A universal method of controlling for confounds in MVPA

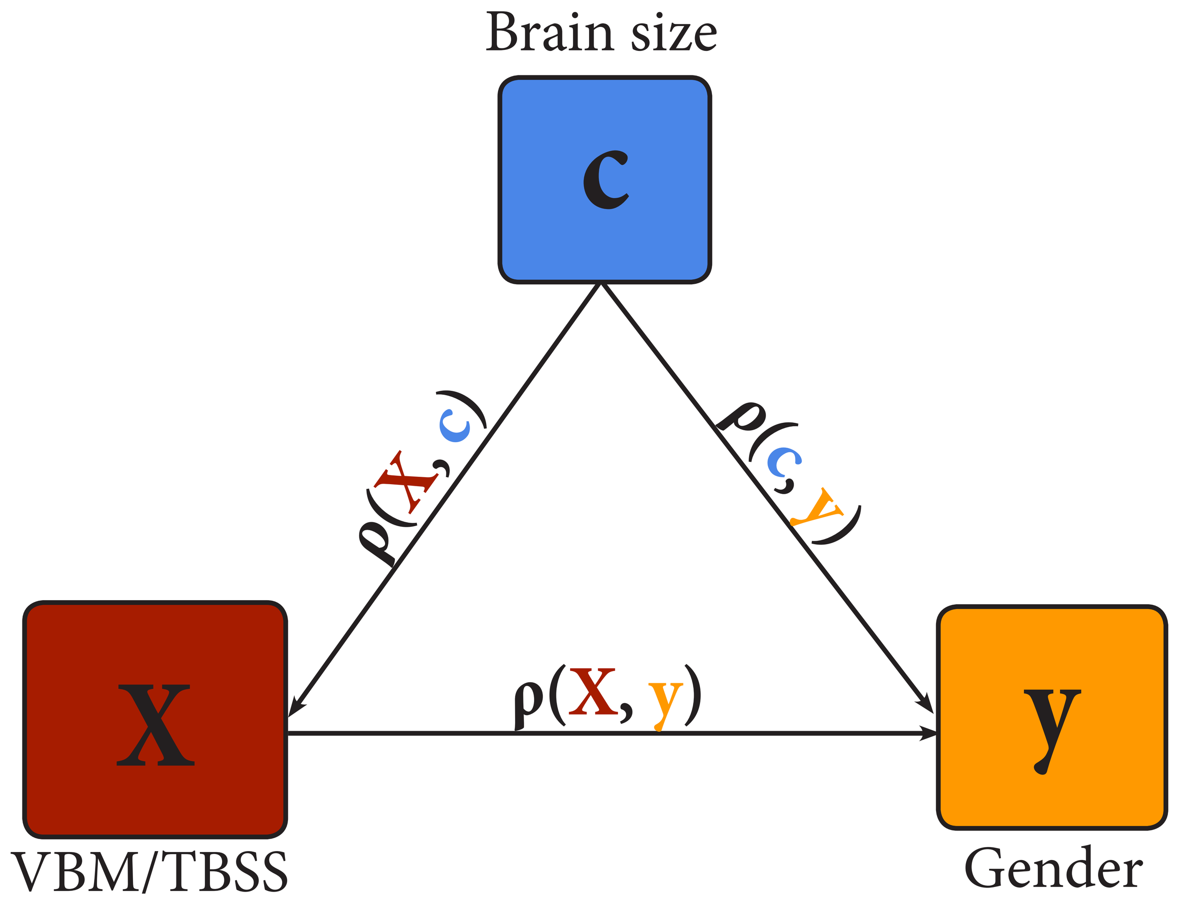
Steven Miletić, Lukas Snoek, & H. Steven Scholte
University of Amsterdam

