Discussion:

Does AI help humans make better decisions? A methodological framework for experimental evaluation.

&

Model complexity for supervised learning: why simple models almost always work best, and why it matters for applied research.

Paper 1: Does AI help humans make better decisions? A methodological framework for experimental evaluation.

Variables

Variables

In what follows all variables are binary.

Variables

In what follows all variables are binary.

A: output of an independent recommendation (e.g. cash bail vs signature bond);

Z: whether the decision maker (e.g. a judge) is exposed to the recommendation A;

D: actual decision (e.g. cash bail vs signature bond), which may differ from A

Y: outcome (e.g., recidivism)

Variables

In what follows all variables are binary.

A: output of an independent recommendation (e.g. cash bail vs signature bond);

Z: whether the decision maker (e.g. a judge) is exposed to the recommendation A;

D: actual decision (e.g. cash bail vs signature bond), which may differ from A

Y: outcome (e.g., recidivism)

Here 1 is bad (e.g, cash bail or recidivism) and 0 is good (e.g, signature bond or no recidivism).

Assumptions (identification)

Assumptions (identification)

The paper proposes a new experimental design to assess the effect of exposing the decision maker to a recommendation system on the accuracy of their decisions.

Assumptions (identification)

The paper proposes a new experimental design to assess the effect of exposing the decision maker to a recommendation system on the accuracy of their decisions.

Assumptions (identification)

The paper proposes a new experimental design to assess the effect of exposing the decision maker to a recommendation system on the accuracy of their decisions.

The design relies on three identification assumptions:

Assumptions (identification)

The paper proposes a new experimental design to assess the effect of exposing the decision maker to a recommendation system on the accuracy of their decisions.

The design relies on three identification assumptions:

Assumptions (identification)

The paper proposes a new experimental design to assess the effect of exposing the decision maker to a recommendation system on the accuracy of their decisions.

The design relies on three identification assumptions:

1. **Exclusion Restriction:** Y(z, d) = Y(d).
   That is, exposure to the recommendation only affects outcomes through the actual decision D. Enforced by **single-blidedness** (e.g. arrestee does not know judge was exposed to A).

Assumptions (identification)

The paper proposes a new experimental design to assess the effect of exposing the decision maker to a recommendation system on the accuracy of their decisions.

The design relies on three identification assumptions:

1. **Exclusion Restriction:** $Y(z, d) = Y(d)$.
   That is, exposure to the recommendation only affects outcomes through the actual decision D. Enforced by **single-blidedness** (e.g. arrestee does not know judge was exposed to A).

2. **Randomization**: $Z \perp\!\!\!\perp \{A, D(z), Y(d)\}$.
   Exposure to recommendation is randomized.

Assumptions (identification)

The paper proposes a new experimental design to assess the effect of exposing the decision maker to a recommendation system on the accuracy of their decisions.

The design relies on three identification assumptions:

1. **Exclusion Restriction:** $Y(z, d) = Y(d)$.
   That is, exposure to the recommendation only affects outcomes through the actual decision D. Enforced by **single-blidedness** (e.g. arrestee does not know judge was exposed to A).

2. **Randomization**: $Z \perp\!\!\!\perp \{A, D(z), Y(d)\}$.
   Exposure to recommendation is randomized.

3. **Consistency:** $D = D(Z)$, $Y = Y(D(Z))$.
   This assumes, e.g., no interference accross units, nor different versions of the exposure.

Assumptions (identification)

The paper proposes a new experimental design to assess the effect of exposing the decision maker to a recommendation system on the accuracy of their decisions.

The design relies on three identification assumptions:

Q: Do we also need explicit assumptions about A here?

1. **Exclusion Restriction:** Y(z, d) = Y(d).
   That is, exposure to the recommendation only affects outcomes through the actual decision D. Enforced by **single-blidedness** (e.g. arrestee does not know judge was exposed to A).

2. **Randomization**: $Z \perp\!\!\!\perp \{A, D(z), Y(d)\}$.
   Exposure to recommendation is randomized.

3. **Consistency:** D = D(Z), Y = Y(D(Z)).
   This assumes, e.g., no interference accross units, nor different versions of the exposure.

Measuring accuracy of decisions

Measuring accuracy of decisions

Consider a judge deciding on whether to grant a signature bond vs setting a cash bail.

Measuring accuracy of decisions

Consider a judge deciding on whether to grant a signature bond vs setting a cash bail.

Measuring accuracy of decisions

Consider a judge deciding on whether to grant a signature bond vs setting a cash bail.

The judge's decision is **correct** if $Y(0) = D$, namely, if the judge:

Measuring accuracy of decisions

Consider a judge deciding on whether to grant a signature bond vs setting a cash bail.

The judge's decision is **correct** if $Y(0) = D$, namely, if the judge:

Measuring accuracy of decisions

Consider a judge deciding on whether to grant a signature bond vs setting a cash bail.

The judge's decision is **correct** if $Y(0) = D$, namely, if the judge:

- grants a signature bond ($D = 0$) and the subject does not commit a new crime ($Y(0) = 0$). This outcome is observed.

Measuring accuracy of decisions

Consider a judge deciding on whether to grant a signature bond vs setting a cash bail.

The judge's decision is **correct** if $Y(0) = D$, namely, if the judge:

- grants a signature bond ($D = 0$) and the subject does not commit a new crime ($Y(0) = 0$). This outcome is observed.

- does not grant a signature bond ($D = 1$) while the subject would have commited a new crime if granted ($Y(0) = 1$). This outcome is not observed.

