bestknn: Best k for kNN Classification

Shahidul M. Islam
University of Western Ontario

A. I. McLeod
University of Western Ontario

Abstract

The function `bestknn` uses the pseudolikelihood method to select the best k for kNN classification.

Keywords: classification.

1. Introduction

Strauss (1992)

Holmes and Adams (2002, 2003)

(Islam 2008)

Ripley (1994, 1996)

```
R> library(bestknn)
R> data(trainSyn)
R> s.tr <- trainSyn
R> Xtr <- s.tr[, 1:2]
R> ytr <- ifelse(s.tr[, 3] == 0, -1, 1)
R> knnselect(X = Xtr, Y = ytr, maxk = 100)


[1] 66


R> knnselect(X = Xtr, Y = ytr, maxk = 100)
```
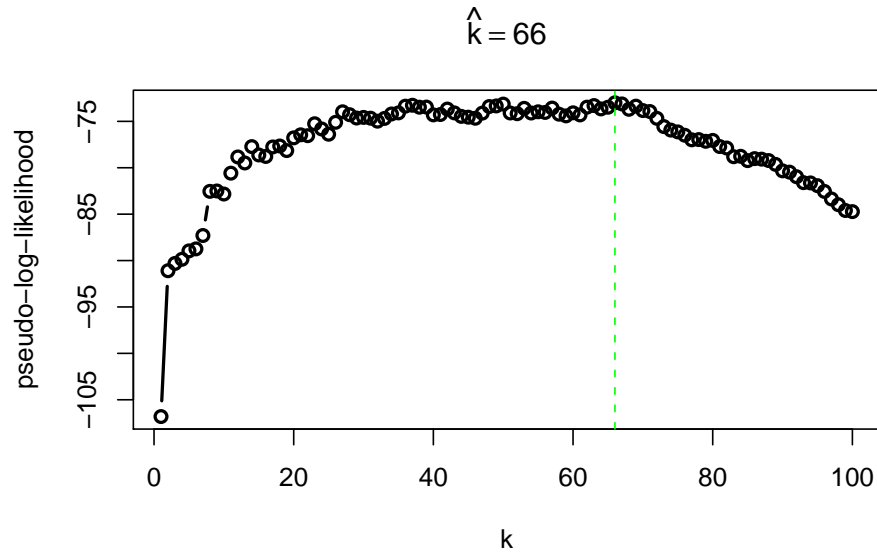
2. Concluding Remarks

Concluding remarks etc

References

Holmes CC, Adams NM (2002). "A probabilistic nearest neighbour method for statistical pattern recognition." *Journal of the Royal Statistical Society, B*, **64**, 295–306.

```
[1] 66
```


Figure 1: Optimal k for Ripley's synthetic training data

Holmes CC, Adams NM (2003). "Likelihood inference in nearest-neighbour classification models." *Biometrika*, **90**, 99–112.

Islam SM (2008). *Periodicity, change detection and prediction in microarrays.* Doctor of philosophy, The University of Western Ontario, London, Ontario.

Ripley BD (1994). "Neural networks and related methods for classification." *Journal of the Royal Statistical Society, B*, **56**, 409–456.

Ripley BD (1996). *Pattern recognition and neural networks.* Cambridge University Press, Cambridge.

Strauss D (1992). "The many faces of logistic regression." *The American Statistician*, **46**, 321–327.

Affiliation:

Shahidul M. Islam
University of Western Ontario

A.I. McLeod
E-mail: aimcleod@uwo.ca