Extreme Learning Machines for Variance-Based Global Sensitivity Analysis

John Darges, Alen Alexanderian, and Pierre Gremaud

Department of Mathematics
North Carolina State University

March 16, 2022

Supported by NSF through awards DMS-1745654 and DMS 1953271.

Motivations

- Consider a model $y = f(\mathbf{x})$ where $y \in \mathbb{R}$, and $\mathbf{x} \in [0,1]^d$ has independent uniformly distributed entries
- Sobol' indices are invaluable tools for GSA:

$$S_k = \frac{\mathrm{var}[f_k(x_k)]}{\mathrm{var}[f(\mathbf{x})]}, \quad S_k^{\mathrm{tot}} = 1 - \frac{\mathrm{var}[f_{-k}(\mathbf{x}_{-k})]}{\mathrm{var}[f(\mathbf{x})]}$$

- Approximation using Monte Carlo (MC) methods is **intractable when f is expensive to evaluate**
- MC can be avoided using surrogates with analytically known Sobol' indices (e.g. polynomial chaos, Gaussian processes)
- Can neural networks work as surrogates with analytic formulas for Sobol' indices?

Motivations

- Consider a model $y = f(\mathbf{x})$ where $y \in \mathbb{R}$, and $\mathbf{x} \in [0,1]^d$ has independent uniformly distributed entries
- Sobol' indices are invaluable tools for GSA:

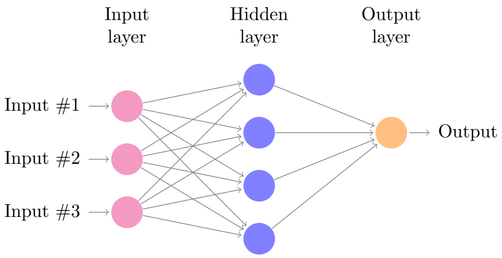
$$S_k = \frac{\mathrm{var}[f_k(x_k)]}{\mathrm{var}[f(\mathbf{x})]}, \quad S_k^{\mathrm{tot}} = 1 - \frac{\mathrm{var}[f_{-k}(\mathbf{x}_{-k})]}{\mathrm{var}[f(\mathbf{x})]}$$

- Approximation using Monte Carlo (MC) methods is **intractable when f is expensive to evaluate**
- MC can be avoided using surrogates with analytically known Sobol' indices (e.g. polynomial chaos, Gaussian processes)
- Can neural networks work as surrogates with analytic formulas for Sobol' indices? YES

Extreme Learning Machines

An ELM has the form $\hat{f}(\mathbf{x}) = \boldsymbol{\beta}^T \left(\phi\left(\mathbf{W}\mathbf{x} + \mathbf{b}\right)\right)$

$\mathbf{W}$ - inner layer weight matrix
$\mathbf{b}$ - inner layer biases
$\boldsymbol{\beta}$ - output weights
ϕ - activation function



- $\mathbf{W}, \mathbf{b}$ independently sampled randomly (e.g. from standard normal distribution)
- Solve the L_2 regularized linear least squares problem to find output weights

$$\underset{\boldsymbol{\beta}}{\arg\min}\ \frac{1}{2}\|\mathbf{H}\boldsymbol{\beta} - \mathbf{y}\|_2^2 + \frac{\alpha}{2}\|\boldsymbol{\beta}\|_2^2$$

- Regularization parameter α determined by L-curve method or generalized cross validation

Variance-based GSA with ELMs

- Integration of ELM surrogate should be easy if we want formulas for Sobol' indices
- Common machine learning activation functions (e.g. sigmoid, ReLU) do not make integration easy
- Theory tells us our activation function can be any smooth non-polynomial function
- Setting $\phi(t) = e^t$, we can derive analytic formulas for Sobol' indices in terms of $\mathbf{b}, \mathbf{W}$, and β
- After training ELM, we can obtain Sobol' indices for free:

$$S(\hat{f}) = S(\mathbf{b}, \mathbf{W}, \beta), \quad S^{\text{tot}}(\hat{f}) = S^{\text{tot}}(\mathbf{b}, \mathbf{W}, \beta)$$

