

1. Record Nr.	UNINA9910552748103321
Autore	Eguchi Shinto
Titolo	Minimum Divergence Methods in Statistical Machine Learning : From an Information Geometric Viewpoint // by Shinto Eguchi, Osamu Komori
Pubbl/distr/stampa	Tokyo : , : Springer Japan : , : Imprint : Springer, , 2022
ISBN	9784431569220 4431569227
Edizione	[1st ed. 2022.]
Descrizione fisica	1 online resource (224 pages)
Disciplina	006.31
Soggetti	Statistics Computer science - Mathematics Mathematical statistics Statistics in Engineering, Physics, Computer Science, Chemistry and Earth Sciences Statistical Theory and Methods Probability and Statistics in Computer Science
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Note generali	Includes index.
Nota di contenuto	Information geometry -- Information divergence -- Maximum entropy model -- Minimum divergence method -- Unsupervised learning algorithms -- Regression model -- Classification. .
Sommario/riassunto	This book explores minimum divergence methods of statistical machine learning for estimation, regression, prediction, and so forth, in which we engage in information geometry to elucidate their intrinsic properties of the corresponding loss functions, learning algorithms, and statistical models. One of the most elementary examples is Gauss's least squares estimator in a linear regression model, in which the estimator is given by minimization of the sum of squares between a response vector and a vector of the linear subspace hulled by explanatory vectors. This is extended to Fisher's maximum likelihood estimator (MLE) for an exponential model, in which the estimator is provided by minimization of the Kullback-Leibler (KL) divergence between a data distribution and a parametric distribution of the exponential model in an empirical analogue. Thus, we envisage a

geometric interpretation of such minimization procedures such that a right triangle is kept with Pythagorean identity in the sense of the KL divergence. This understanding sublimates a dualistic interplay between a statistical estimation and model, which requires dual geodesic paths, called m-geodesic and e-geodesic paths, in a framework of information geometry. We extend such a dualistic structure of the MLE and exponential model to that of the minimum divergence estimator and the maximum entropy model, which is applied to robust statistics, maximum entropy, density estimation, principal component analysis, independent component analysis, regression analysis, manifold learning, boosting algorithm, clustering, dynamic treatment regimes, and so forth. We consider a variety of information divergence measures typically including KL divergence to express departure from one probability distribution to another. An information divergence is decomposed into the cross-entropy and the (diagonal) entropy in which the entropy associates with a generative model as a family of maximum entropy distributions; the cross entropy associates with a statistical estimation method via minimization of the empirical analogue based on given data. Thus any statistical divergence includes an intrinsic object between the generative model and the estimation method. Typically, KL divergence leads to the exponential model and the maximum likelihood estimation. It is shown that any information divergence leads to a Riemannian metric and a pair of the linear connections in the framework of information geometry. We focus on a class of information divergence generated by an increasing and convex function U , called U -divergence. It is shown that any generator function U generates the U -entropy and U -divergence, in which there is a dualistic structure between the U -divergence method and the maximum U -entropy model. We observe that a specific choice of U leads to a robust statistical procedure via the minimum U -divergence method. If U is selected as an exponential function, then the corresponding U -entropy and U -divergence are reduced to the Boltzmann-Shannon entropy and the KL divergence; the minimum U -divergence estimator is equivalent to the MLE. For robust supervised learning to predict a class label we observe that the U -boosting algorithm performs well for contamination of mislabel examples if U is appropriately selected. We present such maximal U -entropy and minimum U -divergence methods, in particular, selecting a power function as U to provide flexible performance in statistical machine learning. .