Universiteit van Amsterdam

Spinoza Centre for Neuroimaging

Introduction

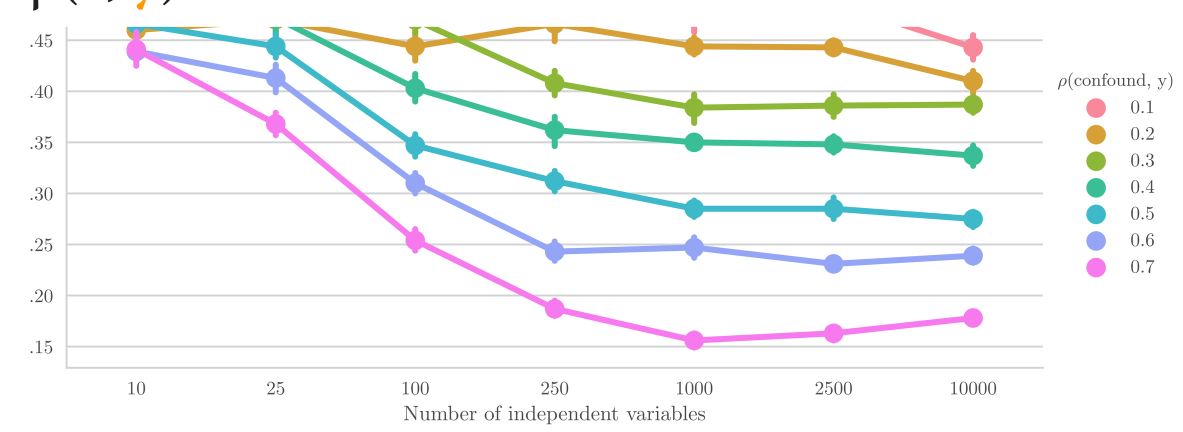
- Contrary to mass-univariate analyses, where confounds are widely controlled for, it is **unclear how to handle confounds** in MVPA
- This poses a serious threat to the **generalizability of MVPA results** in both clinical and fundamental research - especially because MVPA is arguably **more sensitive** to confounds[1]
- Here, we show how a previously proposed[2,3] method of dealing with confounds ("confound regression") leads to bias[4], and causes **below-chance accuracy**[3]
- We introduce a **universal and unbiased method** of dealing with confounds in MVPA



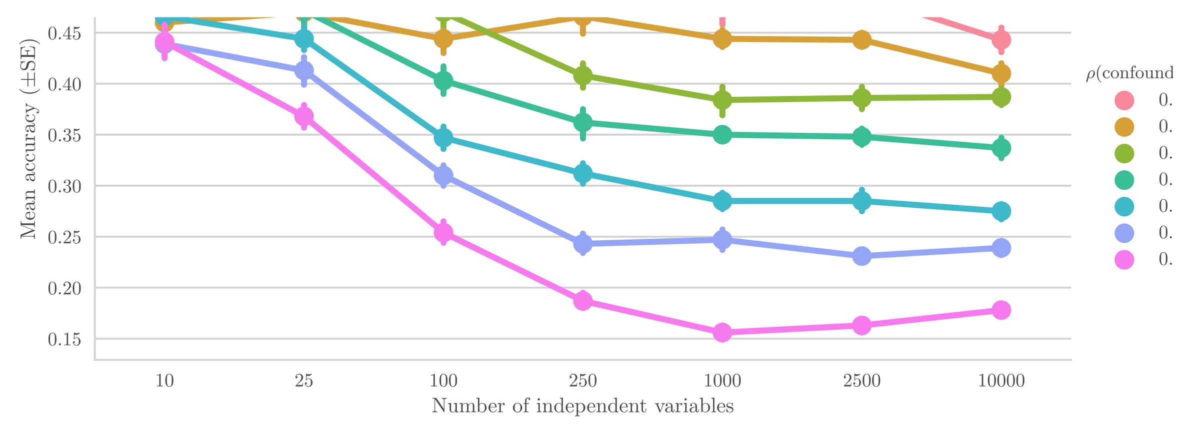
What's the problem?

Following the example to predict gender (y) from **VBM and TBSS-data (X)** in the face of the "confound" brain size (c)...