Measuring accuracy of decisions

Consider a judge deciding on whether to grant a signature bond vs setting a cash bail.

The judge's decision is **correct** if $Y(0) = D$, namely, if the judge:

- grants a signature bond ($D = 0$) and the subject does not commit a new crime ($Y(0) = 0$). This outcome is observed.

- does not grant a signature bond ($D = 1$) while the subject would have commited a new crime if granted ($Y(0) = 1$). This outcome is not observed.

Measuring accuracy of decisions

Consider a judge deciding on whether to grant a signature bond vs setting a cash bail.

The judge's decision is **correct** if Y(0) = D, namely, if the judge:

- grants a signature bond (D = 0) and the subject does not commit a new crime (Y(0) = 0). This outcome is observed.

- does not grant a signature bond (D = 1) while the subject would have commited a new crime if granted (Y(0) = 1). This outcome is not observed.

Conversely, the judge's decision is **incorrect** if Y(0) ≠ D, namely, if the judge:

Measuring accuracy of decisions

Consider a judge deciding on whether to grant a signature bond vs setting a cash bail.

The judge's decision is **correct** if $Y(0) = D$, namely, if the judge:

- grants a signature bond ($D = 0$) and the subject does not commit a new crime ($Y(0) = 0$). This outcome is observed.

- does not grant a signature bond ($D = 1$) while the subject would have commited a new crime if granted ($Y(0) = 1$). This outcome is not observed.

Conversely, the judge's decision is **incorrect** if $Y(0) \neq D$, namely, if the judge:

- grants a signature bond ($D = 0$) and the subject commits a new crime ($Y(0) = 1$). This outcome is observed.

Measuring accuracy of decisions

Consider a judge deciding on whether to grant a signature bond vs setting a cash bail.

The judge's decision is **correct** if $Y(0) = D$, namely, if the judge:

- grants a signature bond ($D = 0$) and the subject does not commit a new crime ($Y(0) = 0$). This outcome is observed.

- does not grant a signature bond ($D = 1$) while the subject would have commited a new crime if granted ($Y(0) = 1$). This outcome is not observed.

Conversely, the judge's decision is **incorrect** if $Y(0) \neq D$, namely, if the judge:

- grants a signature bond ($D = 0$) and the subject commits a new crime ($Y(0) = 1$). This outcome is observed.

- does not grant a signature bond ($D = 1$) while the subject would not have commited a new crime if granted ($Y(0) = 0$). This outcome is not observed.

Measuring accuracy of decisions

	D = 0 (Signature Bond)	D = 1 (Cash Bail)
Y(0) = 0 (no recidivism)	**Correct Decision.** Observed	Incorrect Decision. Unobserved.
Y(0) = 1 (recidivism)	Incorrect Decision. Observed	**Correct Decision.** Unobserved

Assigning a loss to each entry allows us to derive several measures of classification ability.

Measuring accuracy of decisions

	D = 0 (Signature Bond)	**D = 1 (Cash Bail)**
Y(0) = 0 (no recidivism)	Correct Decision. **Observed**	Incorrect Decision. Unobserved.
Y(0) = 1 (recidivism)	Incorrect Decision. **Observed**	Correct Decision. Unobserved

Assigning a loss to each entry allows us to derive several measures of classification ability.

(Partial) Identification

(Partial) Identification

We have three possible decision-making systems.

(Partial) Identification

We have three possible decision-making systems.

(Partial) Identification

We have three possible decision-making systems.

- **D(0)**: decisions **without** exposure to recommendation A;

(Partial) Identification

We have three possible decision-making systems.

- **D(0)**: decisions **without** exposure to recommendation A;
- **D(1)**: decisions **with** exposure to recommendation A;

(Partial) Identification

We have three possible decision-making systems.

- **D(0)**: decisions **without** exposure to recommendation A;
- **D(1)**: decisions **with** exposure to recommendation A;
- **D <- A**: A overrules decision D.

(Partial) Identification

We have three possible decision-making systems.

- **D(0)**: decisions **without** exposure to recommendation A;
- **D(1)**: decisions **with** exposure to recommendation A;
- **D <- A**: A overrules decision D.

(Partial) Identification

We have three possible decision-making systems.

- **D(0)**: decisions **without** exposure to recommendation A;
- **D(1)**: decisions **with** exposure to recommendation A;
- **D <- A**: A overrules decision D.

Note we observe Y(0) only when D = 0. Thus, we cannot identify the accuracy under D = 1 for any of these decision-making systems. Surprisingly, however, under the experimental design we can identify the **difference** in correct decisions when D = 1!

(Partial) Identification

We have three possible decision-making systems.

- **D(0)**: decisions **without** exposure to recommendation A;
- **D(1)**: decisions **with** exposure to recommendation A;
- **D <- A**: A overrules decision D.

Note we observe Y(0) only when D = 0. Thus, we cannot identify the accuracy under D = 1 for any of these decision-making systems. Surprisingly, however, under the experimental design we can identify the **difference** in correct decisions when D = 1!

$$P(Y(0)=1, D(0)=1) + P(Y(0)=1, D(0)=0) = P(Y(0)=1, D(1)=1) + P(Y(0)=1, D(1)=0)$$

(Partial) Identification

We have three possible decision-making systems.

- **D(0)**: decisions **without** exposure to recommendation A;
- **D(1)**: decisions **with** exposure to recommendation A;
- **D <- A**: A overrules decision D.

