Nonparametric variable importance using an augmented neural network with multi-task learning

Jean Feng, Brian Williamson, Marco Carone, Noah Simon

ICML 2018

12 July 2018

Motivation

What is the importance of different medical tests and measurements for predicting patient survival?

$X_1 =$ heart rate

$X_2 =$ blood sugar

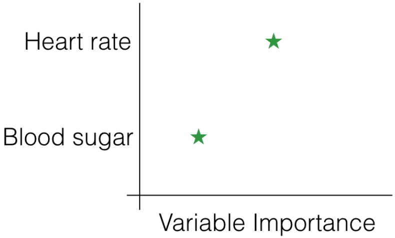
$Y =$ patient survival

Motivation

"What is the importance of different medical tests and measurements for predicting patient survival?"

We need:

- A definition of variable importance with minimal assumptions

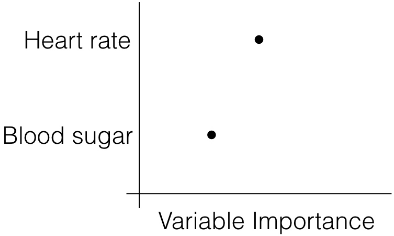


Motivation

"What is the importance of different medical tests and measurements for predicting patient survival?"

We need:

- A definition of variable importance with minimal assumptions
- A method that:
  - Estimates variable importance

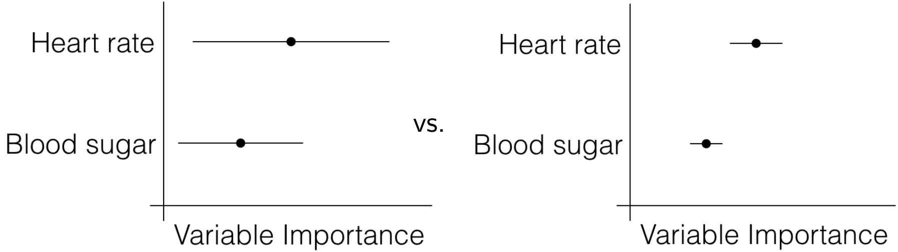


Motivation

"What is the importance of different medical tests and measurements for predicting patient survival?"

We need:

- A definition of variable importance with minimal assumptions
- A method that:
  - Estimates variable importance
  - Provides valid uncertainty assessment for our estimates, e.g., confidence intervals



Motivation

> *"What is the importance of different medical tests and measurements for predicting patient survival?"*

We need:

- A definition of variable importance with minimal assumptions
- A method that:
  - Estimates variable importance
  - Provides valid uncertainty assessment for our estimates, e.g., confidence intervals
  - Is computationally efficient

Existing variable importance methods

Method	Nonparametric estimation method	Uncertainty assessment	Computationally efficient
Existing variable importance measures in the **neural network** literature	✓	—	✓
Nonparametric variable importance using existing estimation procedures	✓	✓	—

Contribution

Statistical inference on

nonparametric variable importance

using

an **augmented** neural network

Contribution

Statistical inference on

nonparametric variable importance

using

an **augmented** neural network

✓ Uncertainty assessment

Contribution

Statistical inference on

nonparametric variable importance

using

an **augmented** neural network

✓ Uncertainty assessment

✓ Computationally efficient

Nonparametric variable importance

For random variables $(X, Y) \sim P_0$, we define the importance of

X_s: covariates X with indices in $s \subseteq \{1, 2, \ldots, p\}$ relative to

$X_{(-s)}$: covariates with indices in s removed

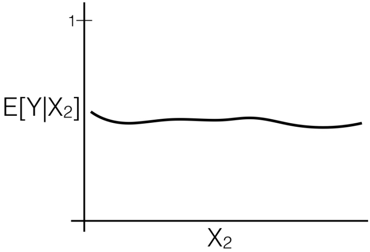
for predicting Y as

$$\psi_{0,s} := \frac{E_{P_0}(\{E_{P_0}[Y \mid X] - E_{P_0}[Y \mid X_{(-s)}]\}^2)}{\mathrm{Var}_{P_0}(Y)}.$$

Nonparametric variable importance: example

What is the importance of X_1 = heart rate vs. X_2 = blood sugar for predicting Y = patient survival?

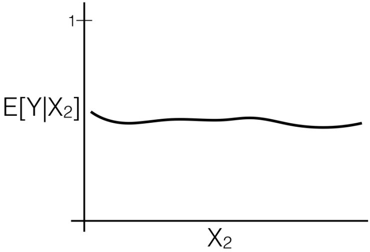
Importance of X_1:

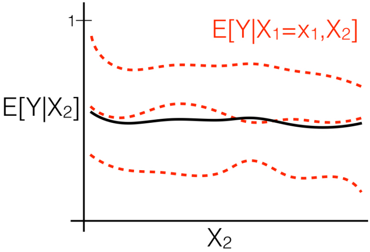


Nonparametric variable importance: example

What is the importance of X_1 = heart rate vs. X_2 = blood sugar for predicting Y = patient survival?

