Detection of Causal SNPs in GWAS Using Saliency of Neural Networks

Alison Durst

Faculty of Informatics and Data Science, University of Regensburg

Women in Data Science Worldwide

- Current state-of-the-art methods are not always able to detect causal single nucleotide polymorphisms (SNPs) in GWAS Data due to high correlation between SNPs.
- Saliency of Neural Networks (NN) correctly ranks signal features. However, when true number of signals is unknown, saliency is not able to distinguish signal from noise.
- Our method "local sample-weighting NN" (losawNN) increases the separation between signal and noise features in saliency maps, simplifying choosing a cutoff.

Causal SNPs in GWAS Data

- Genome-Wide Association Studies (GWAS) identify genetic variants associated with traits by computing p-values from univariate linear regression models.
- High SNP-to-SNP correlations (e.g., due to Linkage Disequilibrium) and the high dimensionality of data make identifying causal SNPs difficult.

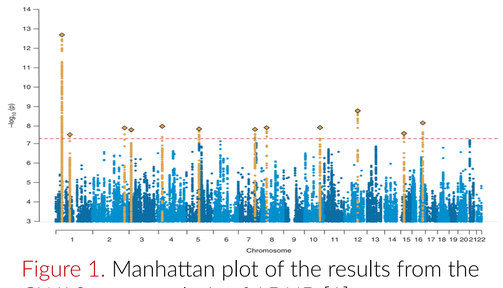


Figure 1. Manhattan plot of the results from the GWAS meta-analysis of ADHD.[1]

FineMapping on Correlated Features

- SuSiE [4] is a state-of-the-art FineMapping approach to pinpoint probable causal variants in GWAS.
- When multiple signals are correlated to a single noise feature, SuSiE fails to differentiate between signal and correlated noise features.
- SuSiE should detect feature 1 and 3 as the only signal features with same probability.
- Correlated noise feature 0 is identified as a signal feature with the same probability as true signals.

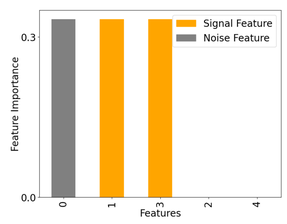


Figure 2. Feature importance of linear response model on correlated features with SuSiE.

$$y = f(x) + \varepsilon = x_1 + x_3 + \varepsilon,$$
$$\varepsilon \sim N(0, \sigma^2), \sigma^2 = \mathrm{Var}(f(X))$$

The dataset X contains 10,000 samples, each with 5 features representing a SNP (0, 1, or 2 alleles) drawn from a discrete uniform distribution satisfying the correlation matrix on the right.

	0	1	2	3	4
0	1.00	0.69	0.18	0.72	0.17
1	0.69	1.00	0.13	0.14	0.17
2	0.18	0.13	1.00	0.15	0.12
3	0.72	0.14	0.15	1.00	0.13
4	0.17	0.17	0.12	0.13	1.00

FineMapping is **not** able to distinguish signal from noise SNPs in this setting.

Saliency in low-dimensional scenario

- Saliency accurately identifies signal features in both correlated and independent settings across various response types in a scenario with 5 features.
- The presence of correlation inflates the feature importance scores of non-signal features that are correlated with true signal features.
- In the locally spiky sparse (LSS) response model with interactions, saliency scores fail to provide a clear threshold for distinguishing signal features from noise.

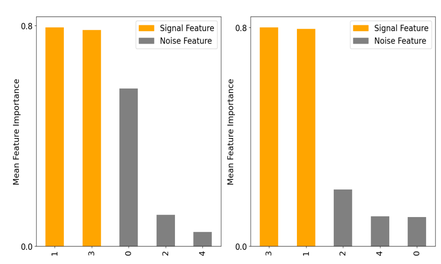


Figure 3. 5 features and LSS response on correlated (left) and independent data (right).

