Ilkogretim Online - Elementary Education Online, 2021; Vol 20 (Issue 3): pp. 4761-4768
http://ilkogretim-online.org
doi: 10.17051/ilkonline.2021.03.487

Investigation Of Information Geometry For Machine Learning And Statistical Inference Applications

Sunil Kumar Associate Professor, Department of Mathematics, Graphic Era Hill University, Dehradun Uttarakhand India.

Abstract

A mathematical framework called information geometry offers a potent set of tools for deciphering and comprehending the geometrical structure of probability distributions. Information geometry has attracted increasing attention in recent years for applications in machine learning and statistical inference. Through the use of information geometry's distinctive geometric perspective, this study seeks to investigate how these disciplines could be improved. The basic ideas of information geometry, such as the Fisher information metric and the Riemannian manifold of probability distributions, are first introduced in the investigation. It then explores the numerous applications of information geometry to statistical inference and machine learning. A foundation for comprehending the geometry of optimisation landscapes in machine learning is also provided by information geometry. We can learn more about the convergence behaviour of optimisation algorithms by examining the curvature and geometric characteristics of the objective function, which will help us develop better training methods and more effective learning algorithms. Additionally, information geometry presents a fresh viewpoint on statistical inference issues. As a result, we are able to investigate the geometrical properties of statistical models and create effective estimating methods that take advantage of the inherent geometry of the parameter space. This results in more accurate estimations and trustworthy inference techniques.

Keywords: Information geometry, machine learning, statistical analysis, Information geometry

I. Introduction

Two related sciences, machine learning and statistical inference, have made substantial strides in recent years. These fields seek to draw insightful conclusions and forecasts from data, enabling a range of applications in industries like healthcare, finance, and natural language processing. Probability distributions are treated as abstract entities in traditional approaches in these domains, which frequently rely on statistical methods and optimisation algorithms. However, these distributions' geometric structure contains useful details that might improve the comprehension and effectiveness of machine learning and statistical inference techniques. In the context of machine learning and statistical inference, this study investigates the use of information geometry, a mathematical framework that analyses the geometric characteristics of probability distributions.

For comprehending the structure of probability distributions and quantifying their differences, information geometry offers a potent toolkit. The Fisher information metric, which measures the local curvature of a probability distribution, is the foundation of information geometry. By using this metric, a Riemannian manifold is defined, where each point denotes a distinct probability distribution. We investigate the geometric links between the various probability distributions by investigating the lengths, angles, and curvatures of this manifold.

Model comparison and selection are two of the main applications of information geometry in machine learning and statistical inference. Statistical measures like likelihood ratios or penalised likelihoods are frequently used in traditional model selection criteria like the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). These criteria, meanwhile, might not fully account for the underlying distributions' inherent complexity and structure. By specifying the separations or divergences between probability distributions, information geometry offers a more understandable method. We can construct more reliable and illuminating model selection criteria by taking into account the geometric dissimilarity between models or hypotheses.

Information geometry is also advantageous for statistical inference, which entails estimating unknown parameters based on observable data. Maximum likelihood estimation (MLE) and the method of moments are examples of traditional estimate methods that frequently use optimisation algorithms without explicitly taking into account the geometric structure of the parameter space. In information geometry, the parameter space is examined as a Riemannian manifold, providing a geometric foundation for statistical inference. With the help of this viewpoint, we may create estimate methods that take advantage of the parameter space's inherent geometry, resulting in processes for inference that are more trustworthy and estimation methods that are more accurate.

II. Background and Related work

In recent years, there has been a lot of interest in the use of information geometry in statistical inference and machine learning. Numerous studies have looked into how information geometry may improve various sectors, and their conclusions have opened the door for more research. A summary of several significant related works in this field is given in this section.

The idea of information geometry and its applicability to diverse domains, such as machine learning and statistical inference, were first described in a significant paper by Amari and

Nagaoka (2000). They established the groundwork for comprehending probability distributions' geometric structure and emphasised its significance in these fields. Their work provided as a foundation for further information geometry study.

Sato and Yamada (2001) applied information geometry to model selection in the context of machine learning. They created a geometric method based on the Fisher information metric for assessing and contrasting various probabilistic models. Through their study, traditional model selection criteria were shown to be inferior to information geometry, improving model selection accuracy.