Note we observe Y(0) only when D = 0. Thus, we cannot identify the accuracy under D = 1 for any of these decision-making systems. Surprisingly, however, under the experimental design we can identify the **difference** in correct decisions when D = 1!

$$P(Y(0)=1, D(0)=1) + P(Y(0)=1, D(0)=0) = P(Y(0)=1, D(1)=1) + P(Y(0)=1, D(1)=0)$$

$$\implies P(Y(0)=1, D(1)=1) - P(Y(0)=1, D(0)=1) = P(Y(0)=1, D(0)=0) - P(Y(0)=1, D(1)=0)$$

(Partial) Identification

We have three possible decision-making systems.

- **D(0)**: decisions **without** exposure to recommendation A;
- **D(1)**: decisions **with** exposure to recommendation A;
- **D <- A**: A overrules decision D.

Note we observe Y(0) only when D = 0. Thus, we cannot identify the accuracy under D = 1 for any of these decision-making systems. Surprisingly, however, under the experimental design we can identify the **difference** in correct decisions when D = 1!

$$P(Y(0)=1, D(0)=1) + P(Y(0)=1, D(0)=0) = P(Y(0)=1, D(1)=1) + P(Y(0)=1, D(1)=0)$$

$$\implies P(Y(0)=1, D(1)=1) - P(Y(0)=1, D(0)=1) = P(Y(0)=1, D(0)=0) - P(Y(0)=1, D(1)=0)$$

$$\implies P(Y(0)=1, D(1)=1) - P(Y(0)=1, D(0)=1) = P(Y=1, D=0 \mid Z=0) - P(Y=1, D=0 \mid Z=1)$$

(Partial) Identification

We have three possible decision-making systems.

- **D(0)**: decisions **without** exposure to recommendation A;
- **D(1)**: decisions **with** exposure to recommendation A;
- **D <- A**: A overrules decision D.

Note we observe Y(0) only when D = 0. Thus, we cannot identify the accuracy under D = 1 for any of these decision-making systems. Surprisingly, however, under the experimental design we can identify the **difference** in correct decisions when D = 1!

$$P(Y(0)=1,D(0)=1)+P(Y(0)=1,D(0)=0)=P(Y(0)=1,D(1)=1)+P(Y(0)=1,D(1)=0)$$

$$\implies P(Y(0)=1,D(1)=1)-P(Y(0)=1,D(0)=1)=P(Y(0)=1,D(0)=0)-P(Y(0)=1,D(1)=0)$$

$$\implies P(Y(0)=1,D(1)=1)-P(Y(0)=1,D(0)=1)=P(Y=1,D=0\,|\,Z=0)-P(Y=1,D=0\,|\,Z=1)$$

First line is due to the the law of total probability; second line just rearrange terms; the third line follows from randomization, exclusion and consistency. In words, we can impute the difference in recidivism under D = 1 by the observed difference in recidivism among the released.

(Partial) Identification

(Partial) Identification

For the case where **D <- A** (that is, A overrules decision D) point identification is not possible.

The authors instead derive bounds on the difference, and the bounds can be very informative if the observed D aligns with the recommendation A.

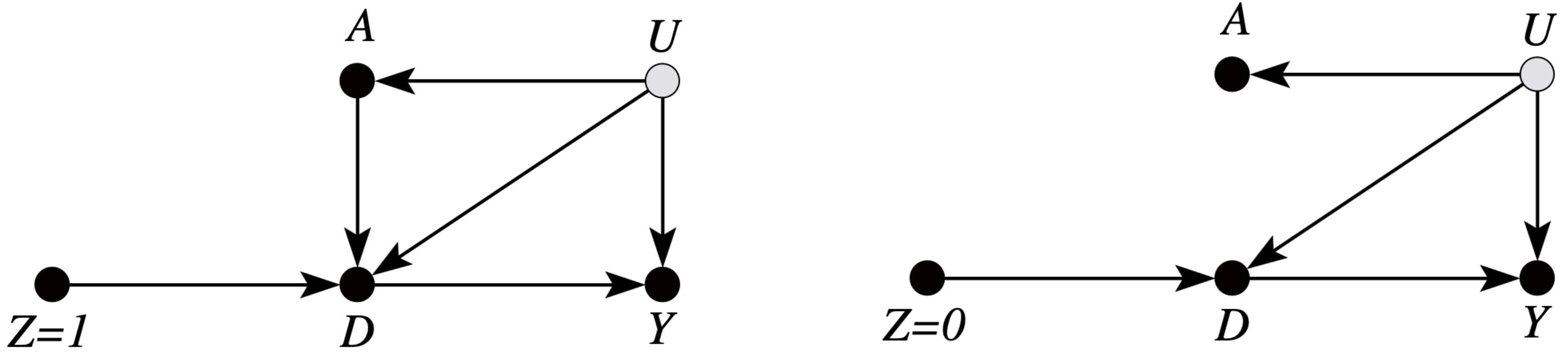
Question: it seems to me no explicit assumptions were made regarding A, except that Z is independent of A. For instance, A could be confounded with D and Y. The only additional implicit assumption seems to be that Z fully determines whether A affects D, and in its turn, D fully mediates the effect of Z on Y. What exactly is buying us informative bounds? Some intuition would be useful.

(Partial) Identification

For the case where **D <- A** (that is, A overrules decision D) point identification is not possible.

The authors instead derive bounds on the difference, and the bounds can be very informative if the observed D aligns with the recommendation A.