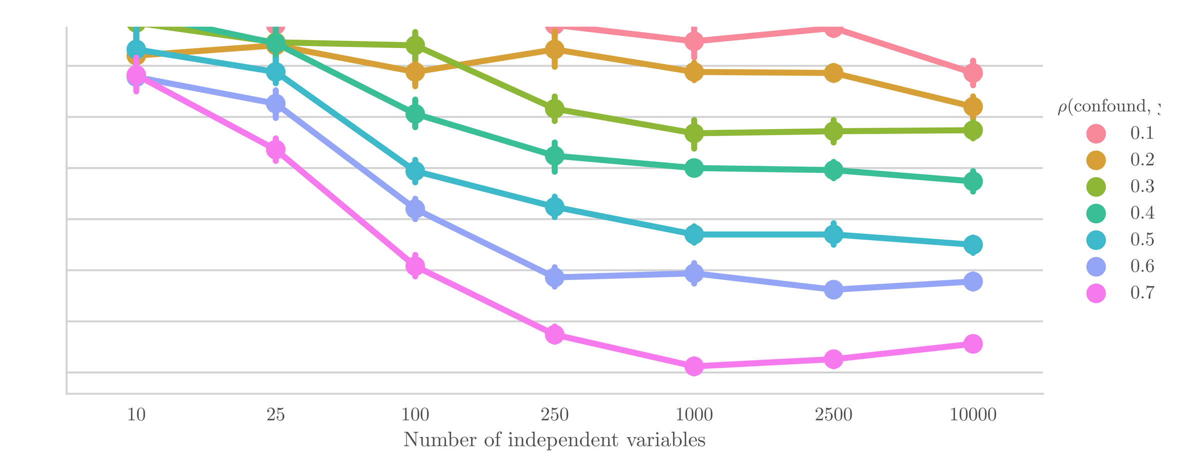
We know that brain size truly confounds ρ(c, y)...[2]



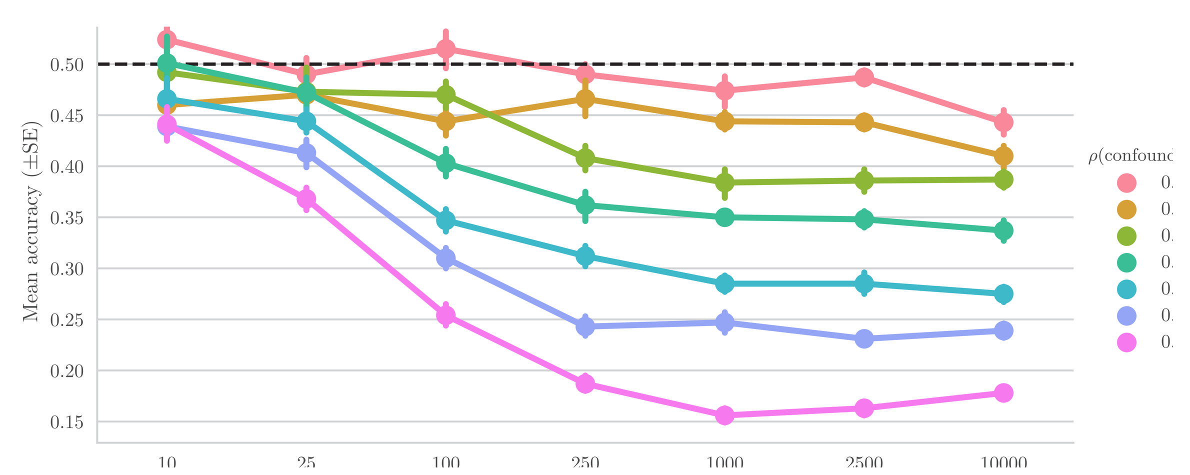
So, $\rho(X, y)_{\text{uncorrected}}$ is biased...



...and is related to our data...