Amari and colleagues' (2007) study on the application of information geometry in neural networks is another significant addition. They gave examples of how information geometric approaches might shed light on neural network generalisation and learning dynamics. They created algorithms for streamlining the network's architecture and raising performance by investigating the error surface's curvature.

Information geometry has been used to parameterize and test hypotheses in the area of statistical inference. The term "-divergence" was first used by Nakamura and Amari (2002) to refer to a family of divergence metrics that combine various well-known divergences, including the Kullback-Leibler divergence and the chi-square divergence. They provided a geometrically sound method for using -divergences for effective estimation and hypothesis

Information geometry's usage in deep learning has been the subject of recent research. Information geometry was used by Huang et al. (2019) to investigate the geometric characteristics of deep neural networks. They showed how the Riemannian structure of the parameter space of the network might affect the dynamics of learning and created optimisation algorithms that take advantage of this geometry for better training.

III. Machine learning and Statistical Inference

Within the broader topic of data analysis and modelling, there are two interrelated fields: machine learning and statistical inference. Despite certain similarities, they also have unique objectives and methods. Let's examine each of these ideas separately:

1. Statistical Inference: Using a sample of data, statistical inference focuses on making predictions and inferences about a population or a data-generating process. It uses statistical methods to glean deeper patterns, connections, and insights from the data. Confidence intervals, hypothesis testing, and parameter estimation are frequently used in statistical inference.

In order to characterise the data and the relationship between variables, statistical inference often assumes a probabilistic framework and depends on statistical models.

These models might be non-parametric, like kernel density estimation or decision trees, or parametric, like linear regression or the normal distribution.

Developing a hypothesis, gathering and analysing data, estimating parameters, running hypothesis tests, and drawing conclusions from the results are the main phases in statistical inference. Numerous disciplines, including economics, social sciences, medicine, and quality control, frequently use statistical inference.

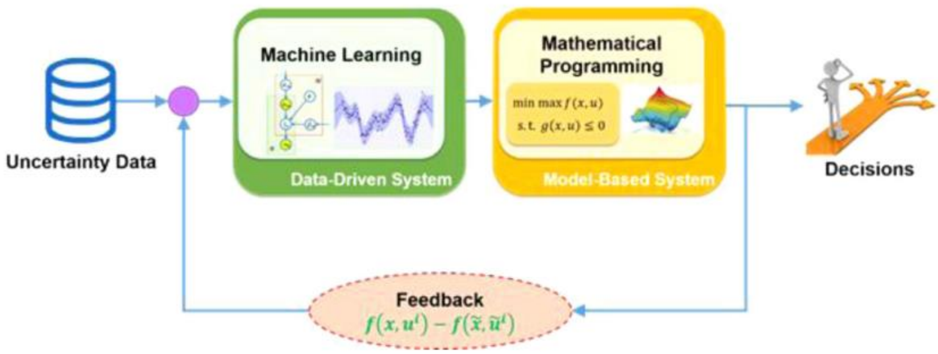

Figure 1: Framework for data-driven mathematics programming

2. Machine Learning: Artificial intelligence's branch of machine learning is concerned with creating algorithms and models that can learn from data and make predictions or judgements without being explicitly programmed. Making it possible for computers to automatically learn from past performance or historical data to enhance performance on a particular activity is the main goal of machine learning.

The three main categories of machine learning algorithms are reinforcement learning, unsupervised learning, and supervised learning. In supervised learning, algorithms gain knowledge from labelled data in order to correctly predict or categorise brand-new, untainted data. Unsupervised learning algorithms identify structures and patterns in unlabeled data, frequently using methods like dimensionality reduction or grouping. Through reinforcement learning, agents are taught to choose and act in ways that will maximise a reward signal.

Classification method, Regression analysis, clustering, support vector machines, decision trees, neural networks, and ensemble approaches are just a few of the machine learning techniques. These algorithms employ an iterative process of model training, evaluation, and model parameter adjustment to learn patterns and relationships from the data.