Question: it seems to me no explicit assumptions were made regarding A, except that Z is independent of A. For instance, A could be confounded with D and Y. The only additional implicit assumption seems to be that Z fully determines whether A affects D, and in its turn, D fully mediates the effect of Z on Y. What exactly is buying us informative bounds? Some intuition would be useful.



Empirical Analysis

Empirical Analysis

The paper illustrates the previous methods in an RCT that randomizes whether judges receive a recommendation from the Public Safety Assessment (PSA), in Dane County, Wisconsin.

Empirical Analysis

The paper illustrates the previous methods in an RCT that randomizes whether judges receive a recommendation from the Public Safety Assessment (PSA), in Dane County, Wisconsin.

Empirical Analysis

The paper illustrates the previous methods in an RCT that randomizes whether judges receive a recommendation from the Public Safety Assessment (PSA), in Dane County, Wisconsin.

Here I was very confused about what the PSA actually is, and the paper deffer this discussion to other references. I think this is important enough to have its own space on the current paper.

Empirical Analysis

The paper illustrates the previous methods in an RCT that randomizes whether judges receive a recommendation from the Public Safety Assessment (PSA), in Dane County, Wisconsin.

Here I was very confused about what the PSA actually is, and the paper deffer this discussion to other references. I think this is important enough to have its own space on the current paper.

Empirical Analysis

The paper illustrates the previous methods in an RCT that randomizes whether judges receive a recommendation from the Public Safety Assessment (PSA), in Dane County, Wisconsin.

Here I was very confused about what the PSA actually is, and the paper deffer this discussion to other references. I think this is important enough to have its own space on the current paper.

In particular, the paper calls PSA "AI recommendations". However, PSA seems to be a simple classification system where you ask binary questions about the subject (eg., is the subject younger than 23) and then you assign points to it.

Empirical Analysis

The paper illustrates the previous methods in an RCT that randomizes whether judges receive a recommendation from the Public Safety Assessment (PSA), in Dane County, Wisconsin.

Here I was very confused about what the PSA actually is, and the paper deffer this discussion to other references. I think this is important enough to have its own space on the current paper.

In particular, the paper calls PSA "AI recommendations". However, PSA seems to be a simple classification system where you ask binary questions about the subject (eg., is the subject younger than 23) and then you assign points to it.

Empirical Analysis

The paper illustrates the previous methods in an RCT that randomizes whether judges receive a recommendation from the Public Safety Assessment (PSA), in Dane County, Wisconsin.

Here I was very confused about what the PSA actually is, and the paper deffer this discussion to other references. I think this is important enough to have its own space on the current paper.

In particular, the paper calls PSA "AI recommendations". However, PSA seems to be a simple classification system where you ask binary questions about the subject (eg., is the subject younger than 23) and then you assign points to it.

This doesn't seem to be an AI in the usual meaning that people attribute to the word? Is this description really accurate?

Empirical Analysis

Empirical Analysis

New Criminal Arrest: Points		
PSA FACTOR	RESPONSE	POINTS
Age at current arrest	23 or older	0
	22 or younger	2
Pending charge at the time of the arrest	No	0
	Yes	3
Prior misdemeanor conviction	No	0
	Yes	1
Prior felony conviction	No	0
	Yes	1

Empirical Analysis

New Criminal Arrest: Points		
PSA FACTOR	**RESPONSE**	**POINTS**
Age at current arrest	23 or older	0
	22 or younger	2
Pending charge at the time of the arrest	No	0
	Yes	3
Prior misdemeanor conviction	No	0
	Yes	1
Prior felony conviction	No	0
	Yes	1

Perhaps it makes more sense to frame the empirical example in terms of assessing the quality of **PSA** recommendations instead of **AI** recommendations?

Paper 2: Model complexity for supervised learning: why simple models almost always work best, and why it matters for applied research.

Background

Background

Machine Learning (ML) algorithms (e.g, Random Forests, Deep Neural Networks) have seen a surge in interest in many social sciences.

Background

Machine Learning (ML) algorithms (e.g, Random Forests, Deep Neural Networks) have seen a surge in interest in many social sciences.

Background

Machine Learning (ML) algorithms (e.g, Random Forests, Deep Neural Networks) have seen a surge in interest in many social sciences.

For instance, recent developments in semi-parametric inference (DML, targeted learning) allows one to seamelessly integrate ML algorithms in estimation and inference of causal effect estimation.

Background

Machine Learning (ML) algorithms (e.g, Random Forests, Deep Neural Networks) have seen a surge in interest in many social sciences.

For instance, recent developments in semi-parametric inference (DML, targeted learning) allows one to seamlessly integrate ML algorithms in estimation and inference of causal effect estimation.

Background

Machine Learning (ML) algorithms (e.g, Random Forests, Deep Neural Networks) have seen a surge in interest in many social sciences.

For instance, recent developments in semi-parametric inference (DML, targeted learning) allows one to seamlessly integrate ML algorithms in estimation and inference of causal effect estimation.

Apparently, however, such methods have not gained a lot of traction in Political Science. Why? This sets background of the paper.

Background

Machine Learning (ML) algorithms (e.g, Random Forests, Deep Neural Networks) have seen a surge in interest in many social sciences.

For instance, recent developments in semi-parametric inference (DML, targeted learning) allows one to seamelessly integrate ML algorithms in estimation and inference of causal effect estimation.

Apparently, however, such methods have not gained a lot of traction in Political Science. Why? This sets background of the paper.

Background

Machine Learning (ML) algorithms (e.g, Random Forests, Deep Neural Networks) have seen a surge in interest in many social sciences.