Importance of X_1:

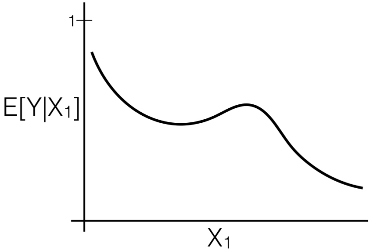




Nonparametric variable importance: example

What is the importance of X_1 = heart rate vs. X_2 = blood sugar for predicting Y = patient survival?

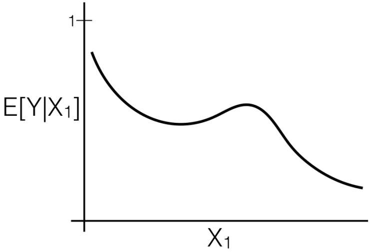
Importance of X_2:

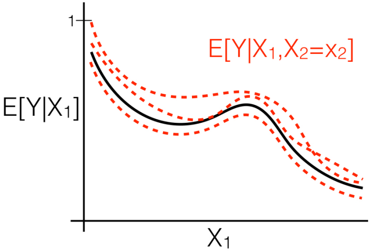


Nonparametric variable importance: example

What is the importance of X_1 = heart rate vs. X_2 = blood sugar for predicting Y = patient survival?

Importance of X_2:





Statistical Inference

Using a **population-based**, **model-agnostic** variable importance measure allows us to perform **formal statistical inference**.

Statistical Inference

Using a **population-based**, **model-agnostic** variable importance measure allows us to perform **formal statistical inference**.

We estimate $\psi_{0,s}$ via the estimator $\hat{\psi}_{n,s}$, defined as

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\{2Y_i - \hat{E}[Y \mid X_i] - \hat{E}[Y \mid X_{(-s),i}]\}\{\hat{E}[Y \mid X_i] - \hat{E}[Y \mid X_{(-s),i}]\}}{\widehat{\text{Var}}(Y)},$$

Statistical Inference

Using a **population-based**, **model-agnostic** variable importance measure allows us to perform **formal statistical inference**.

We estimate $\psi_{0,s}$ via the estimator $\hat{\psi}_{n,s}$, defined as

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\{2Y_i - \hat{E}[Y \mid X_i] - \hat{E}[Y \mid X_{(-s),i}]\}\{\hat{E}[Y \mid X_i] - \hat{E}[Y \mid X_{(-s),i}]\}}{\widehat{\mathrm{Var}}(Y)},$$

- $\hat{\psi}_{n,s}$ is an asymptotically efficient estimator.
- We can construct **confidence intervals** centered at $\hat{\psi}_{n,s}$.

Statistical inference for multiple variable groups

To do statistical inference on the importance of multiple variable groups

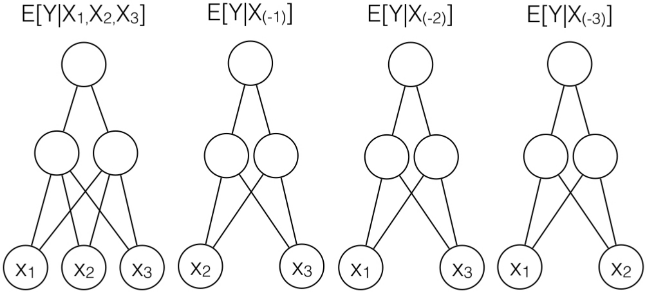
$$\mathcal{S} = \{s_i\}_{i=1}^{N} \text{ for } s_i \subseteq \{1, ..., p\},$$

we need estimates of:

- $E[Y|X]$
- $E[Y|X_{(-s)}]$ for each $s \in \mathcal{S}$

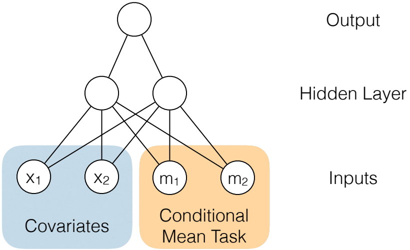
Estimating conditional means: multiple models

Existing methods estimate each of the conditional means separately, which is **computationally expensive** if $|\mathcal{S}|$ is large.



Estimating conditional means: augmented neural network

Instead we propose using a single neural network to estimate $E[Y|X]$ as well as all $E[Y|X_{(-s)}]$ for all $s \in \mathcal{S}$.



