On Robust Training of Regression Neural Networks

Jan Kalina, Petra Vidnerová

The Czech Academy of Sciences, Institute of Computer Science

Introduction

- Neural networks are commonly used for models with a large number of variables (or even for functional data)
- Here: **nonlinear regression** of unknown form
- Model $Y_i = \varphi(X_i) + e_i$ for $i = 1, \dots, n$, where $e_1, \dots, e_n$ are mutually independent and $X_i \in \mathbb{R}^p$
- Standard training of common types of neural networks minimize the sum of squared residuals
- Thus, they may be heavily influenced by contamination (**outliers**)
- Multi-layer perceptron neural networks (MLP)
- Radial basis function networks (RBF)
- Back-MLP, Back-RBF: available robust versions, based on backward subsample selection (outlier elimination) [1]

Neural networks with a robust loss function

- $\theta \in \mathbb{R}^Q$ = the vector of parameters of the neural network
- $\hat{\theta} \in \mathbb{R}^Q$ = estimate of θ given by the neural network
- $u_i(\hat{\theta})$ = residual for $i = 1, \dots, n$
- $u^2_{(1)}(\hat{\theta}) \leq \dots \leq u^2_{(n)}(\hat{\theta})$
- $|u(\hat{\theta})|_{(1)} \leq \dots \leq |u(\hat{\theta})|_{(n)}$
- Regularization parameter $\lambda > 0$ (found by cross validation)

LTS-MLP, LTS-RBF
Neural networks with the loss function of the least trimmed squares (LTS) defined for a fixed h ($n/2 \leq h < n$):

$$\arg\min_{\hat{\theta} \in \mathbb{R}^N} \left\{ \sum_{i=1}^{h} u^2_{(i)}(\hat{\theta}) + \lambda \sum_{j=1}^{Q} |\hat{\theta}_i| \right\}$$

LTA-MLP, LTA-RBF
Neural networks with the loss function of least trimmed absolute error (LTA):

$$\arg\min_{\hat{\theta} \in \mathbb{R}^N} \left\{ \sum_{i=1}^{h} |u(\hat{\theta})|_{(i)} + \lambda \sum_{j=1}^{Q} |\hat{\theta}_i| \right\}$$

Contact Information

Email: petra@cs.cas.cz, kalina@cs.cas.cz

Interquantile neural networks

Quantile regression neural networks: QMLP(τ), QRBF(τ)

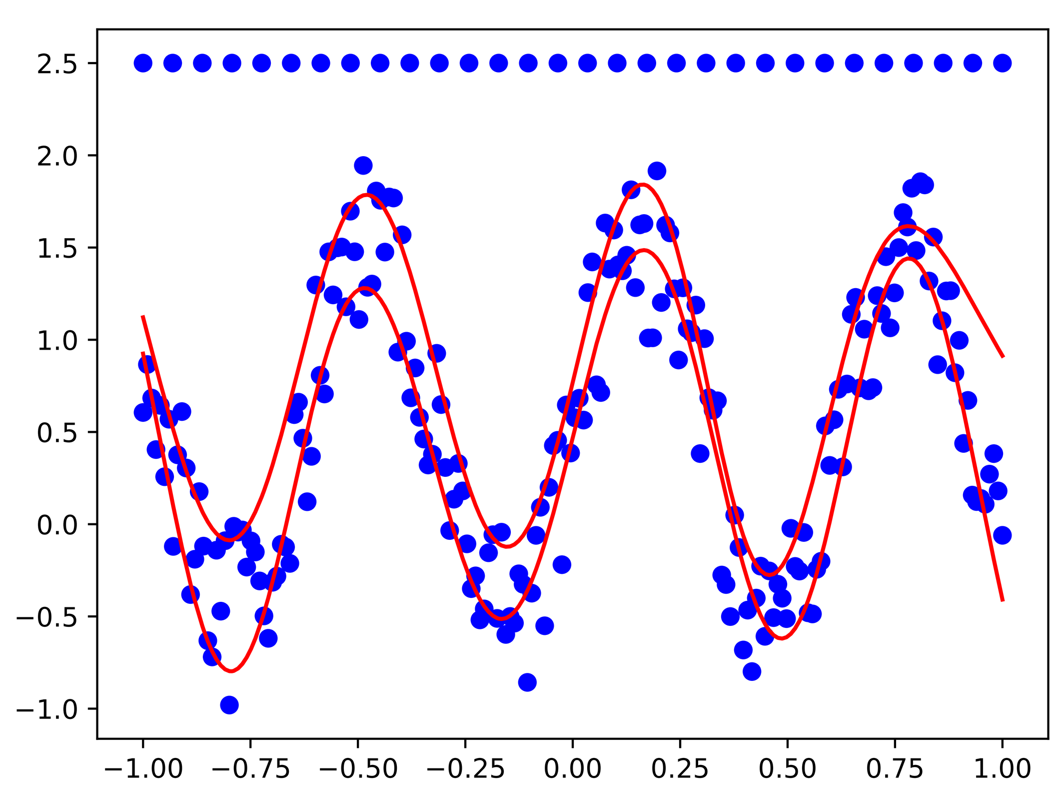
$$\arg\min_{\hat{\theta} \in \mathbb{R}^Q} \sum_{i=1}^{n} \rho_\tau \left(u_i(\hat{\theta}) \right),$$