For instance, recent developments in semi-parametric inference (DML, targeted learning) allows one to seamelessly integrate ML algorithms in estimation and inference of causal effect estimation.

Apparently, however, such methods have not gained a lot of traction in Political Science. Why? This sets background of the paper.

The paper has two main goals:

Background

Machine Learning (ML) algorithms (e.g, Random Forests, Deep Neural Networks) have seen a surge in interest in many social sciences.

For instance, recent developments in semi-parametric inference (DML, targeted learning) allows one to seamlessly integrate ML algorithms in estimation and inference of causal effect estimation.

Apparently, however, such methods have not gained a lot of traction in Political Science. Why? This sets background of the paper.

The paper has two main goals:

Background

Machine Learning (ML) algorithms (e.g, Random Forests, Deep Neural Networks) have seen a surge in interest in many social sciences.

For instance, recent developments in semi-parametric inference (DML, targeted learning) allows one to seamelessly integrate ML algorithms in estimation and inference of causal effect estimation.

Apparently, however, such methods have not gained a lot of traction in Political Science. Why? This sets background of the paper.

The paper has two main goals:

- **Description/Explanation:** it tries to understand the reasons why ML algorithms has not gained traction in Political Science. Here the ideas of *intrinsic dimension* and *data curation* play a key role.

Background

Machine Learning (ML) algorithms (e.g, Random Forests, Deep Neural Networks) have seen a surge in interest in many social sciences.

For instance, recent developments in semi-parametric inference (DML, targeted learning) allows one to seamelessly integrate ML algorithms in estimation and inference of causal effect estimation.

Apparently, however, such methods have not gained a lot of traction in Political Science. Why? This sets background of the paper.

The paper has two main goals:

- **Description/Explanation:** it tries to understand the reasons why ML algorithms has not gained traction in Political Science. Here the ideas of *intrinsic dimension* and *data curation* play a key role.

- **Prescription:** it makes recommendations regarding the use of ML algorithms, such as not using htem in lieu of OLS.

Claims

Claims

Here's my interpretation of the claims of the paper:

Claims

Here's my interpretation of the claims of the paper:

Claims

Here's my interpretation of the claims of the paper:

1. That the adoption of ML methods in Political Science has been slow, or slower than expected.

Claims

Here's my interpretation of the claims of the paper:

1. That the adoption of ML methods in Political Science has been slow, or slower than expected.

2. That the *intrinsic dimension* of Political Science datasets is low, and for that reason ML methods are a poor match for Political Science data.

Claims

Here's my interpretation of the claims of the paper:

1. That the adoption of ML methods in Political Science has been slow, or slower than expected.

2. That the *intrinsic dimension* of Political Science datasets is low, and for that reason ML methods are a poor match for Political Science data.

3. That *data curation* is one reason why Political Science datasets have low *intrinsic dimension.*

Claims

Here's my interpretation of the claims of the paper:

1. That the adoption of ML methods in Political Science has been slow, or slower than expected.

2. That the *intrinsic dimension* of Political Science datasets is low, and for that reason ML methods are a poor match for Political Science data.

3. That *data curation* is one reason why Political Science datasets have low *intrinsic dimension.*

4. That (2) and (3) may be reasons for (1).

Claims

Here's my interpretation of the claims of the paper:

1. That the adoption of ML methods in Political Science has been slow, or slower than expected.

2. That the *intrinsic dimension* of Political Science datasets is low, and for that reason ML methods are a poor match for Political Science data.

3. That *data curation* is one reason why Political Science datasets have low *intrinsic dimension.*

4. That (2) and (3) may be reasons for (1).

5. Recommendations regarding the use of ML methods (e.g. just use OLS).

Claims

Here's my interpretation of the claims of the paper:

1. That the adoption of ML methods in Political Science has been slow, or slower than expected.

2. That the *intrinsic dimension* of Political Science datasets is low, and for that reason ML methods are a poor match for Political Science data.

3. That *data curation* is one reason why Political Science datasets have low *intrinsic dimension.*

4. That (2) and (3) may be reasons for (1).

5. Recommendations regarding the use of ML methods (e.g. just use OLS).

Has the adoption of ML methods been slow?

Has the adoption of ML methods been slow?

The paper claims the adoption of ML methods in Political Science has been slow, or slower than expected.

Has the adoption of ML methods been slow?

The paper claims the adoption of ML methods in Political Science has been slow, or slower than expected.

The paper seems to outsource the above claim to references from 2004 and 2018. Though intuitively I agree with the claim, I felt that not enough evidence for this was provided. As far as I understand, these papers did not catalog the usage of ML methods in Political Science. Moreover, the references are not that recent anyomore.

Has the adoption of ML methods been slow?

The paper claims the adoption of ML methods in Political Science has been slow, or slower than expected.

The paper seems to outsource the above claim to references from 2004 and 2018. Though intuitively I agree with the claim, I felt that not enough evidence for this was provided. As far as I understand, these papers did not catalog the usage of ML methods in Political Science. Moreover, the references are not that recent anyomore.

Has the adoption of ML methods been slow?

The paper claims the adoption of ML methods in Political Science has been slow, or slower than expected.

The paper seems to outsource the above claim to references from 2004 and 2018. Though intuitively I agree with the claim, I felt that not enough evidence for this was provided. As far as I understand, these papers did not catalog the usage of ML methods in Political Science. Moreover, the references are not that recent anyomore.

Given that this is one of the main motivations for the paper, it seems it should be better justified.

Has the adoption of ML methods been slow?

