1. **Fundamentals of Differential Geometry(DG):** We studied about the fundamentals of differential geometry, namely the definitions of manifolds($M,\ g,\ \nabla$), tangent/cotangent spaces($T_p M,\ T_p^* M$), parallel transport($v_{c(t)} = \prod_{c(0)\to c(t)}^{\nabla} v \in T_{c(t)} M$), covariant derivatives($\nabla_X Y$), the metric tensor(g) and other such basic concepts in geometry which will be important in the next chapters.

2. **Conjugate Connection Manifolds (CCMs)($M,\ g,\ \nabla,\ \nabla^*$)**: Manifolds equipped with a metric tensor and a pair of conjugate affine connections that preserve the metric under dual parallel transport. These manifolds capture the dualistic nature of many information-theoretic quantities.

3. **Statistical Manifolds: Derived from CCMs($M,\ g,\ C$)**, statistical manifolds are endowed with a metric tensor and the Amari-Chentsov cubic tensor, encoding the skewness of the underlying statistical model.

4. **Dually Flat Manifolds($M,\ \nabla^2 F,\ {}^F\nabla,\ {}^F\nabla^*$)**: Obtained from Bregman divergences, these manifolds exhibit a flat structure with respect to both affine connections, enabling powerful geometric results such as Pythagorean projections. One may use the fisher information matrix as the metric tensor if they choose the entropy as their potential function.

Connection to Information sciences

While studying these mathematically rich aspects of information geometry, it has led us to get a more fundamental and better view of information sciences as a whole, gaining new perspectives on machine learning, optimal transport theory, statistical divergences and such.

Planned Work and Explorations

Building upon the theoretical foundations, We plan to investigate the following aspects and applications of information geometry:

1. **Natural Gradient Descent**[2]: Leveraging the Riemannian gradient on statistical manifolds, we will explore the formulation and advantages of natural gradient descent algorithms for optimization problems in machine learning and statistics, particularly exploring their invariance to parametrization. Similar work are introduced such as the mirror gradient descent.[3]

2. **Parallel Transport and Invariance**: we will study the parallel transport of vectors on information manifolds and its implications for invariance properties in statistical inference tasks. The invariance of the Fisher metric and f-divergences under specific transformations will be analyzed. We will also study geodesics in such manifolds[4] and their connections to optimal transport theory.

3. **Statistical Mixture Clustering**: Utilizing the dually flat structure of mixture family manifolds, we aim to implement algorithms for clustering statistical mixtures with shared component distributions.

4. **Connections to Other Fields**: Throughout the project, we will explore the connections between information geometry and other domains, including optimization, control theory, and mathematical physics, where geometric structures play a crucial role.

References

[1] F. Nielsen, “An elementary introduction to information geometry,” *Entropy*, vol. 22, no. 10, p. 1100, Sep. 2020, doi: 10.3390/e22101100.

[2] “Natural Gradient Descent - Agustinus Kristiadi.” Accessed: May 21, 2024. [Online]. Available: https://agustinus.kristia.de/techblog/2018/03/14/natural-gradient/

[3] G. Raskutti and S. Mukherjee, “The Information Geometry of Mirror Descent.” arXiv, Apr. 29, 2014. Accessed: May 21, 2024. [Online]. Available: http://arxiv.org/abs/1310.7780

[4] V. Seguy and M. Cuturi, “Principal Geodesic Analysis for Probability Measures under the Optimal Transport Metric,” in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., 2015. Accessed: May 21, 2024. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2015/hash/f26dab9bf6a137c3b6782e562794c2f2-Abstract.html