MANCHESTER
1824
The University of Manchester

Limits to classification performance relating Kullback-Leibler Divergence to Cohen's Kappa

Lisa Crow, Stephen Watts

Department of Physics and Astronomy, The University of Manchester, UK

Stephen.Watts@manchester.ac.uk

1. Introduction

Performance of Machine Learning (ML) classification algorithms is evaluated using training data and cross-validation.

Q. How do you know if one has reached the best possible performance ?

A. The Kullback-Leibler Divergence through the Chernoff-Stein Lemma gives the rate at which performance improves due to the underlying pdf of the two classes

Key step - write the **Confusion Matrix** to be compatible with the Chernoff-Stein Lemma.

Classified as → 1, 2; Input Classification; Row Sum

$$\begin{array}{c|cc|c} & 1 & 2 & \\ 1 & N_1^T \rightarrow & N_1^L & N_1 \\ 2 & N_2^L & \leftarrow N_2^T & N_2 \end{array}$$

True and Leaked entries for the two classes with N_1 and N_2 entries

Total entries, $N = N_1 + N_2$

$$N_1^T = N_1 - N_1^L \qquad N_1^L = \frac{N_1 N_2}{N}\exp(-K_{12})$$

$$N_2^L = \frac{N_1 N_2}{N}\exp(-K_{21}) \qquad N_2^T = N_2 - N_2^L$$

2. Methods

The confusion matrix is obtained from the classification algorithm. From this extract K_{12} , K_{21} and **Cohen's Kappa**, κ , which is the overall classification efficiency correcting for random chance. From this the parameter, K is calculated, with, $K = -\log_2(1-\kappa)$.

This parameter can be shown to be approximated by the Resistor Average Distance - R(P,Q) - which is the parallel combination of the Kullback-Leibler Divergences, D(P|Q) and D(Q|P). These divergences are independently estimated from the same training data. kNN estimate of R(P,Q) is called the Class Distance Resistance, CDR. One finds that, $\kappa \simeq 1 - 2^{-CDR(bits)}$

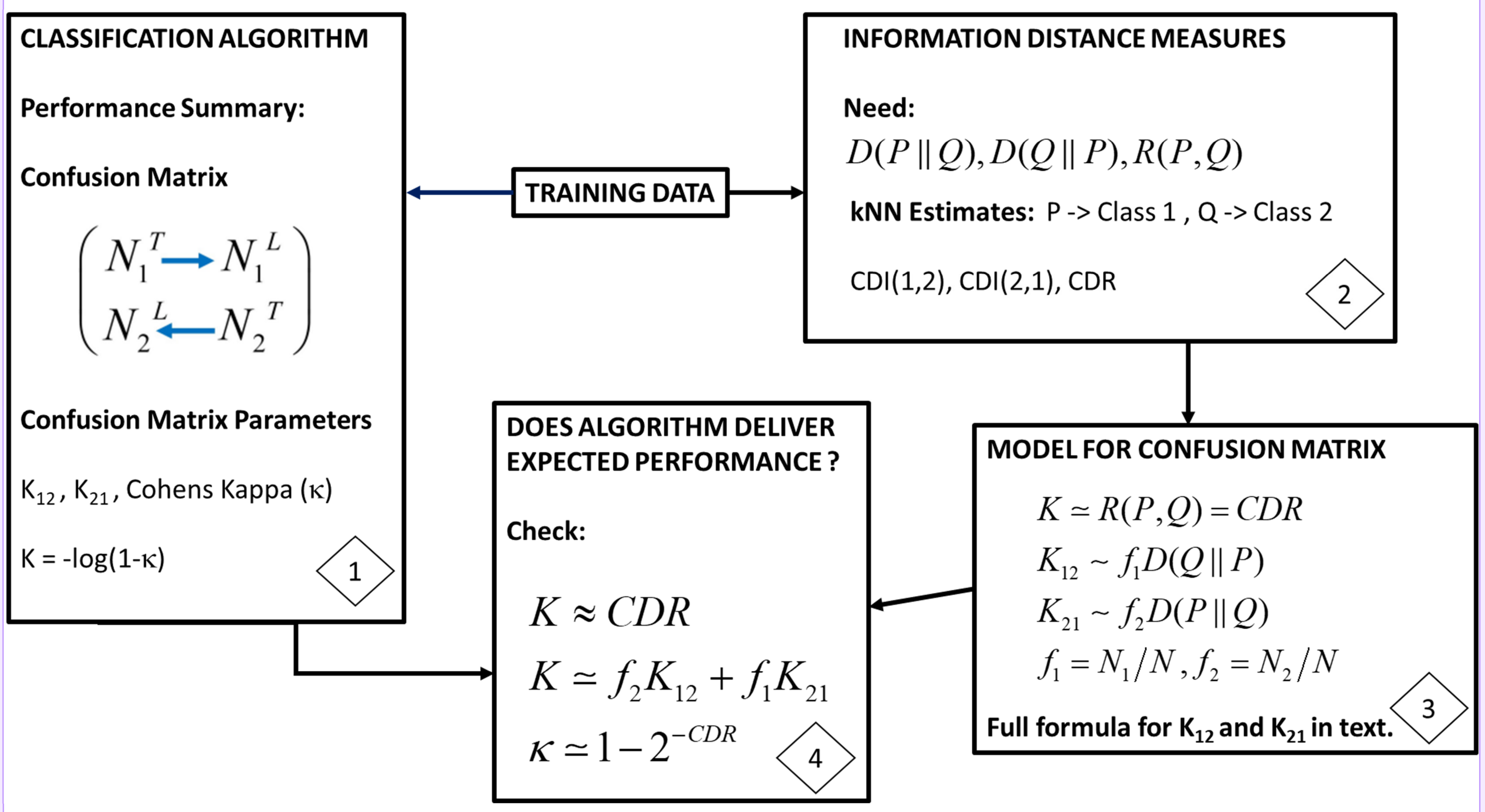

Figure 1 Methodology to compare performance of the classification algorithm with the expectation from information distance measures.