The paper claims the adoption of ML methods in Political Science has been slow, or slower than expected.

The paper seems to outsource the above claim to references from 2004 and 2018. Though intuitively I agree with the claim, I felt that not enough evidence for this was provided. As far as I understand, these papers did not catalog the usage of ML methods in Political Science. Moreover, the references are not that recent anyomore.

Given that this is one of the main motivations for the paper, it seems it should be better justified.

Has the adoption of ML methods been slow?

The paper claims the adoption of ML methods in Political Science has been slow, or slower than expected.

The paper seems to outsource the above claim to references from 2004 and 2018. Though intuitively I agree with the claim, I felt that not enough evidence for this was provided. As far as I understand, these papers did not catalog the usage of ML methods in Political Science. Moreover, the references are not that recent anyomore.

Given that this is one of the main motivations for the paper, it seems it should be better justified.

Perhaps a new updated survey on the usage of ML methods would be appropriate?

Has the adoption of ML methods been slow?

The paper claims the adoption of ML methods in Political Science has been slow, or slower than expected.

The paper seems to outsource the above claim to references from 2004 and 2018. Though intuitively I agree with the claim, I felt that not enough evidence for this was provided. As far as I understand, these papers did not catalog the usage of ML methods in Political Science. Moreover, the references are not that recent anyomore.

Given that this is one of the main motivations for the paper, it seems it should be better justified.

Perhaps a new updated survey on the usage of ML methods would be appropriate?

Intrinsic dimension and the appropriateness of ML.

Intrinsic dimension and the appropriateness of ML.

The authors introduce the idea of intrinsic dimension of a prediction problem to the Political Science audience.

Intrinsic dimension and the appropriateness of ML.

The authors introduce the idea of intrinsic dimension of a prediction problem to the Political Science audience.

Intrinsic dimension and the appropriateness of ML.

The authors introduce the idea of intrinsic dimension of a prediction problem to the Political Science audience.

Briefly, the intrinsic dimension of data X to predict Y is defined as the smallest number of non-zero weights in a Deep Neural Network (DNN) such what we can still recover E[Y|X] (this is because the CEF is the best predictor in the mean squared error sense).

Intrinsic dimension and the appropriateness of ML.

The authors introduce the idea of intrinsic dimension of a prediction problem to the Political Science audience.

Briefly, the intrinsic dimension of data X to predict Y is defined as the smallest number of non-zero weights in a Deep Neural Network (DNN) such what we can still recover E[Y|X] (this is because the CEF is the best predictor in the mean squared error sense).

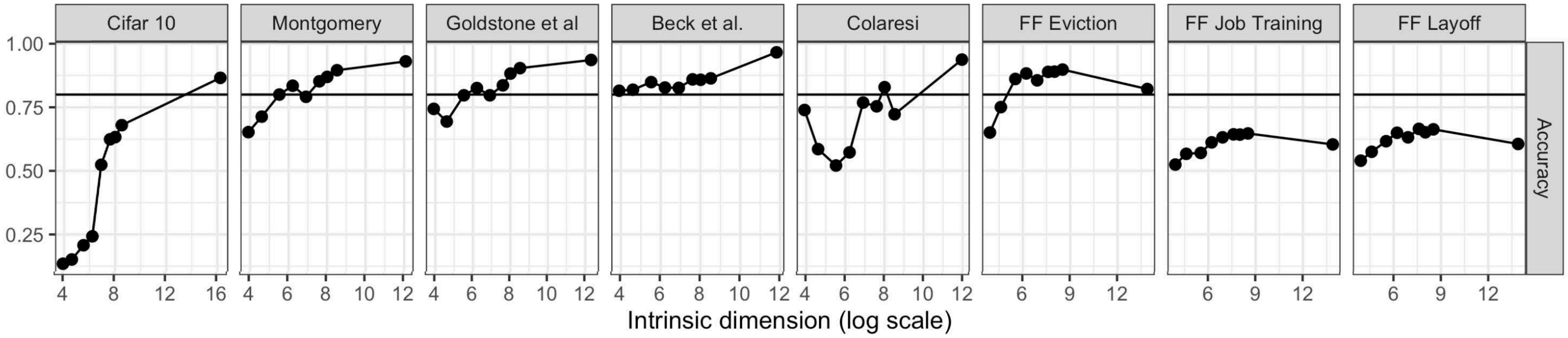
Intrinsic dimension and the appropriateness of ML.

The authors introduce the idea of intrinsic dimension of a prediction problem to the Political Science audience.

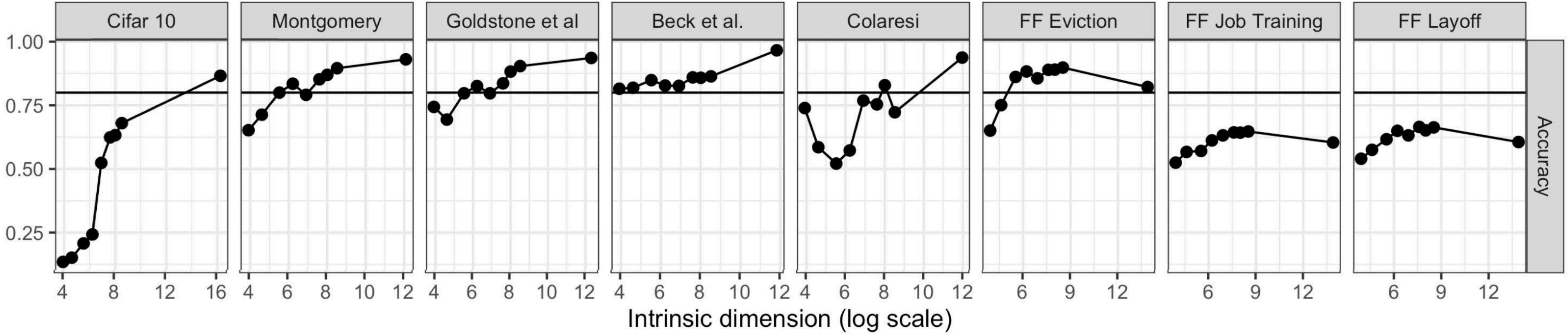
Briefly, the intrinsic dimension of data X to predict Y is defined as the smallest number of non-zero weights in a Deep Neural Network (DNN) such what we can still recover E[Y|X] (this is because the CEF is the best predictor in the mean squared error sense).