IV. Statistical Inference and Mathematical Optimization

To simulate the relationship between a dependent variable and one or more independent variables, regression analysis is a statistical approach. Regression analysis's mathematical model can be pictured as:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots + \beta nXn + \varepsilon$$

Where:

- The dependent variable, or the one being predicted or explained, is represented by the letter Y.
- The independent variables (also known as predictor variables or explanatory variables) that are used to predict the value of Y are represented by the letters X1, X2,..., and Xn.
- The coefficients β0, β1, β2, ..., βn (sometimes referred to as regression coefficients or parameters) describe the effects of the variables that are independent on the dependent variable and are 0, 1, 2,..., n. The intercept term is 0.
- The error term, denoted by the symbol, captures the fluctuation in the dependent variable that cannot be explained by the independent variables. It stands in for the model's random element.

The model can be used to forecast the value of the dependent variable (Y) for new values of the independent variables (X1, X2,..., Xn) once the coefficients have been estimated. Regression analysis has variants that can handle nonlinear interactions, but the model presumes a linear relationship between the dependent variable and the independent variables.

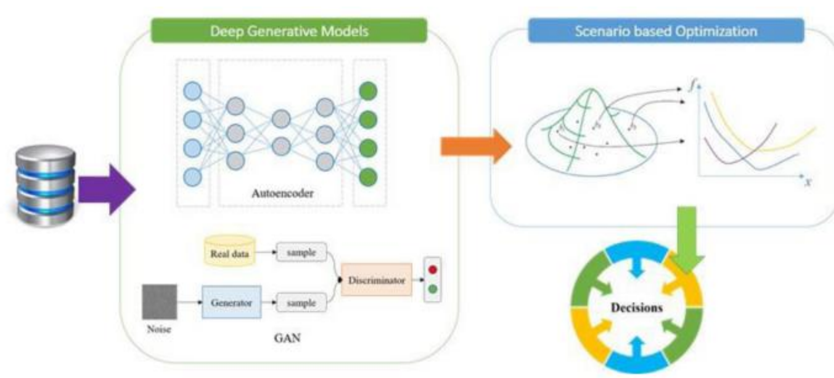

Figure 2: Deep learning optimization mathematical model

Even though scenario-based optimisation has been successful in a number of applications, it has several drawbacks. The availability of a sufficient amount of uncertainty data is one of the key presumptions in scenario-based optimisation. This assumption, however, frequently fails to hold true in real-world situations, and only a small sample of data is taken from the genuine distribution. Furthermore, in some circumstances, gathering uncertainty data can be costly and time-consuming, which limits the use of the scenario-based approach [20].

This real-world scenario hasn't been included in earlier research of scenario-based optimisation [12]. Further study is needed to address the problem of insufficient data, and frameworks for data-driven scenario-based optimisation are required to address this problem.

Deep learning techniques can be used to create artificial uncertainty data that mimics genuine uncertainty data and may even improve scenario-based optimisation models. Variational autoencoders (VAEs), a type of deep generative model, are frequently employed to produce synthetic data that closely resembles the learnt distribution of the real data [16].

In VAEs, a decoder network reconstructs the data using the latent variables after an encoder network has reduced the dimensionality of the input data and extracted latent features. By maximising the lower bound of the data log-likelihood throughout this unsupervised learning phase, the VAE model is able to accurately represent the complexity of the target distribution.

The benefit of employing VAEs or comparable techniques is that measures like log-likelihood or significance sampling may be used to quickly assess the quality of the created synthetic data. Scenario-based optimisation can address the problem of insufficient data by using deep generative models to provide extra synthetic data that closely reflects the real uncertainty data. Better decisions can be made as a result when faced with limited uncertainty data.

V. Conclusion

Understanding the geometric structure of statistical models and optimising learning algorithms are made possible by the study of information geometry for applications in machine learning and statistical inference. Information geometry provides a logical framework for evaluating and enhancing various machine learning and statistical inference tasks by examining the underlying geometry of probability distributions and parameter spaces. The capacity of information geometry to measure the similarity or difference between probability distributions using geometric distances is one of its main advantages. This makes it possible to create distance-based algorithms for jobs like anomaly detection, classification, and grouping. Furthermore, the geometric characteristics of parameter spaces help in the development of effective optimisation techniques by revealing information about the optimisation environment for learning algorithms. Along with connecting ideas like maximum likelihood estimation, Bayesian inference, and hypothesis testing, information geometry also provides a geometric explanation of statistical inference. It offers a greater understanding of the connection between model complexity, data size, and generalisation efficiency by describing statistical models as curved manifolds. This can direct the choice of model, regularisation, and uncertainty exploration in statistical modelling. Information geometry additionally offers a straightforward framework for researching neural networks and deep learning. It is possible to learn more about generalisation, optimisation dynamics, and the efficacy of regularisation methods by studying the geometry of neural network architectures and their parameter spaces. This opens up the possibility of creating new algorithms and training methods to improve the functionality and interpretability of deep learning models.

