Full Title of Article

Author Name1 ABC@SAMPLE.COM
Address 1

Author Name2 XYZ@SAMPLE.COM
Address 2

Editor: Cheng Soon Ong and Tu Bao Ho

Abstract

This paper aims at bringing out the merits of using the Laplace distribution over the Normal distribution. The choice of the best distribution is objectively made using the minimum message length (MML) principle. The message length for transmitting data using a Laplace distribution is derived and its parameters are estimated. This method of transmission is compared with that of transmistting the data using the Normal distribution. This is explored in the context of superposition of protein structures. The optimal superposition of protein structures (described using a Normal distribution) minimizing the L2 norm is computed using the Kearsley's method and the superposition minimizing the L1 norm (described using a Laplace model) is approximated using Monte Carlo simulation. These two are compared with respect to their model complexity and the overall fit to the data using MML.

Keywords: Laplace, Normal, MML, Kearsley, Monte Carlo simulation

1. Introduction

Normal distribution is widely used in modelling a set of data whose true distribution is unknown. In many problems, the objective function is formulated as a sum of squares, (the L2 norm) and this function is minimized or maximized depending on the application. Normal distribution has a huge impact on the cost function because of the squared nature of the individual terms. If there are outliers in the dataset, the final inference might be skewed to accommodate the outliers in the model description. The Laplace distribution, however, is robust to outliers as the objective function involves the sum of the absolute values of the difference of the individual terms (L1 norm). The choice of selecting a Laplace over Normal is investigated in this paper. This selection is made by formulating the objective function using Minimum Message Length (MML). The distribution which results in the best compression of data is chosen to be the best model.

Normal distribution is expressed in terms of squared difference from the mean μ (1),

$$\mathrm{pdf}(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \tag{1}$$

and hence objective functions based on minimizing the total least squares result in a closed form analytical solution. On the contrary, because of the mathematical nature of the Laplace

(2) which is expressed as the absolute difference from the mean, does not offer a closed form solution.

$$\text{pdf}(x) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right) \tag{2}$$

That being said, we however, believe that it has potential benefits. Laplace distribution is a model of choice in areas as diverse as signal processing (Eltoft et al., 2006), image denoising (Rabbani et al., 2006), gene expression studies (Bhowmick et al., 2006), market risk prediction (Haas et al., 2005), and machine learning (Cord et al., 2006). In most of these applications, a mixture model of Laplace distributions is used.

The main contribution of this work is the derivation of the message length expression and estimation of the MML parameters for the Laplace distribution. The general procedure to formulate the message length expression for transmitting data using some statistical model is outlined in Wallace and Freeman (1987). The MML method of estimating parameters for a number of distributions has been well established (Wallace, 2005). This general approach is followed in the derivation of the message length expression for the Laplace distribution.

We demonstrate its use in two cases. The first scenario involves data being randomly generated from an underlying Laplace distribution. This data is then fitted using both Normal and Laplace models. We use the derived formulation of MML for modelling the data using Laplace distribution. It is observed that the compression in message length is better when compared to transmitting it over a Normal distribution. Since the original distribution from which the data was generated happens to be Laplace, it is reasonable to guess that a Laplace model fits better. In general, if the data is generated from an underlying true distribution, modelling that data using the same statistical model results in a better fit. This is done as a validation check to ensure that the MML formulation for the Laplace case is consistent with this observation.

Further, we demonstrate its use by applying to the problem of optimal superposition. Examples of protein structures are considered and they are superposed using Kearsley's transformation. This results in an orientation of the proteins which minimizes the total least squares deviations. This optimal superposition is then encoded using a Normal distribution. A related superposition which minimizes the total absolute value of the deviations is then computed using a Monte Carlo simulation. This optimal superposition is encoded using a Laplace distribution.

2. The Minimum Message Length (MML) Framework

MML is an Bayesian inference method based on information theory. Using Shannon's insight (Shannon, 1948), encoding of an event E with probability $P(E)$ can be optimally done in $\log_2(P(E))$ bits.

Acknowledgments

Acknowledgements should go at the end, before appendices and references.

References

Debjani Bhowmick, AC Davison, Darlene R. Goldstein, and Yann Ruffieux. A laplace mixture model for identification of differential expression in microarray experiments. *Biostatistics*, 7(4):630–641, 2006.

Aurélien Cord, Christophe Ambroise, and Jean-Pierre Cocquerez. Feature selection in robust clustering based on laplace mixture. *Pattern Recogn. Lett.*, 27(6):627–635, April 2006.

T. Eltoft, Taesu Kim, and Te-Won Lee. On the multivariate laplace distribution. *Signal Processing Letters, IEEE*, 13(5), 2006.

Markus Haas, Stefan Mittnik, and Marc Paolella. Modeling and predicting market risk with laplace-gaussian mixture distributions. 2005.

H. Rabbani, M. Vafadust, and S. Gazor. Image denoising based on a mixture of laplace distributions with local parameters in complex wavelet domain. In *Image Processing, 2006 IEEE International Conference on*, pages 2597–2600, 2006.

Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 623–, july, october 1948.

C. S Wallace. *Statistical and Inductive Inference using Minimum Message Length*. Information Science and Statistics. SpringerVerlag, 2005.

C. S. Wallace and P. R. Freeman. Estimation and inference by compact coding. *Journal of the Royal Statistical Society. Series B (Methodological)*, 49(3):pp. 240–265, 1987.

Appendix A. First Appendix

This is the first appendix.

Appendix B. Second Appendix

This is the second appendix.