One contribution of this paper is to provide evidence that the intrinsic dimension of Political Science datasets is low, at least when compared to image data (CIFAR-10).

Intrinsic dimension and the appropriateness of ML.



Intrinsic dimension and the appropriateness of ML.



While it is true that the intrinsic dimension of CIFAR-10 is larger than the focus datasets, it does not seem obvious to me that the intrinsic dimension of Political Science datasets is low. In particular, if I understand the scales correctly, the dimensions seem to be in the order of 3,000 to 5,000, and in some cases 162,000 (Montgomery, Goldstone and Beck).

Is this really low?

Intrinsic dimension and the appropriateness of ML.

Intrinsic dimension and the appropriateness of ML.

The paper concludes that linear models should be preferred over non-linear models based on the intrinsic dimension evidence as before.

Intrinsic dimension and the appropriateness of ML.

The paper concludes that linear models should be preferred over non-linear models based on the intrinsic dimension evidence as before.

Intrinsic dimension and the appropriateness of ML.

The paper concludes that linear models should be preferred over non-linear models based on the intrinsic dimension evidence as before.

However, this claim does not seem to follow from the evidence, here are some reasons:

Intrinsic dimension and the appropriateness of ML.

The paper concludes that linear models should be preferred over non-linear models based on the intrinsic dimension evidence as before.

However, this claim does not seem to follow from the evidence, here are some reasons:

Intrinsic dimension and the appropriateness of ML.

The paper concludes that linear models should be preferred over non-linear models based on the intrinsic dimension evidence as before.

However, this claim does not seem to follow from the evidence, here are some reasons:

The absolute value of the intrinsic dimension did not seem low. At leas to me, it is not obvious that a linear model would have performed better than a properly regularized DNN or Random Forest in the focus datasets. Why not compare linear models against ML methods directly, instead of indirectly via instrisic dimension?

Intrinsic dimension and the appropriateness of ML.

The paper concludes that linear models should be preferred over non-linear models based on the intrinsic dimension evidence as before.

However, this claim does not seem to follow from the evidence, here are some reasons:

The absolute value of the intrinsic dimension did not seem low. At leas to me, it is not obvious that a linear model would have performed better than a properly regularized DNN or Random Forest in the focus datasets. Why not compare linear models against ML methods directly, instead of indirectly via instrisic dimension?

Intrinsic dimension and the appropriateness of ML.

The paper concludes that linear models should be preferred over non-linear models based on the intrinsic dimension evidence as before.

However, this claim does not seem to follow from the evidence, here are some reasons:

The absolute value of the intrinsic dimension did not seem low. At leas to me, it is not obvious that a linear model would have performed better than a properly regularized DNN or Random Forest in the focus datasets. Why not compare linear models against ML methods directly, instead of indirectly via instrisic dimension?

Moreover, it seems to me that mispecification concerns can still happen even if the intrinsic dimension is low. For example, I could have a model with low intrinsic dimension (say, E[Y|X] = X^2) and yet, if I don't know the correct specification, using linear models would still give me biased estimates (say, if I'm estimating a causal effect), while using a data adaptive algorithms would protect me against misspecification.

Intrinsic dimension and the appropriateness of ML.

The paper concludes that linear models should be preferred over non-linear models based on the intrinsic dimension evidence as before.

However, this claim does not seem to follow from the evidence, here are some reasons:

The absolute value of the intrinsic dimension did not seem low. At leas to me, it is not obvious that a linear model would have performed better than a properly regularized DNN or Random Forest in the focus datasets. Why not compare linear models against ML methods directly, instead of indirectly via instrisic dimension?

Moreover, it seems to me that mispecification concerns can still happen even if the intrinsic dimension is low. For example, I could have a model with low intrinsic dimension (say, $E[Y|X] = X^2$) and yet, if I don't know the correct specification, using linear models would still give me biased estimates (say, if I'm estimating a causal effect), while using a data adaptive algorithms would protect me against misspecification.

Intrinsic dimension and the adoption of ML.

Intrinsic dimension and the adoption of ML.

The paper conjectures that the low intrinsic dimension is a possible reason for the low adoption of ML methods in Political Science.

Intrinsic dimension and the adoption of ML.

The paper conjectures that the low intrinsic dimension is a possible reason for the low adoption of ML methods in Political Science.

Intrinsic dimension and the adoption of ML.

The paper conjectures that the low intrinsic dimension is a possible reason for the low adoption of ML methods in Political Science.

There seems to be many other alternative explanations for this low adoption!

Intrinsic dimension and the adoption of ML.

The paper conjectures that the low intrinsic dimension is a possible reason for the low adoption of ML methods in Political Science.