Genetic Oscillator

- Biochemical model describing circadian rhythm regulation
- ODE system (expensive to solve)
- 16 reaction rate parameters are uncertain
- Each parameter uniformly distributed in interval ± 5 of respective nominal value
- Study average concentration in time of species R as QoI:

$$f(\mathbf{x}) = \frac{1}{T}\int_0^T R(t;\mathbf{x})\,dt$$

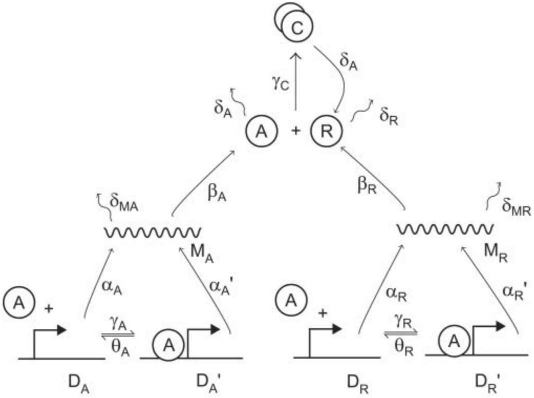


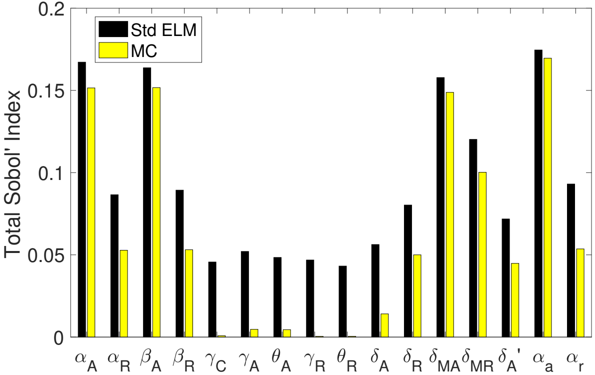
Image credit[1]

[1] J.G. Vilar, H.Y. Kueh, N. Barkai, and S. Leibler. Mechanisms of noise-resistance in genetic oscillators. 2002.

GSA Using ELM Surrogate

Experimental setup

- 3000 training size
- Points sampled via Latin hypercube sampling
- 1000 neurons
- Regularization parameter $\alpha = 10^{-4}$ from L-curve method



ELM surrogate overestimates total indices compared to MC[2]

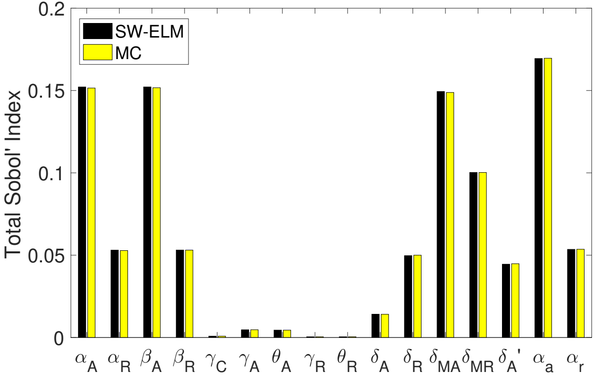
[2] M. Merritt, A. Alexanderian, and P.A. Gremaud. Multiscale global sensitivity analysis for stochastic

Sparse-Weight ELM

- **Issue:** ELM may overestimate the influence of higher order ANOVA terms
- **Idea:** We can reduce influence of higher order terms by making inner weight matrix sparse
- Sparse weight matrix $\mathbf{W}_s = \mathbf{B} \circ \mathbf{W}$, where

$$B_{ij} = \begin{cases} 0 & \text{with probability } p, \\ 1 & \text{with probability } 1-p \end{cases},$$

- Choose p by testing which value gives the best surrogate error on a validation set



Note: SW-ELM performs well with FAR fewer training points

Summary

- We use ELM as a quick and easy tool for variance-based GSA
- With exponential activation function, we derive analytic expressions of Sobol' indices for uniformly and normally distributed inputs
- After training surrogate, we obtain Sobol' indices for no additional cost
- We developed sparse weight ELM to improve GSA performance without sacrificing ELM's speed and simplicity

References

[1] J. Darges, A. Alexanderian, and P. Gremaud. Extreme learning machines for variance-based global sensitivity analysis [preprint]. 01 2022. URL https://arxiv.org/abs/2201.05586.

[2] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew. Extreme learning machine: Theory and applications. *Neurocomputing*, 70(1):489–501, 2006. ISSN 0925-2312. URL https://doi.org/10.1016/j.neucom.2005.12.126. Neural Networks.

[3] M. Merritt, A. Alexanderian, and P. A. Gremaud. Multiscale global sensitivity analysis for stochastic chemical systems. *Multiscale Modeling & Simulation*, 19(1):440–459, 2021. doi: 10.1137/20M1323989. URL https://doi.org/10.1137/20M1323989.

[4] J. M. Vilar, H. Y. Kueh, N. Barkai, and S. Leibler. Mechanisms of noise-resistance in genetic oscillators. *Proceedings of the National Academy of Sciences*, 99(9):5988–5992, 2002.