Method applied to Monte Carlo data and four datasets. Results section here shows the four datasets.

Written up in paper of same title : **https://arxiv.org/abs/2403.01571**

3. Results

Data	N_1/N_2	Total	Continuous/ Discrete	Machine Learning (WEKA Software, [5])
Breast Cancer [1]	212/357	569	30/0	Simple Logistic Regression
Bankruptcy [2]	107/143	250	0/6	J48 Decision Tree
Particle [3]	3736/1264	5000	8/0	Random Forest
Heart Disease (CHD)[4]	302/160	462	8/1	Logistic Regression

Kappa Scale

- Very Good
- Good
- Moderate
- Fair
- Poor

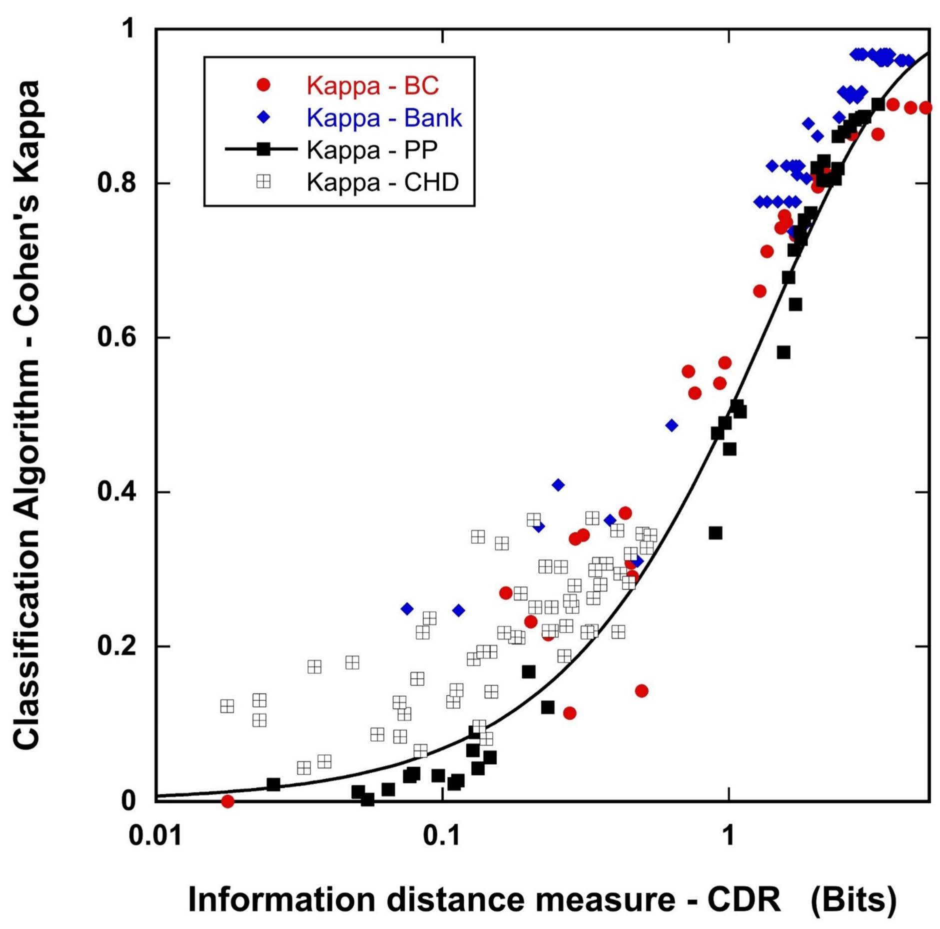

Figure 2. This figure combines the result of the classification algorithm performance using Cohen's Kappa versus the independently calculated Class Distance Resistance (CDR). CDR is an estimate of the Resistor Average Distance, which is an information distance measure. The curve gives the relation, $\kappa = 1 - 2^{-CDR}$. The points lie on this curve which indicates that the classification algorithm is performing as well as can be expected. The Kappa scale on the left-hand side is from D. G. Altman.

4. Conclusions

- ❑ First time that the performance of a Machine Learning Algorithm checked against predicted best case confusion matrix, estimated using the actual underlying probability density functions of the two classes from the same training data.
- ❑ Method applies to discrete, continuous or mixed data.
- ❑ Any algorithm, no matter how clever, can better this performance limit. For example, the coronary heart disease (CHD) data is only able to deliver " Fair" performance. This type of data is useful for risk analysis but not prediction.
- ❑ Apply to multi-class data by taking the classes in pairs.
- ❑ The formulation leads naturally to methods to understand imbalanced class machine learning classification.

References and Acknowledgements

1. W. Street, W. Wolberg, O. Mangasarian (1993), Int. Symp. On Electronic Imaging: Science and Technology, 1905, 861-870.
2. Kim and Han (2003) Expert Systems with Applications, 25, 637-646.
3. S.J. Watts & L. Crow, (2019). Nuclear Instruments and Methods in Physics Research Section A, 940, 441-447.
4. J. E. Rossouw et al., (1983) South African Medical Journal, 64, 430-436.
5. E. Frank, M. Hall, I. Witten (2016). " The WEKA Workbench.." Morgan Kaufmann, Fourth Edition.

LC thanks UKRI (STFC) and The University of Manchester for research student funding. SW thanks the Leverhulme Trust for support with an Emeritus Fellowship.