There seems to be many other alternative explanations for this low adoption!

Intrinsic dimension and the adoption of ML.

The paper conjectures that the low intrinsic dimension is a possible reason for the low adoption of ML methods in Political Science.

There seems to be many other alternative explanations for this low adoption!

As a developer of statistical packages, one that readily comes to mind is simply the availability of easy to use and reliable software.

Intrinsic dimension and the adoption of ML.

The paper conjectures that the low intrinsic dimension is a possible reason for the low adoption of ML methods in Political Science.

There seems to be many other alternative explanations for this low adoption!

As a developer of statistical packages, one that readily comes to mind is simply the availability of easy to use and reliable software.

Intrinsic dimension and the adoption of ML.

The paper conjectures that the low intrinsic dimension is a possible reason for the low adoption of ML methods in Political Science.

There seems to be many other alternative explanations for this low adoption!

As a developer of statistical packages, one that readily comes to mind is simply the availability of easy to use and reliable software.

For instance, if you want to use DNN for estimating causal effects using debiased machine learning, this is still non-trivial, buggy, and not as easy and reliable than to do run OLS.

Intrinsic dimension and the adoption of ML.

The paper conjectures that the low intrinsic dimension is a possible reason for the low adoption of ML methods in Political Science.

There seems to be many other alternative explanations for this low adoption!

As a developer of statistical packages, one that readily comes to mind is simply the availability of easy to use and reliable software.

For instance, if you want to use DNN for estimating causal effects using debiased machine learning, this is still non-trivial, buggy, and not as easy and reliable than to do run OLS.

Practical Recommendation: Just do OLS

Practical Recommendation: Just do OLS

As someone who loves OLS, I do love this recommendation!
(ML can be seens as just OLS anyway, with feature expansions).

Practical Recommendation: Just do OLS

As someone who loves OLS, I do love this recommendation!

(ML can be seens as just OLS anyway, with feature expansions).

Practical Recommendation: Just do OLS

As someone who loves OLS, I do love this recommendation!
(ML can be seens as just OLS anyway, with feature expansions).

But while I think one should always run a simple OLS for a baseline comparison, some aspects of the methodology were not clear.

Practical Recommendation: Just do OLS

As someone who loves OLS, I do love this recommendation!
(ML can be seens as just OLS anyway, with feature expansions).

But while I think one should always run a simple OLS for a baseline comparison, some aspects of the methodology were not clear.

Practical Recommendation: Just do OLS

As someone who loves OLS, I do love this recommendation!
(ML can be seens as just OLS anyway, with feature expansions).

But while I think one should always run a simple OLS for a baseline comparison, some aspects of the methodology were not clear.

For example, how do I map the intrinsic dimension to the decision of running OLS?

Practical Recommendation: Just do OLS

As someone who loves OLS, I do love this recommendation!
(ML can be seens as just OLS anyway, with feature expansions).

But while I think one should always run a simple OLS for a baseline comparison, some aspects of the methodology were not clear.

For example, how do I map the intrinsic dimension to the decision of running OLS?

Practical Recommendation: Just do OLS

As someone who loves OLS, I do love this recommendation!
(ML can be seens as just OLS anyway, with feature expansions).

But while I think one should always run a simple OLS for a baseline comparison, some aspects of the methodology were not clear.

For example, how do I map the intrinsic dimension to the decision of running OLS?

I don't understand how this would be superior to the recommendation of:

Practical Recommendation: Just do OLS

As someone who loves OLS, I do love this recommendation!
(ML can be seens as just OLS anyway, with feature expansions).

But while I think one should always run a simple OLS for a baseline comparison, some aspects of the methodology were not clear.

For example, how do I map the intrinsic dimension to the decision of running OLS?

I don't understand how this would be superior to the recommendation of:

Practical Recommendation: Just do OLS

As someone who loves OLS, I do love this recommendation!
(ML can be seens as just OLS anyway, with feature expansions).

But while I think one should always run a simple OLS for a baseline comparison, some aspects of the methodology were not clear.

For example, how do I map the intrinsic dimension to the decision of running OLS?

I don't understand how this would be superior to the recommendation of:

just do cross-validation.

Practical Recommendation: Just do OLS

As someone who loves OLS, I do love this recommendation!
(ML can be seens as just OLS anyway, with feature expansions).

But while I think one should always run a simple OLS for a baseline comparison, some aspects of the methodology were not clear.

For example, how do I map the intrinsic dimension to the decision of running OLS?

I don't understand how this would be superior to the recommendation of:

just do cross-validation.

Practical Recommendation: Just do OLS

As someone who loves OLS, I do love this recommendation!
(ML can be seens as just OLS anyway, with feature expansions).

But while I think one should always run a simple OLS for a baseline comparison, some aspects of the methodology were not clear.

For example, how do I map the intrinsic dimension to the decision of running OLS?

I don't understand how this would be superior to the recommendation of:

just do cross-validation.

If linear models are superior to non-linear ones in the particular dataset you are using, shouldn't proper cross-validation capture that?

Practical Recommendation: Just do OLS

As someone who loves OLS, I do love this recommendation!
(ML can be seens as just OLS anyway, with feature expansions).

But while I think one should always run a simple OLS for a baseline comparison, some aspects of the methodology were not clear.

For example, how do I map the intrinsic dimension to the decision of running OLS?

I don't understand how this would be superior to the recommendation of:

just do cross-validation.

If linear models are superior to non-linear ones in the particular dataset you are using, shouldn't proper cross-validation capture that?