Conditional mean	Input to NN x_1 node	x_2 node	m_1 node	m_2 node
$E[Y\|X_1 = x_1, X_2 = x_2]$	x_1	x_2	0	0
$E[Y\|X_1 = x_1]$	x_1	—	0	1
$E[Y\|X_2 = x_2]$	—	x_2	1	0

Augmented neural network

We can train the augmented NN by minimizing a **multi-task loss**:

$$\text{Loss for estimating } E(Y|X) + \sum_{s \in \mathcal{S}} \text{Loss for estimating } E[Y|X_{(-s)}]$$

Computational gain: Training time is about the same as training time for a single network in the multiple network approach.

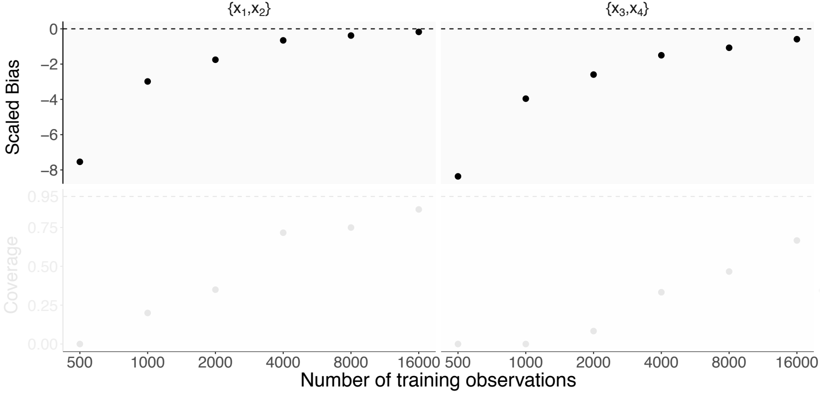
Performance gain: The augmented NN regularizes the estimates, which translates into better finite-sample performance.

Simulation study

For random variables $(Y, X_1, ..., X_6)$ with distribution

$$Y = X_1 \sin(X_1 + 2X_2)\cos(X_3 + 2X_4) + \epsilon$$

estimate importance of $\{x_1, x_2\}$ and $\{x_3, x_4\}$:

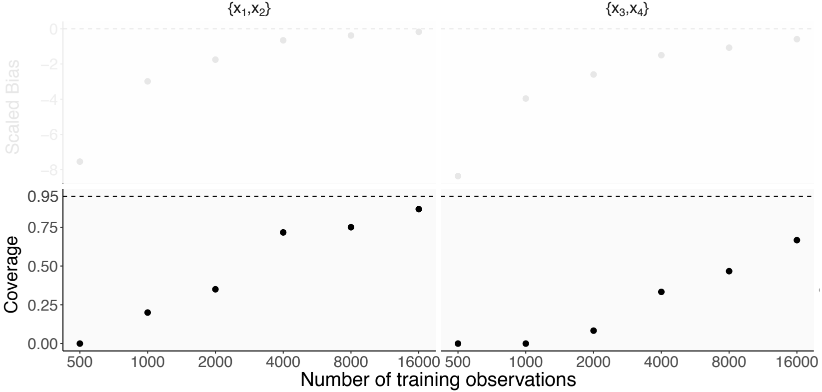


Simulation study

For random variables $(Y, X_1, ..., X_6)$ with distribution

$$Y = X_1 \sin(X_1 + 2X_2) \cos(X_3 + 2X_4) + \epsilon$$

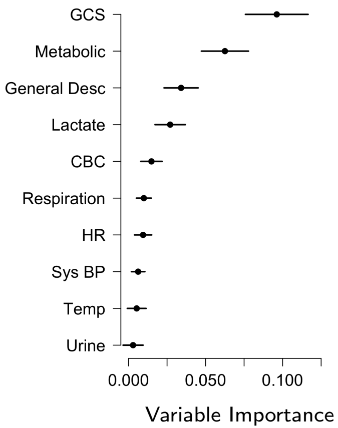
estimate importance of $\{x_1, x_2\}$ and $\{x_3, x_4\}$:



Predicting in-hospital mortality of ICU patients

PhysioNet/CinC Challenge 2012: *4000 patient records with medical test measurements over 48 hours*

Goal: understand the importance of medical tests for predicting in-hospital mortality of ICU patients



⇐ Glasgow Coma Score (GCS): measures patient consciousness, highest scoring item in SAPS II

Conclusion

- We propose estimating a **nonparametric variable importance** measure, enabling us to assess the uncertainty of our estimates.
- We show how an **augmented neural network** structure enables us to estimate conditional means quickly with improved finite-sample performance.

Thank you!

Poster #139

https://github.com/jjfeng/nnet_var_import