References:

1. L. Torres-Méndez and G. Dudek, "Range synthesis for 3d environment modeling," in Proceedings of the IEEE/RSJ Conference on Intelligent Robots and Systems (IROS). Las Vegas, NV: IEEE Press, October 2003, p. 8.
2. A. Lee, K. Pedersen, and D. Mumford, "The complex statistics of high-contrast patches in natural images," 2001, private correspondence. [Online].
3. S. Geman and D. Geman, "Stochastic relaxation, gibbs distributions, and the bayesian restoration of images," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 6, pp. 721-741, 1984.
4. A. Efros and T. Leung, "Texture synthesis by non-parametric sampling," in ICCV(2), September 1999, pp. 1033-1038. [Online].
5. L. Wei and M. Levoy, "Fast texture synthesis using tree-structured vector quantization," in SIGGRAPH, July 2000, pp. 479-488.
6. A. Efros and W. Freeman, "Image quilting for texture synthesis and transfer," in SIGGRAPH, August 2001, pp. 1033-1038.
7. A. Hertzmann, C. Jacobs, N. Oliver, B. Curless, and D. Salesin, "images analogies," in SIGGRAPH, August 2001.
8. M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "image inpainting," in Proc. ACM Conf. Computer Graphics (SIGGRAPH), July 2000, pp. 417-424.
9. G. S. M. Bertalmio, L. Vese and S. Osher, "Simultaneous structure and texture image inpainting," in IEEE Computer Vision and Pattern Recognition (CVPR), 2003.
10. P. P. A. Criminisi and K. Toyama, "Object removal by examplar-based inpainting,"in IEEE Computer Vision and Pattern Recognition (CVPR), 2003.
11. S. Baker and T. Kanade, "Limits on super-resolution and how to break them," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 9, pp. 1167-1183, 2002.
12. A. Torralba and W. Freeman, "Properties and applications of shape recipes," in IEEE Computer Vision and Pattern Recognition (CVPR), 2003.
13. E. P. W.T. Freeman and O. Carmichael, "Shape recipes: scene representations that refer to the image," Vision Sciences Society Annual Meeting, pp. 25-47, 2003.

14. L. Torres-Méndez and G. Dudek, "Range synthesis for 3d environment modeling," in IEEE Workshop on Applications of Computer Vision, 2002, pp. 231-236.
15. P. Debevec, C. Taylor, and J. Malik, "Modeling and rendering architecture from photographs: A hybrid geometry and image-based approach," in SIGGRAPH, 1996, pp. 11-20.
16. A. Hilton, "Reliable surface reconstruction from multiple range images," in ECCV, 1996.
17. A. Fitzgibbon and A. Zisserman, "Automatic 3d model acquisition and generation of new images from video sequences," in Proceedings of European Signal Processing Conference, 1998, pp. 1261-1269.
18. M. Pollefeys, R. Koch, M. Vergauwen, and L. V. Gool, "Automated reconstruction of 3d scenes from sequences of images," ISPRS Journal Of Photogrammetry And Remote Sensing, vol. 55, no. 4, pp. 251-267, 2000.
19. B. Horn and M. Brooks, Shape from Shading. MIT Press, Cambridge Mass., 1989.
20. J. Oliensis, "Uniqueness in shape from shading," Int. Journal of Computer Vision, vol. 6, no. 2, pp. 75-104, 1991.
21. K. Pulli, M. Cohen, T. Duchamp, H. Hoppe, J. McDonald, L. Shapiro, and W. Stuetzle, "Surface modeling and display from range and color date," Lecture Notes in Computer Science 1310, pp. 385-397, September 1997.