Variance explained by different model components does not behave like a Venn diagram: Why variance decomposition provides misleading intuitions

Jinkang (Derrick) Xiang, Marieke Mur*, Jörn Diedrichsen* (jxiang27, mmur, jdiedric, @ uwo.ca)
Western Institute for Neuroscience, University of Western Ontario
London, Ontario, Canada (* indicates co-senior authorship)

Abstract

When explaining brain responses Y with a set of predictors, the variance of Y is often decomposed into portions explained by each predictor, as a reflection of their contribution. The explained variance is commonly visualized using Venn diagrams. This approach originates from Fisher's ANOVA where some of the variance of a variable Y can be explained by orthogonal predictors, and the variance explained by the predictors together is the sum of the variance explained by each one alone. However, in neuroscience applications, the predictors are often correlated, which could cause the variance explained by two predictors to be smaller than, equal to or greater than the sum of each alone. Variance is not a fixed quantity of the data that can be decomposed, but should be considered in the context of all model components. We provide an alternative to the commonly used Venn diagram to visualize variance explained, and will provide an analytical framework to quantitatively conduct model selection and comparison for RSA, PCM and encoding models.

Keywords: variance decomposition; regression; model selection; representational similarity analysis (RSA); pattern component modeling (PCM); encoding models

Variance decomposition and Venn diagrams

Researchers often rely on the idea of variance decomposition when interpreting the contributions of multiple predictors to explaining brain responses. The underlying intuition – rooted in Fisher's analysis of variance (ANOVA) (Fisher, 1925) – is that we can decompose the variance of the observed data Y into a part that can be explained by predictor X_1, a part that can be explained by predictor X_2, and a residual part that is not explained by either (Fig. 1a). The proportion of variance explained by a certain set of predictors is the R^2 value for the model. In ANOVA, where predictors X_1 and X_2 are orthogonal, the variance explained by the two together equals the sum of the variance explained by each one alone: $R^2_{Y|1\cup 2} = R^2_{Y|1} + R^2_{Y|2}$.

Several studies have sought to extend this logic to the more common case of correlated predictors (Bonner and Epstein, 2018; de Heer et al., 2017; Lescroart et al., 2015). In this case, often the variance explained by two predictors together is smaller than the sum of the variance explained by each predictor alone. The difference is attributed to the "overlapping" proportion of the variance that can be explained by either predictor, that is $R^2_{Y|1,2} = R^2_{Y|1} + R^2_{Y|2} - R^2_{Y|1\cup 2}$. This logic is often visualized using a Venn diagram, subdividing the variance of Y into four parts (Fig. 1b). While intuitively appealing, this approach can lead to incorrect conclusions about how models with multiple correlated predictors behave. In the following sections, we outline several scenarios where this intuition breaks down.

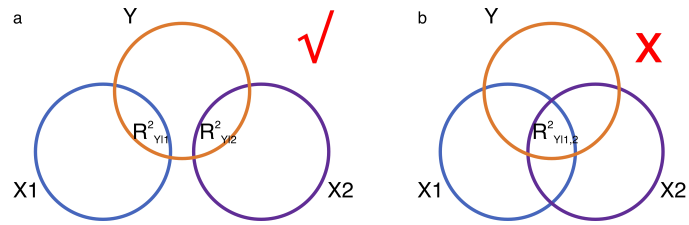

Figure 1: The logic of variance decomposition breaks down for correlated predictors. (a) In the case of uncorrelated predictors, the variance of Y (orange circle) can be partitioned into a portion explained by X_1 ($R^2_{Y|1}$), a portion explained by X_2 ($R^2_{Y|2}$), and a residual portion not explained by either predictor ($1 - R^2_{Y|1\cup 2}$). **(b)** In the case of correlated predictors, one can attempt to partition the variance of Y into the same portions as in panel a. The discrepancy between $R^2_{Y|1\cup 2}$ and the sum of $R^2_{Y|1}$ and $R^2_{Y|2}$ is attributed to the overlap in variance explained by the two predictors ($R^2_{Y|1,2}$).