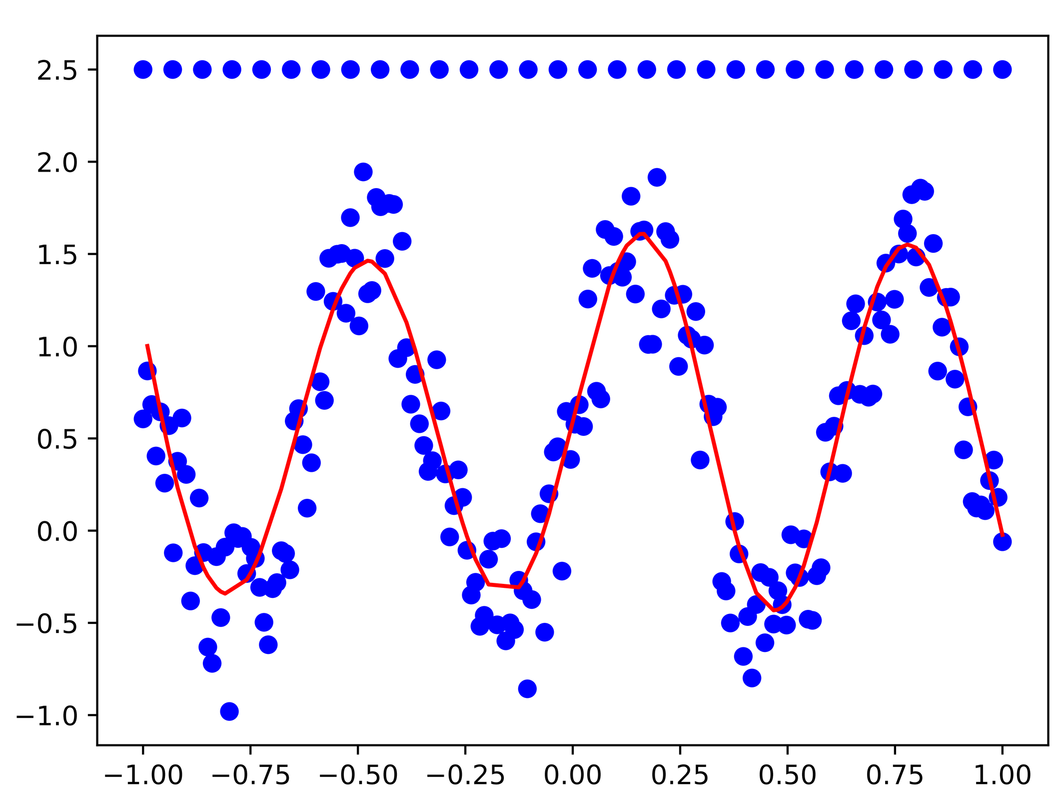
where $\rho_\tau(x) = x\,(\tau - \mathbb{1}[x < 0])$ for $x \in \mathbb{R}$, $\tau \in (0, 1)$

Interquantile neural networks (IQ-MLP, IQ-RBF)

- Choose suitable constants $\tau_1 \in (0,1)$ and $\tau_2 \in (0,1)$
- Fit a standard RBF network only for such measurements, for which
$$\hat{Y}_i^{\mathrm{QRBF}(\tau_1)} \leq Y_i \leq \hat{Y}_i^{\mathrm{QRBF}(\tau_2)}$$
where $\hat{Y}_i^{\mathrm{QRBF}(\tau)}$ are fitted values of QRBF(τ)

Illustrative example – IQ-RBF (with $p = 1$)





The method is robust to (severe) data contamination.

Benchmark (real) datasets

	Dataset		
	TTCI	Boston housing	Auto MPG
Method	MSE/TMSE	MSE/TMSE	MSE/TMSE
	Versions of MLP		
MLP	0.41/0.14	57.9/5.3	60.8/28.9
Back-MLP [1]	0.44/0.12	65.1/4.3	72.8/15.0
LTS-MLP	0.43/0.12	67.2/4.5	69.4/14.3
LTA-MLP	0.43/0.12	66.8/4.5	69.6/14.1
IQ-MLP	0.44/0.12	67.7/4.2	70.1/13.8
	Versions of RBF network		
RBF	0.39/0.14	52.7/4.4	46.9/17.2
Back-RBF [1]	0.43/0.12	59.7/3.9	51.0/13.3
LTS-RBF	0.45/0.12	60.3/4.1	52.7/12.9
LTA-RBF	0.45/0.12	61.1/4.1	53.2/12.7
IQ-RBF	0.44/**0.11**	60.8/**3.7**	52.3/**12.2**

- MSE = Mean square error
- TMSE = Trimmed MSE (robust)
- The novel methods: Remarkable improvement compared to standard training

Conclusions

- Several novel methods [2] turn out to be suitable for various types of contamination (also under no contamination)
- The interquantile approach seems to be the most promising (also for heteroscedastic models)
- Future work: Hard trimming may be replaced by implicit weights, metalearning may propose suitable weights [3]

References

[1] Kalina J., Vidnerová P. (2019): Robust training of radial basis function neural networks. Proceedings 18th International Conference ICAISC 2019. Springer, Cham, 113–124.

[2] Kalina J., Vidnerová P. (2020): On robust training of regression neural networks. Functional and High-Dimensional Statistics and Related Fields (IWFOS 2020). Springer, Cham, 145–152.

[3] Kalina J., Neoral A., Vidnerová P. (2021): Effective automatic method selection for nonlinear regression modelling. International Journal of Neural Systems. Online first.