NeuraMorph
Validation results

P. Baillehache

October 4, 2020

Contents

Introduction

This document introduces the validation results of the NeuraMorph algorithm.

Validation has been performed by using NeuraMorph on data sets publicly available on Internet, and comparing the accuracy of prediction made by NeuraMorph with the publicly available results achieved by other algorithms.

To avoid "lucky strike", results introduced here are produced with an instance of the NeuraMorph model selected amongst models produced with a systematic exploration of hyper parameters of the model, and based on its average of prediction accuracy over 5 different random splits of training/test samples.

1 Breast Cancer Wisconsin (Diagnostic) Data Set

Source:
   https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+(Diagnostic)

1.1 NeuraMorph results

Training parameters:

Nb inputs/outputs	30/2
Nb samples	569
Percentage of training samples	90%
weakUnitThreshold	0.900000
depth	5
maxLvlDiv	1
nbMaxInputsUnit	2
nbMaxUnitDepth	21
order	2
Time training	17.1s

Results:

Bias prediction (min/avg/sigma/max) and accuracy	c.00 [0.000,0.000,0.000,0.000] 100.00% c.01 [0.000,0.000,0.000,0.000] 100.00% .... [0.000,0.000,0.000,0.000] 100.00%
Bias training (min/avg/sigma/max) and accuracy	c.00 [0.000,0.008,0.062,2.000] 99.61% c.01 [0.000,0.008,0.062,2.000] 99.61% .... [0.000,0.011,0.062,2.828] 99.61%

1.2 Concurrent results

1.2.1 Multisurface method

"Dr. Wolberg [...] in collaboration with Prof. Mangasarian and two of his graduate students, Rudy Setiono and Kristin Bennett, [developped] a classifier [...] using the multisurface method (MSM) of pattern separation [...] that successfully diagnosed 97% of new cases."

Source:
http://pages.cs.wisc.edu/ olvi/uwmp/cancer.html

1.2.2 Additive logistic regression

"J. Friedman, T. Hastie, R. Tibshirani (1998). Additive Logistic Regression: a Statistical View of Boosting. Stanford University."

Predictive accuracy: 96.28%
Source:
https://www.openml.org/f/77 https://projecteuclid.org/euclid.aos/1016218223

2 Iris Data Set

Source:
  http://archive.ics.uci.edu/ml/datasets/Iris/

2.1 NeuraMorph results

Training parameters:

Nb inputs/outputs	4/3
Nb samples	150
Percentage of training samples	90%
weakUnitThreshold	0.975000
depth	2
maxLvlDiv	1
nbMaxInputsUnit	2
nbMaxUnitDepth	11
order	1
Time training	0.1s

Results:

Bias prediction (min/avg/sigma/max) and accuracy	c.00 [0.000,0.000,0.000,0.000] 100.00% c.01 [0.000,0.000,0.000,0.000] 100.00% c.02 [0.000,0.000,0.000,0.000] 100.00% .... [0.000,0.000,0.000,0.000] 100.00%
Bias training (min/avg/sigma/max) and accuracy	c.00 [0.000,0.000,0.000,0.000] 100.00% c.01 [0.000,0.015,0.086,2.000] 99.26% c.02 [0.000,0.015,0.086,2.000] 99.26% .... [0.000,0.021,0.000,2.828] 99.26%

2.2 Concurrent results

2.2.1 Additive logistic regression

"J. Friedman, T. Hastie, R. Tibshirani (1998). Additive Logistic
Regression: a Statistical View of Boosting. Stanford University."

Predictive accuracy: 94.67%
Source:
https://www.openml.org/f/77 https://projecteuclid.org/euclid.aos/1016218223

2.2.2 Neural networks

"Classification of Iris data set, Primoz Potocnik, Vitaly Borovinskiy"

Predictive accuracy:
Multilayer perceptron: 96.825% Radial basis function network: 96.825% Probabilistic neural network: 95.238%
Source:
http://lab.fs.uni-lj.si/lasin/wp/IMIT_files/neural/doc/seminar8.pdf

3 Annealing Data Set

Source:
https://archive.ics.uci.edu/ml/datasets/Annealing

3.1 NeuraMorph results

Training parameters:

Nb inputs/outputs	77/5
Nb samples	798
Percentage of training samples	90%
weakUnitThreshold	0.900000
depth	4
maxLvlDiv	4
nbMaxInputsUnit	2
nbMaxUnitDepth	3
order	2
Time training	14.1s

Results:

Bias prediction (min/avg/sigma/max) and accuracy	c.00 [0.000,0.076,0.196,2.000] 96.20% c.01 [0.000,0.025,0.113,2.000] 98.73% c.02 [0.000,0.000,0.000,0.000] 100.00% c.03 [0.000,0.101,0.226,2.000] 94.94% c.04 [0.000,0.000,0.000,0.000] 100.00% .... [0.000,0.143,0.196,2.828] 94.94%
Bias training (min/avg/sigma/max) and accuracy	c.00 [0.000,0.050,0.158,2.000] 97.50% c.01 [0.000,0.011,0.075,2.000] 99.44% c.02 [0.000,0.003,0.037,2.000] 99.86% c.03 [0.000,0.036,0.135,2.000] 98.19% c.04 [0.000,0.000,0.000,0.000] 100.00% .... [0.000,0.071,0.158,2.828] 97.50%

3.2 Concurrent results

3.2.1 Additive logistic regression

"J. Friedman, T. Hastie, R. Tibshirani (1998). Additive Logistic Regression: a Statistical View of Boosting. Stanford University."

Predictive accuracy: 99.67%
Source:
https://www.openml.org/f/77 https://projecteuclid.org/euclid.aos/1016218223

3.2.2 Rule based classifier

"Analysis of Different Classification Algorithms Applied to Anneal Dataset Using Data Mining Techniques, V.Kamalakkannan, Dr.D.

Ramyachitra, Computer Science Department, Bharathiar University, Coimbatore, Tamil Nadu, India"

Predictive accuracy:
Naive Bayes: 97.66%
SMO: 97.01%
OneR: 83.65%
J48: 95.88%

Source:
http://www.istpublications.com/temp/16_V_Kamalakkannan_Et_al_.pdf

4 Arrhythmia Data Set

Source:

https://archive.ics.uci.edu/ml/datasets/arrhythmia

4.1 NeuraMorph results

Training parameters:

Nb inputs/outputs	325/13
Nb samples	452
Percentage of training samples	90%
weakUnitThreshold	0.900000
depth	4
maxLvlDiv	0
nbMaxInputsUnit	2
nbMaxUnitDepth	1
order	2
Time training	60.8s

Results:

Bias prediction (min/avg/sigma/max) and accuracy	c.00 [0.000,0.267,0.369,2.000] 86.67% c.01 [0.000,0.044,0.151,2.000] 97.78% c.02 [0.000,0.133,0.261,2.000] 93.33% c.03 [0.000,0.044,0.151,2.000] 97.78% c.04 [0.000,0.044,0.151,2.000] 97.78% c.05 [0.000,0.000,0.000,0.000] 100.00% c.06 [0.000,0.000,0.000,0.000] 100.00% c.07 [0.000,0.000,0.000,0.000] 100.00% c.08 [0.000,0.044,0.151,2.000] 97.78% c.09 [0.000,0.000,0.000,0.000] 100.00% c.10 [0.000,0.000,0.000,0.000] 100.00% c.11 [0.000,0.000,0.000,0.000] 100.00% c.12 [0.000,0.489,0.500,2.000] 75.56% .... [0.000,0.754,0.337,2.828] 73.33%
Bias training (min/avg/sigma/max) and accuracy	c.00 [0.000,0.206,0.322,2.000] 89.68% c.01 [0.000,0.103,0.227,2.000] 94.84% c.02 [0.000,0.049,0.157,2.000] 97.54% c.03 [0.000,0.039,0.140,2.000] 98.03% c.04 [0.000,0.118,0.243,2.000] 94.10% c.05 [0.000,0.074,0.192,2.000] 96.31% c.06 [0.000,0.025,0.111,2.000] 98.77% c.07 [0.000,0.133,0.258,2.000] 93.37% c.08 [0.000,0.010,0.070,2.000] 99.51% c.09 [0.000,0.010,0.070,2.000] 99.51% c.10 [0.000,0.074,0.192,2.000] 96.31% c.11 [0.000,0.020,0.099,2.000] 99.02% c.12 [0.000,0.565,0.532,2.000] 71.74% .... [0.000,1.008,0.948,2.828] 64.37%

4.2 Concurrent results

4.2.1 Additive logistic regression

"J. Friedman, T. Hastie, R. Tibshirani (1998). Additive Logistic
Regression: a Statistical View of Boosting. Stanford University."