Variance decomposition breaks down for correlated predictors

Suppression

Suppression occurs when the variance in Y explained by predictors X_1 and X_2 together exceeds the sum of the variance explained by each individually, that is, $R^2_{Y|1} + R^2_{Y|2} < R^2_{Y|1\cup 2}$. For example, suppose X_2 explains some variance in Y, while X_1 does not. If X_1 offsets components of X_2 that do not explain Y, the linear combination of X_1 and X_2 can explain more variance in Y than X_2 alone. In such cases, the variance in Y that is attributed to overlapping contributions from X_1 and X_2 – i.e., $R^2_{Y|1,2}$, as estimated by the decomposition logic – is negative. Indeed, suppression occurs for half of possible pairwise correlations between Y, X_1, and X_2. A concrete example of suppression is provided in *Mathematical Details*.

Explained variance summing does not imply uncorrelated predictors or predictions

It is often assumed that if the variance in Y explained by individual predictors X_1 and X_2 adds up to the variance explained by their linear combination – i.e., $R^2_{Y|1} + R^2_{Y|2} = R^2_{Y|1\cup 2}$ – then the predictors must be uncorrelated. However, this is not the case: this equality can hold even when X_1 and X_2 are correlated, as long as the overlap term $R^2_{Y|1,2}$ is zero. This occurs when the correlations among Y, X_1, X_2 satisfy $r_{12} = 2r_{Y1}r_{Y2}/(r^2_{Y1} + r^2_{Y2})$, which does not require r_{12} to be zero. In addition, when explained variance sums, Y predicted by X_1 and Y predicted by X_2 could still be correlated. See *Mathematical Details* for notation and a worked example.

Variance decomposition breaks down under cross-validation

Consider a model that includes predictors X_1 and X_2. When a third predictor X_3 is added, the model may overfit, resulting in a decrease in explained variance on held-out data during cross-validation. Under a variance decomposition framework, this would require representing X_3 as explaining negative variance – contradicting the intuitive assumptions underlying Venn diagram visualizations.

An alternative to Venn diagrams

We propose an alternative to Venn diagrams for visualizing the variance explained in observed data Y by two predictors. For any X_1, X_2, and Y, Fig. 2a shows the space pf all possible pairwise correlations among them. For a specific variable Y, explaining it with X_1 and X_2 is equivalent to projecting Y onto the hyperplane spanned by the two predictors (Fig. 2b-d). When the projection falls in the blue regions, suppression does not occur (Fig. 2b). In contrast, projections landing in the red regions indicate cases where suppression occurs (Fig. 2c). When X_1 and X_2 are orthogonal, suppression cannot occur, and the projection always falls in a blue region (Fig. 2d). The projections of Y onto X_1, X_2, and their linear combination correspond to predicted values. When all variables are normalized to unit length, the squared lengths of these projections reflect the proportion of variance explained by each model (Fig. 2b-d).

Conclusions

We show that variance decomposition and the associated Venn diagram visualization can lead to misleading intuitions about how models behave when predictors are correlated. Variance is not a fixed quantity of the data that can be decomposed, but should be considered in the context of all model components – particularly when interpreting model performance or comparing models. We propose a more mathematically grounded alternative for visualizing explained variance based on geometric projection. At the main conference, we will present a rigorous framework for evaluating models in the presence of correlated predictors in representational similarity analysis (RSA), pattern component modeling (PCM), and encoding models.

Mathematical Details

Correlations among variables

The correlations among variables Y, X_1, X_2 can be summarized in a symmetric matrix: $\begin{pmatrix} 1 & r_{Y1} & r_{Y2} \\ r_{Y1} & 1 & r_{12} \\ r_{Y2} & r_{12} & 1 \end{pmatrix}$, the determinant of which is non-negative.

Example of suppression

As an example, let $Y = \begin{pmatrix} 1 & 0 \end{pmatrix}^T$, $X_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{pmatrix}^T$, $X_2 = \begin{pmatrix} 0 & 1 \end{pmatrix}^T$. Then we have $R^2_{Y|1} = \frac{1}{2}$, $R^2_{Y|2} = 0$ and $R^2_{Y|1\cup 2} = 1$.