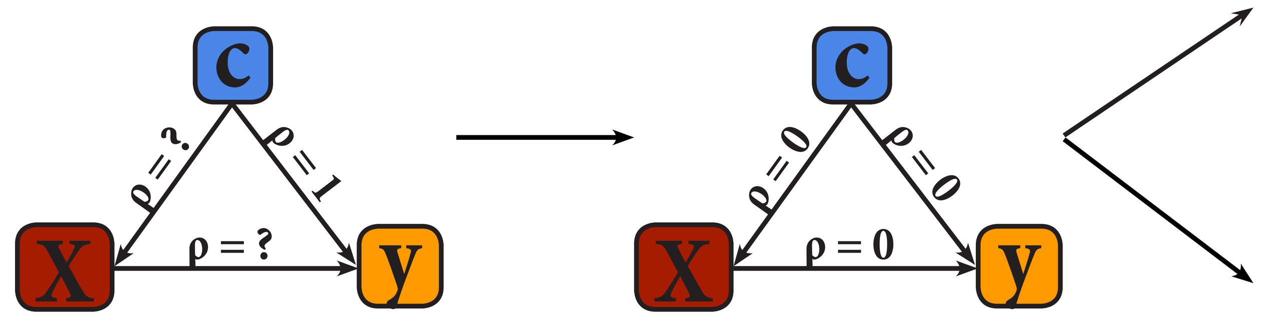
Yet, $\rho(X, y)_{\text{corrected}}$ < chance (50%)

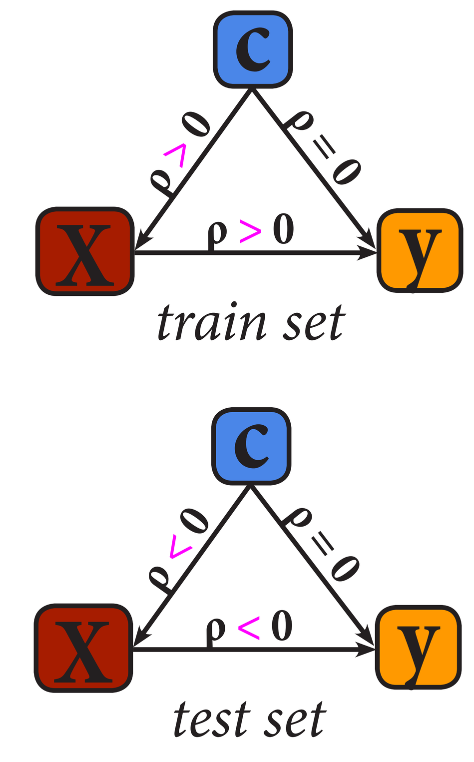


What's going on?

Let's simplify the problem and suppose y = c; thus: ρ(y, c) = 1

After regressing out c from X, correlation ρ(X, y) = 0





Due to random sampling, it may be that:

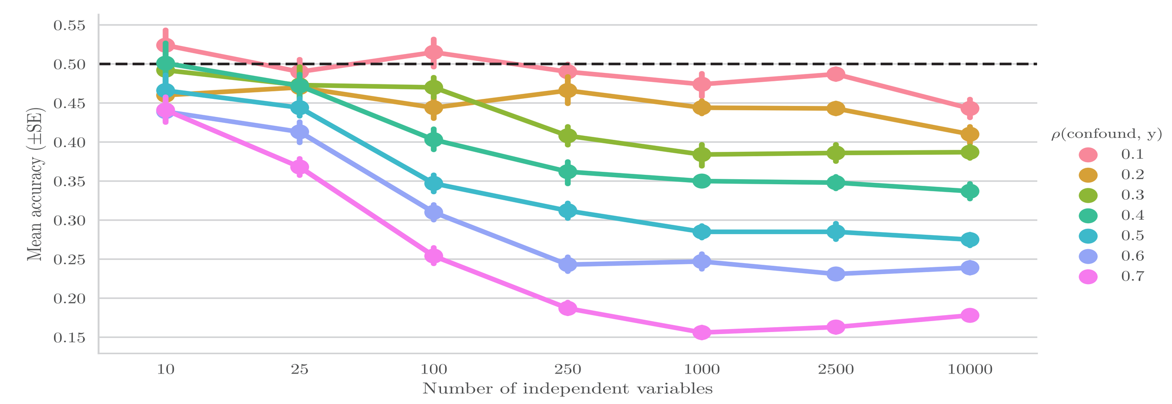
$$\rho(X_{train}, y_{train}) > 0$$

but, because

$$\rho(X_{full}, y_{full}) = 0$$

often:

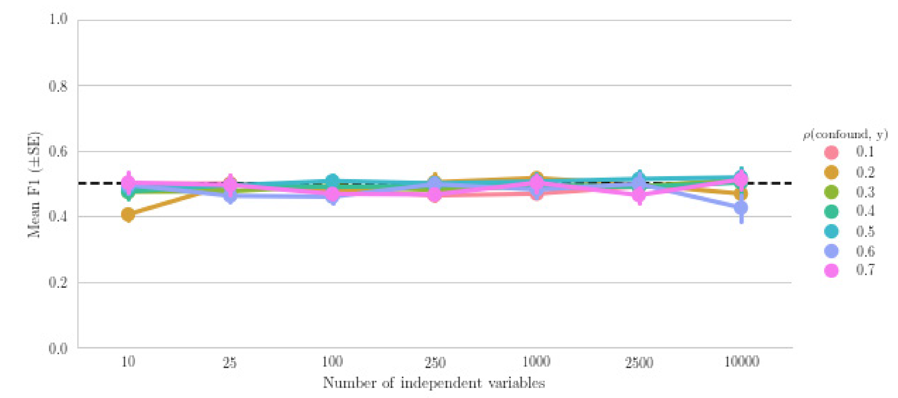
$$\rho(X_{test}, y_{test}) < 0$$



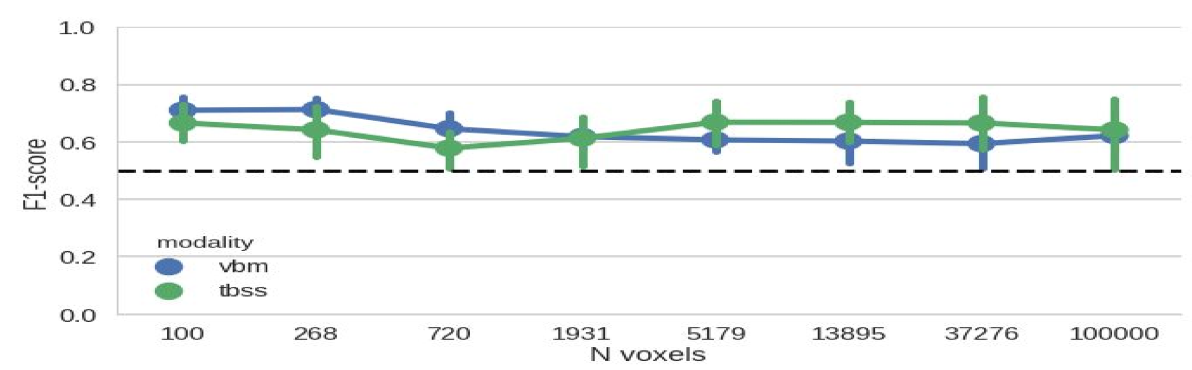
The strength of the negative bias depends on the number of independent variables, and the correlation coefficient between the confound and y

Solution

The problem can easily be solved by regressing out c from X within each fold! In simulations without a correlation ρ(X, y)...



...and in our empirical example, where there is a relation, $\rho(X, y)_{\text{corrected foldwise}}$



Conclusion

- Confound regression introduces bias in cross-validated MVPA pipelines, especially when many voxels are used
- Regressing out confounds foldwise is a universal and simple method, enhancing the generalizability of MVPA results

References

1 Naselaris & Kay, *TICS*, 2015
2 Todd et al., *NeuroImage*, 2013
3 Woolgar et al., *NeuroImage*, 2014
4 Hebart & Baker, *Arxiv*, 2017
5 github.com/lukassnoek/MVCA
6 github.com/lukassnoek/skbold