Information Geometry

G H Lathrom

May 6, 2021

1 Background

Information Geometry is a area which combines Differential Geometry, Probability Theory and Information Theory into one discipline. Because of this, it is somewhat difficult to get a grasp on all of the concepts which go into Information Theory. By molding these three disciplines together we also see connections between Physics and Machine Learning also with implications for Artificial Intelligence. Let's just see where all of these things take us, while supplementing necessary background along the way. One of the books of interest will be [AD08] along with a collection of papers available here. We begin with the connection between Information Geometry and Mechanics using [LZ17] as the guide.

$$\Gamma^{ij}{}_{k} = \cdots$$

2 Mechanics

Suppose we have a manifold Q with local coordinates $(q^1, q^2, \ldots, q^n)$. That is, Q is a collection of charts, $\{U_\alpha\}$, which is an open cover of Q. For each chart U_α we have a diffeomorphism

$$\varphi_\alpha : U_\alpha \to \varphi_\alpha(U_\alpha) \subseteq \mathbb{R}^n,$$

where, together with the natural projection functions from the product space $\mathbb{R}^n$ onto the individual coordinates, $\pi^i : \mathbb{R}^n \to \mathbb{R}$, we obtain the coordinate functions $q^i = \pi^i \circ \varphi_\alpha$. This gives rise to what we have referred to as the local coordinates above for the chart U_α. A **Lagrangian** is simply a function from the tangent bundle into the real numbers, $L : TQ \to \mathbb{R}$.

3 Cohomology

Beginning the discussion of Cohomology we have the notion of a singular chain complex.

$$\cdots \longrightarrow C_{i+1} \xrightarrow{\partial_{i+1}} C_i \xrightarrow{\partial_i} C_{i-1} \longrightarrow \cdots$$

If we take vectors $u_0, u_1, \dots, u_n \in \mathbb{R}^n$ which are affinely independent, that is $\{u_i - u_0\}_{i=1}^n$ are linearly independent, then the simplex is the set

$$C^n = \left\{ p^0 u_0 + p^1 u_1 + \cdots + p^n u_n \;\middle|\; p^i \geq 0 \text{ for all } i \text{ and } \sum_i p^i = 1 \right\}.$$

One could also recognize the n dimensional probability simplex, or the set of all finite probability distributions on n elements, as

$$\Delta^n = \left\{ \mathbf{p} \in \mathbb{R}^n \;\middle|\; p^i \geq 0 \text{ for } i = 0, 1, \dots, n-1 \text{ and } \sum_i p^i = 1 \right\}$$

so that the simplex, C^n may be recognized as the set of all possible expected-values with respect to the probability distributions

$$C^n = \left\{ E[U|\mathbf{p}] = \langle U \rangle_{\mathbf{p}} \mid \mathbf{p} \in \Delta^{n+1} \right\}$$

where U is a random variable on the set $\{u_0, u_1, \dots, u_n\}$ where $\text{Prob}(U = u_i \mid \mathbf{p}) = p^i$.

References

[AD08] K. Arwini and C. Dodson. *Information Geometry.* Lecture Notes in Mathematics 1953. Springer, 2008.

[LZ17] M. Leok and J. Zhang. Connecting information geometry and geometric mechanics. *Entropy*, 2017.