Example of variance summing for correlated predictors

As an example, let $Y = \begin{pmatrix} 1+\sqrt{3} & 1 & 0 & -1 \end{pmatrix}^T$, $X_1 = \begin{pmatrix} 1 & 0 & 1 & 0 \end{pmatrix}^T$, $X_2 = \begin{pmatrix} 1 & 0 & 1 & 1 \end{pmatrix}^T$. Then X_1 and X_2 are correlated, and $R^2_{Y|1\cup 2} = R^2_{Y|1} + R^2_{Y|2}$. In addition, the Y's predicted by X_1 and X_2 are correlated.

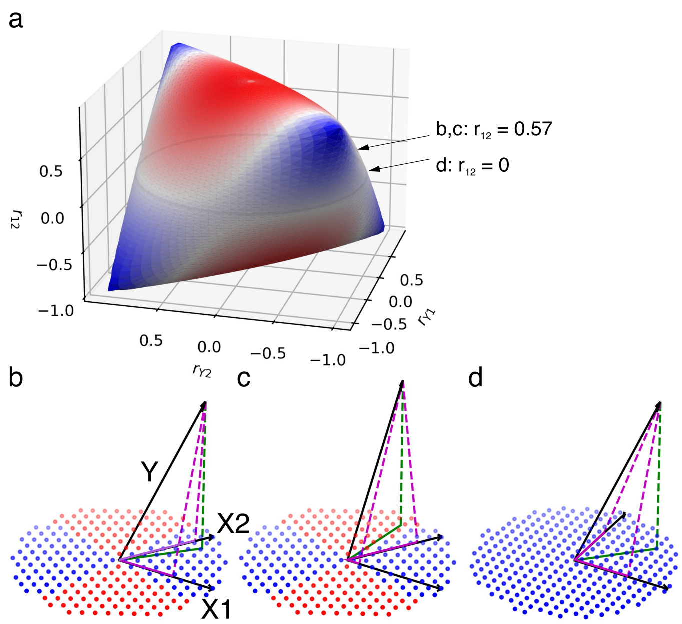

Figure 2: An alternative to Venn diagrams in multiple regression. **(a)** Visualization of all possible combinations of r_{Y1}, r_{Y2} and r_{12}. Suppression occurs in the red regions but not the blue regions. **(b-d)** 3D geometric intuition of regression as projection. Y is projected onto the plane spanned by predictors X_1 and X_2. All variables are normalized to unit length. Solid magenta lines show the predicted Y when using X_1 or X_2 individually. Their squared lengths correspond to $R^2_{Y|1}$ and $R^2_{Y|2}$, respectively. Solid green lines represent predicted Y from the linear combination of both predictors, with squared length equal to $R^2_{Y|1\cup 2}$. Dashed lines indicate residuals. Colored dots represent all possible projections of Y onto the predictor plane, with red regions indicating suppression and blue regions indicating no suppression. **(b)** Example with correlated predictors ($r_{12} = 0.57$). When the projection of Y falls in a blue region, suppression does not occur. **(c)** As in panel b), but with a projection falling in a red region, indicating suppression. **(d)** When X_1 and X_2 are uncorrelated ($r_{12} = 0$), suppression does not occur.

Acknowledgments

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), a Mitacs Globalink Graduate Fellowship, and BrainsCAN at the University of Western Ontario through the Canada First Research Excellence Fund (CFREF).

References

M. F. Bonner and R. A. Epstein. Computational mechanisms underlying cortical responses to the affordance properties of visual scenes. *PLOS Computational Biology*, 14(4):1–31, 04 2018.

W. A. de Heer, A. G. Huth, T. L. Griffiths, J. L. Gallant, and F. E. Theunissen. The Hierarchical Cortical Organization of Human Speech Processing. *Journal of Neuroscience*, 37 (27):6539–6557, July 2017. ISSN 0270-6474, 1529-2401. 10.1523/JNEUROSCI.3267-16.2017. Publisher: Society for Neuroscience Section: Research Articles.

R. A. Fisher. *Statistical Methods for Research Workers*. Oliver & Boyd, Edinburgh, Scotland, 1925.

M. D. Lescroart, D. E. Stansbury, and J. L. Gallant. Fourier power, subjective distance, and object categories all provide plausible models of BOLD responses in scene-selective visual areas. *Frontiers in Computational Neuroscience*, 9, Nov. 2015. ISSN 1662-5188. 10.3389/fncom.2015.00135. Publisher: Frontiers.