Saliency: Gradient-based Feature Importance of NN

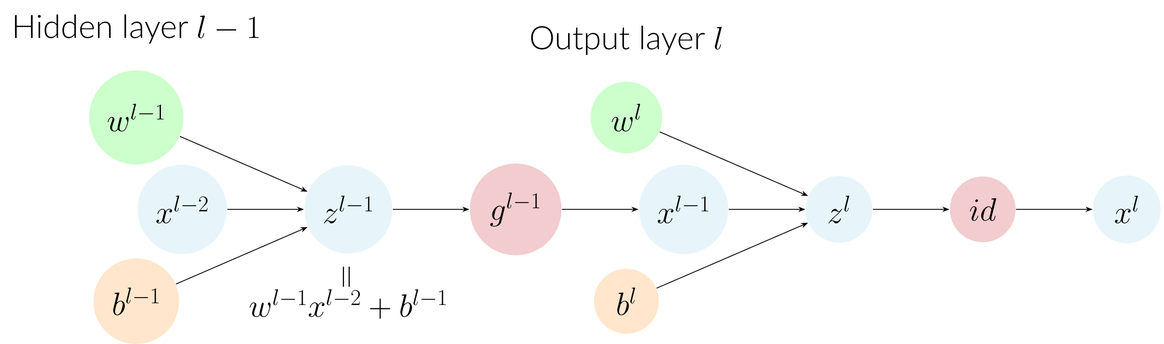
Saliency [3] is a local, gradient-based feature attribution method for NN, which highlights how input features influence prediction. To obtain a global feature importance, the mean is taken over the sample scores.

Given a trained NN $\hat{f} : \mathcal{X} \to \mathcal{Y}$, feature importance of feature $k \in \{0, \dots, P\}$ is given by

$$I_k(\hat{f}) = \frac{1}{|X_{test}|} \sum_{x \in X_{test}} \left| \frac{\partial}{\partial x_k} \hat{f}(x) \right|_{\text{min}=0}^{\text{max}=1},$$

where $|\cdot|_{\text{min}=0}^{\text{max}=1}$ denotes Min-Max scaling values to $[0, 1]$.

Using backpropagation, gradients are fast to compute, but they are not always well-defined when non-differential activation functions are used.



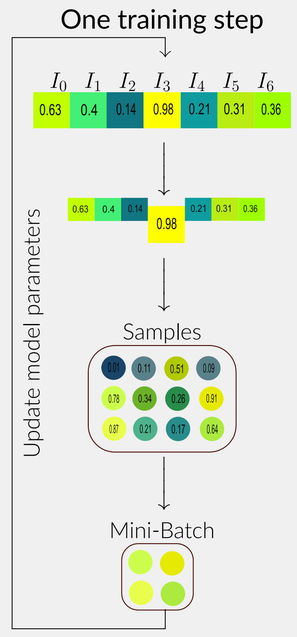
- For example, the ReLU activation function $g(x) = \max\{0, x\}$ is not differentiable in $x = 0$.
- When the activation of sample x_0 is zero in layer $(l-1)$, backpropagation breaks:

$$\frac{\partial \hat{f}}{\partial x^{l-2}}\Big|_{x_0} = \frac{\partial z^{l-1}}{\partial x^{l-2}} \frac{\partial g^{l-1}}{\partial z^{l-1}} \frac{\partial z^l}{\partial x^{l-1}} \frac{\partial id}{\partial z^l}\Big|_{x_0} = \frac{\partial z^{l-1}}{\partial x^{l-2}} \frac{\partial g^{l-1}}{\partial z^{l-1}}\Big|_0$$

Vanilla Saliency approximates gradients using $\frac{\partial g}{\partial z} = \frac{\partial g}{\partial z} \cdot \mathbf{1}(\mathbf{x} > \mathbf{0})$.

Local sample-weighting Neural Network by saliency

Idea: Use inverse probability weighting [2] on samples to locally decorrelate covariates from current most important feature in each training step of mini-batch gradient descent (GD).



Compute saliency scores of features based on current model parameters.

Draw one feature based on multinomial distribution across features with probabilities proportional to the saliency scores.

Calculate sample weights using inverse probability weighting to reduce the local correlation of features with the selected feature.

Draw a mini-batch with probabilities proportional to weights of samples.

Model Agnostic: Apply the locally decorrelated mini-batch approach to **any learner**, that can be retrained on new batches of data with **any feature importance metric**.

Results:losawNN by saliency

- **Accurate identification of signal features** by the losawNN method even under correlation, which is robust across all tested response types.
- **Enhanced separation of signal and noise** is obtained by losawNN. The margin between the mean importance scores of signal and noise features is amplified, leading to clearer separation in feature rankings.
- **Impact of correlation** is mitigated by reducing confounding influences on each mini-batch in training.
- **More reliable inference** about causal SNPs through improved clarity in feature selection.

