Fast low-rank metric learning

Dan-Theodor Oneaţă

Master of Science

Artificial Intelligence

School of Informatics

University of Edinburgh

2011

Abstract

This doctoral thesis will present the results of my work into the reanimation of lifeless human tissues.

Acknowledgements

Many thanks to my mummy for the numerous packed lunches; and of course to Igor, my faithful lab assistant.

Declaration

I declare that this thesis was composed by myself, that the work contained herein is my own except where explicitly stated otherwise in the text, and that this work has not been submitted for any other degree or professional qualification except as specified.

(*Dan-Theodor Oneaţă*)

Table of Contents

Chapter 1

Introduction

k Nearest Neighbours (kNN) is one of the oldest and simplest classification methods. The underlying idea is intuitive and straightforward: classify a given point according to a majority vote of its neighbours; the selected class is the one that is the most represented amongst the k nearest neighbours. This is easy to implement and it usually represents a good way to approach new problems or data sets. Despite its simplicty kNN is still a powerful tool, performing surprisingly well in practice, as most of the simple classifiers Holte (1993).

However, there are also other characteristics that make kNN an interesting method. First of all, kNN makes no assumptions about the underlying structure of the data, but lets the data "speak for themselves". So, the accuracy increases with the number of points in the data set. In fact, it approaches Bayes optimality as the cardinality of the training set approaches infinity and k is sufficiently large. Secondly, kNN constructs only simple local approximations; by combining all these it can represent complex global functions, which have non-linear decision boundaries and which are hardly captured parametric methdods. Laslty, kNN operates in a "lazy" fashion: the training data are just stored and their use is delayed until the the testing. The quasi-inexistent training[1] permits to easily add new training examples. On the other hand the storage is expensive and also the computational cost at testing time is often prohibitive (naively, there are n^2 computations to be done for each different query point). However, the drawback that affects the accuracy is how "nearest" is defined and how this distance is computed.

Neighbourhood Component Analysis (NCA) Goldberger et al. (2004) is a version

---

[1]There can be a training step in which the parameter k is tuned via cross-validation or we can we imagine different technique being applied for dimensionality reduciton

of *k*NN that tries to solve two of its drawbacks by learning a low-rank metric. This improves the accuracy—because the label information is now used for constructing a proper metric—and minimizes the computational cost for testing. However it introduces a training time—we have an objective function that is quadratic in the number of attributes and it has to be optimized using iterative methods.

- *k*-d trees Bentley (1975) were extensively used for accelerating range searches and nearest neighbour searches Moore (1991).

- The main aim of the project is to see if *k*-d trees can be used for reducing NCA's training and testing time without losing accuracy in the classification task. If time permits, we might investigate the use of other variants of tree structure, such as ball trees Omohundro (1989, 1991) (which perform better in high dimensions) or dual trees Gray and Moore (2003) (which scale linearly with the size of the data set). A further idea could be the use of kernels with compact support instead of the Gaussian ones in order to avoid the approximation errors introduced by *k*-d trees.

Bibliography

Bentley, J. L. (1975). Multidimensional binary search trees used for associative searching. *Commun. ACM*, 18:509–517.

Goldberger, J., Roweis, S., Hinton, G., and Salakhutdinov, R. (2004). Neighbourhood components analysis. In *Advances in Neural Information Processing Systems*. MIT Press.

Gray, A. and Moore, A. (2003). Nonparametric density estimation: Toward computational tractability. In *SIAM International Conference on Data Mining*, volume 2003. Citeseer.

Holte, R. (1993). Very simple classification rules perform well on most commonly used datasets. *Machine learning*, 11(1):63–90.

Moore, A. (1991). A tutorial on kd-trees. Extract from PhD Thesis. Available from `http://www.cs.cmu.edu/~awm/papers.html`.

Omohundro, S. (1989). *Five balltree construction algorithms*. Citeseer.

Omohundro, S. (1991). *Bumptrees for efficient function, constraint, and classification learning*. Citeseer.