Robust Training of Radial Basis Function Neural Networks

Jan Kalina & Petra Vidnerová

Introduction

- Nonlinear regression
- Standard training of common types of neural networks may be heavily influenced by contamination (outliers)
- Minimal sum of squared residuals

Robust training based on backward instance selection

The dataset is divided to

1. A smaller set of outliers,
2. A larger set of the remaining (good, consistent) instances.

- Automated computation
- Start with $\lfloor n/2 \rfloor$ observations
- Gradual (sequential) adding observations to a selected subset
- Add the observation with the smallest absolute residual
- Finding the suitable number of outliers: use a quantile (assuming normal errors of the good instances)
- Outliers are deleted

Methods

- RBF = radial basis function network
- RBF-Z = RBF network with a simple outlier detection by means of Z-scores
- RRBF = novel robust RBF network

- MLP = multilayer perceptron
- MLP-Z = MLP with a simple outlier detection by means of Z-scores
- RMLP = novel robust MLP

Two benchmarking (real) datasets

- Mean square eror (MSE)
- (Robust) trimmed MSE (TMSE)

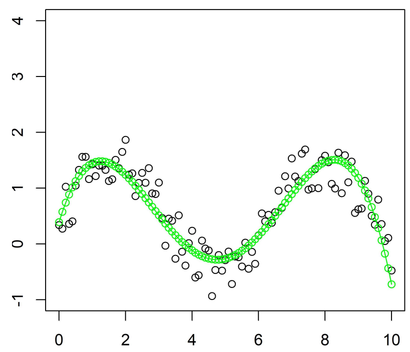
$$\mathrm{TMSE} = \frac{1}{h}\sum_{i=1}^{h} r_{(i)}^2, \quad \text{where } h = \lfloor 3n/4 \rfloor$$

- $r_1, \ldots, r_n$ are residuals
- $r_{(1)}^2 \leq \cdots r_{(n)}^2$

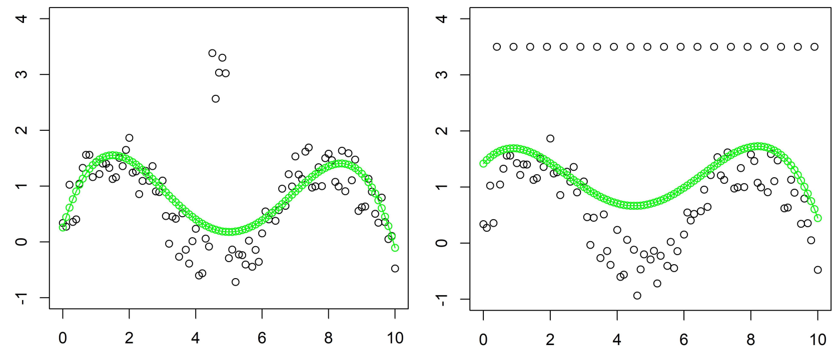
	Dataset			
	Auto MPG		Boston Housing	
Neural	($p = 4$, $n = 392$)		($p = 11$, $n = 506$)	
network	MSE	TMSE	MSE	TMSE
RBF	46.9	17.2	52.7	4.4
RBF-Z	49.1	16.6	56.5	4.3
RRBF	51.0	**13.3**	59.7	**3.9**
MLP	60.8	28.9	57.9	5.3
MLP-Z	68.2	23.6	63.4	5.0
RMLP	72.8	**15.0**	65.1	**4.3**

Simulated data (raw or contaminated)

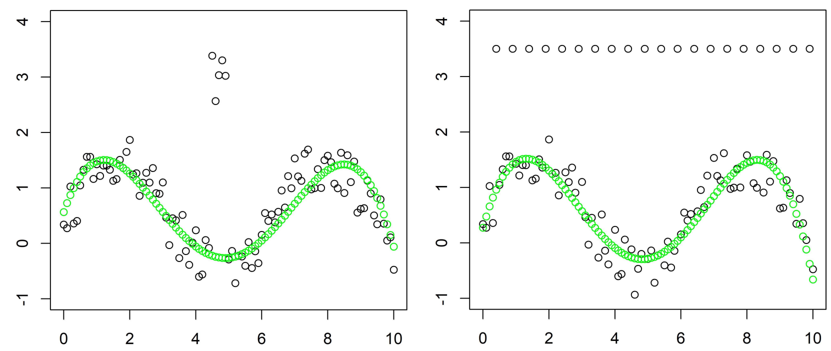
Standard RBF network with 10 RBF units and Gaussian kernel:



Standard RBF network with 10 RBF units and Gaussian kernel:



Novel RRBF network with 10 RBF units and Gaussian kernel:



Conclusions

- Standard training of MLPs and RBF networks vulnerable to outliers
- Novel approach: backward instance selection
- Suitable for various types of contamination (also under no contamination)
- Comprehensible
- Simulated and real data: remarkable improvement compared to standard training
- Computational complexity

References

[1] Cerioli, A., Riani, M., Atkinson, A.C., Corbellini, A.: The power of monitoring: How to make the most of a contaminated multivariate sample. Stat. Methods Appl. **27**, 559–587 (2018)

[2] Kordos, M., Rusiecki, A.: Reducing noise impact on MLP training—Techniques and algorithms to provide noise-robustness in MLP network training. Soft. Comput. **20**, 46–65 (2016)

The Czech Academy of Sciences, Prague petra@cs.cas.cz