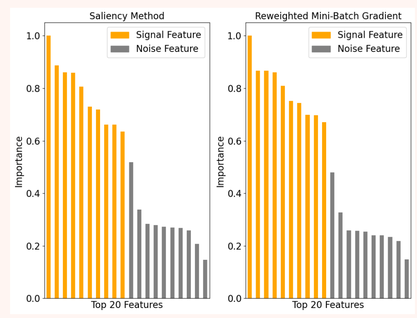


Figure 4. 3000 features and linear response with saliency (left) and decorrelated mini-batch (right).

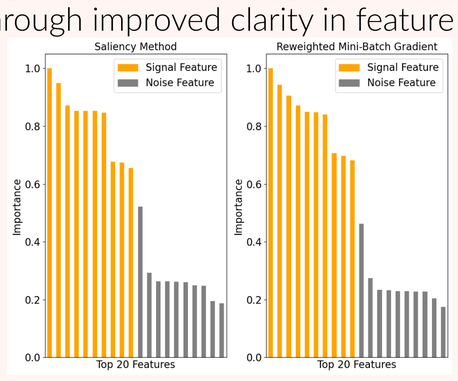


Figure 5. 3000 features and LSS response without interactions with saliency (left) and decorrelated mini-batch (right).

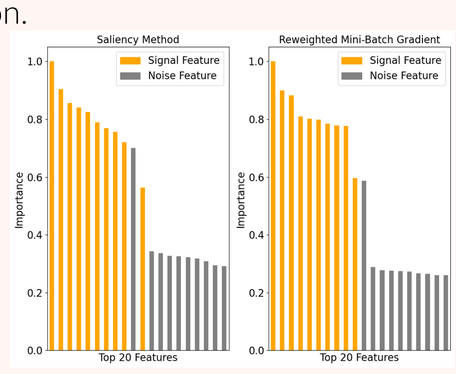


Figure 6. 3000 features and LSS response with interaction with saliency (left) and decorrelated mini-batch (right).

Methods: Simulation Setup

- We specify a 5×5 correlation matrix to define the joint distribution of each block of 5 SNP features, where $X_k \in \{0, 1, 2\}$ indicating the number of allele copies of X_k.
- We repeat this block $\frac{n_{feat}}{5}$ times to reach the desired number of features, sampling n_{obs} per block and concatenating the results, see block-matrix Σ below.
- Two settings are tested with $(n_{feat}, n_{obs}) \in \{(5, 1000), (3000, 50000)\}$.
- Independent data was generated by independently drawing samples from $\mathcal{C}\{0, 2\}$.
- The set S defines the causal features and S_i defines the set of interacting features per interaction i (e.g., $S = \{1, 3, 458, 451, 1000, 1600, 2068, 2069, 2071, 2073\}$).

Response model	$f(X)$	y
Linear Response	$\sum_{s \in S} X_s$	$f(X) + \varepsilon$
LSS Response w/o interaction	$\sum_{s \in S} \mathbf{1}(X_s \geq 2)$	$f(X) + \varepsilon$
LSS Response w/ interactions	$\sum_{i=1}^{5} \prod_{k \in S_i} \mathbf{1}(X_i \geq 1)$	$f(X) + \varepsilon$

$$\Sigma = \begin{pmatrix} & & 0 \\ & & \\ 0 & & \end{pmatrix}$$

- Fixed NN-architecture is used across all experiments.
- Each setup is repeated over 100 trials, using a different covariance matrix Σ with the same block structure.

References

[1] Ditte Demontis et al. "Discovery of the first genome-wide significant risk loci for ADHD". In: *Nature Genetics* 51 (2017), pp. 63–75. URL: https://doi.org/10.1038/s41588-018-0269-7.

[2] Judea Pearl. "An Introduction to Causal Inference". In: *The International Journal of Biostatistics* 6.2 (Feb. 2010), Article 7. DOI: 10.2202/1557-4679.1203. URL: https://www.degruyter.com/document/doi/10.2202/1557-4679.1203/html.

[3] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. "Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps". In: *arXiv preprint* (2014). arXiv: 1312.6034 [cs.CV].

[4] Gao Wang et al. "A Simple New Approach to Variable Selection in Regression, with Application to Genetic Fine Mapping". In: *Journal of the Royal Statistical Society Series B: Statistical Methodology* 82.5 (July 2020), pp. 1273–1300.