Predictive accuracy: 74.12%
Source:
https://www.openml.org/f/77 https://projecteuclid.org/euclid.aos/1016218223

4.2.2 Ensemble classifier, support vector machine and random forest with random sampling

"Arrhythmia Disease Classification and Mobile Based System De-
sign, Soha Samir AbdElMoneem, Hany Hanafy Said, and Amani
Anwar Saad, Computer Engineering Department, Arab Academy
for Science, Technology and Maritime Transport Abukir, Alexan-
dria, Egypt"

Predictive accuracy: 98.18%
Source:
https://iopscience.iop.org/article/10.1088/1742-6596/1447/1/012014/pdf

4.2.3 Support vector machine

"Multiclass Classification of Cardiac Arrhythmia Using Improved
Feature Selection and SVM Invariants Anam Mustaqeem, Syed Muham-
mad Anwar, and Muahammad Majid, University of Engineering and
Technology, Taxila, Pakistan"

Predictive accuracy: 92.07%
Source:
https://alendar.google.com/calendar/r/customday/2020/7/14?pli=1

5 Agaricus Lepiota Data Set

Source:
https://archive.ics.uci.edu/ml/datasets/mushroom

5.1 NeuraMorph results

5.2 Concurrent results

5.2.1 Additive logistic regression

"J. Friedman, T. Hastie, R. Tibshirani (1998). Additive Logistic
Regression: a Statistical View of Boosting. Stanford University."

Predictive accuracy: 100.0%
Source:
https://www.openml.org/f/77 https://projecteuclid.org/euclid.aos/1016218223

5.2.2 Decision tree

"Agaricus Lepiota Analysis, Manikanda Prabhu M and Jerin Stanley
Daniel J"

Predictive accuracy: 99.0%
Source:
https://rpubs.com/ManikandaPrabhuM/JerinDaniel-agaricuslepiota

6 Solar Flare Data Set

Source:

https://www.openml.org/data/get_csv/3597/dataset_116_solar-flare_2.arff

6.1 NeuraMorph results

Training parameters:

Nb inputs/outputs	24/3
Nb samples	1066
Percentage of training samples	90%
weakUnitThreshold	0.800000
depth	2
maxLvlDiv	2
nbMaxInputsUnit	2
nbMaxUnitDepth	3
order	2
Time training	1.0s

Results:

Bias prediction (min/avg/sigma/max) and accuracy	c.00 [0.017,0.587,0.915,5.759] 75.47%
	c.01 [0.001,0.088,0.138,1.959] 94.34%
	c.02 [0.000,0.018,0.000,1.009] 99.06%
	.... [0.036,0.636,0.851,5.759] 89.62%
Bias training (min/avg/sigma/max) and accuracy	c.00 [0.001,0.370,0.517,7.595] 83.02%
	c.01 [0.000,0.070,0.188,4.376] 96.15%
	c.02 [0.000,0.011,0.032,1.852] 99.58%
	.... [0.001,0.400,0.553,7.595] 92.92%

6.2 Concurrent results

6.2.1 Additive logistic regression

"J. Friedman, T. Hastie, R. Tibshirani (1998). Additive Logistic Regression: a Statistical View of Boosting. Stanford University."

Predictive accuracy: 99.15%
Source:
https://www.openml.org/f/77 https://projecteuclid.org/euclid.aos/1016218223

7

Source:
https://

7.1 NeuraMorph results

7.2 Concurrent results

7.2.1 Additive logistic regression

""

Source:
http://

7.3 Validation result

NeuraMorph performs as well as the concurrent algorithm(s) on this data set.