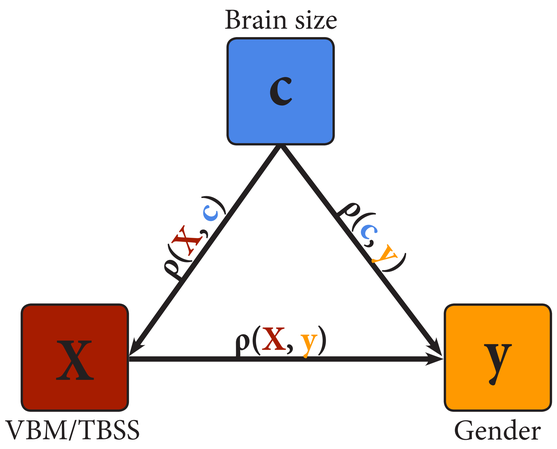
A universal method of controlling for confounds in MVPA

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

Introduction

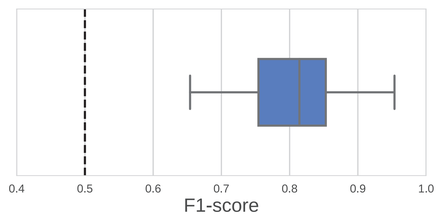
- Contrary to mass-univariate analyses, where confounds are often controlled for, it is **unclear how to handle confounds** in MVPA[1,2]
- This poses a serious threat to the **generalizability of MVPA results** in both clinical and fundamental research - especially because MVPA is arguably **more sensitive**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



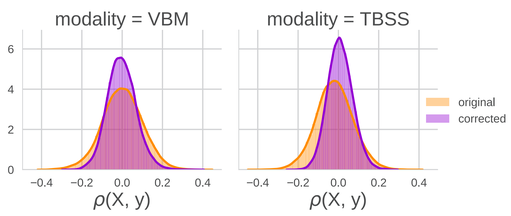
What's the problem?

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

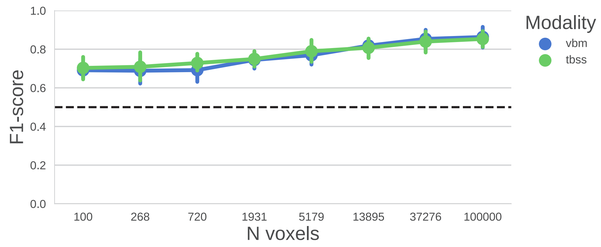
We know that brain size truly confounds ρ(**c**, **y**)...[2]



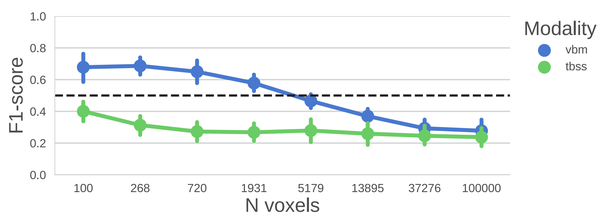
…and is related to our data...



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



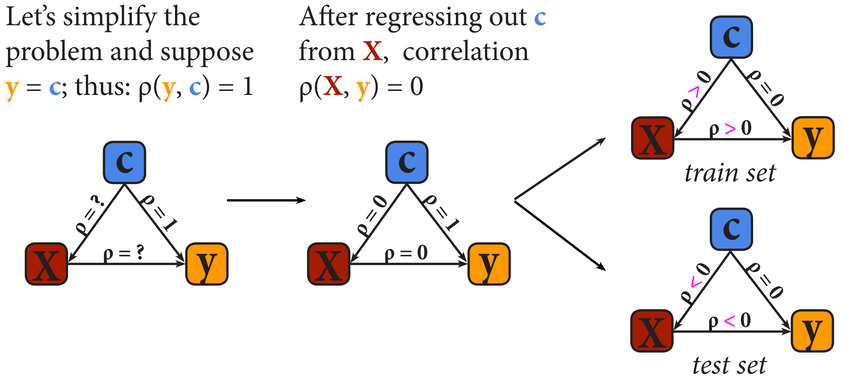
Yet, ρ(**X**, **y**)$_{\text{corrected}}$ < chance (50%)



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:

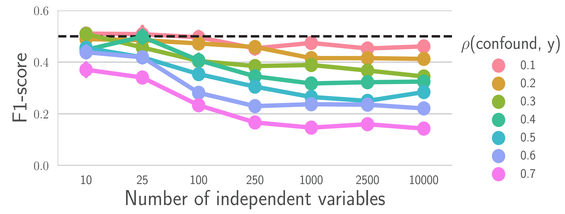
$$\rho(\mathbf{X}_{\text{train}}, \mathbf{y}_{\text{train}}) > 0$$

but, because

$$\rho(\mathbf{X}_{\text{full}}, \mathbf{y}_{\text{full}}) = 0$$

often:

$$\rho(\mathbf{X}_{\text{test}}, \mathbf{y}_{\text{test}}) < 0$$

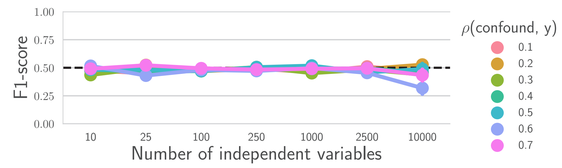


Simulations[5] show how the strength of the bias depends on:

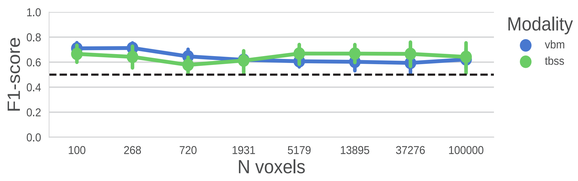
- the number of independent variables
- the correlation coefficient between the confound and y

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, ρ(**X**, **y**)$_{\text{corrected foldwise}}$



Conclusion

- Confound regression introduces bias in cross-validated MVPA pipelines[4], especially when many voxels are used
- Regressing out confounds foldwise is a universal and easy method, improving the generalizability of MVPA results

References

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