

comparing univariate and multivariate techniques with euclidean distance

mike freund

April 24, 2020

notation

For a given subject, task, and parcel, let x_v be the vector of “raw data” (i.e., the contrast estimates obtained from the GLM: e.g., Hi_Lo in Stroop). I’ll call this vector the “activation pattern” (a pattern across vertices). The subscript v indexes the vertex within the parcel, where v in $1, \dots, V$ vertices. I’ll indicate which subject the data are from by the superscript $x_v^{(s)}$, for s in $1, \dots, S$ subjects. E.g., for subject $s = 1$, the activation pattern is indicated by $x_v^{(1)}$.

euclidean distance: multivariate case

The distance between subject 1 and 2’s activation patterns is given by

$$\text{multivariate } d(1, 2) = \sqrt{\sum_{v=1}^V (x_v^{(1)} - x_v^{(2)})^2} \quad (1)$$

But, we will want to use this formula instead:

$$\text{multivariate } d_{\text{scaled}}(1, 2) = \sqrt{\frac{\sum_{v=1}^V (x_v^{(1)} - x_v^{(2)})^2}{V}} \quad (2)$$

This is because the magnitude of d^2 will scale with the number of dimensions V . Our main goal is to compare the sensitivity of d when all dimensions are used, to when only one is used (the mean). Dividing the distances by \sqrt{V} puts these metrics on the same scale.

euclidean distance: univariate case

We will use the same formula for the univariate case. However, now, we will not compute the distances between the pattern vectors $x_v^{(s)}$, but between the *means* of the pattern vectors.

So, first, calculate the across-vertex mean of each pattern vector

$$\bar{x}^{(1)} = \frac{1}{V} \sum_{v=1}^V x_v^{(1)}$$

and

$$\bar{x}^{(2)} = \frac{1}{V} \sum_{v=1}^V x_v^{(2)}$$

Then simply plugging these guys into Equation 2, we get

$$\text{univariate } d_{\text{scaled}}(1, 2) = \sqrt{\frac{\sum_{v=1}^V (\bar{x}^{(1)} - \bar{x}^{(2)})^2}{V}}$$

But because “ V ”, in this case, actually equals 1 (the mean is a scalar value), we can remove the denominator and summation operation

$$\text{univariate } d_{\text{scaled}}(1, 2) = \sqrt{(\bar{x}^{(1)} - \bar{x}^{(2)})^2}$$

which is equal to the absolute value of the difference between means:

$$\text{univariate } d_{\text{scaled}}(1, 2) = |\bar{x}^{(1)} - \bar{x}^{(2)}| \quad (3)$$

inferential statistics to perform

There are three things we want to know:

1. Perform the within-between subject ‘fingerprinting’ procedure using the *univariate* d (Equation 3). I.e., contrast the average within-subject, between-run *univariate* d with the average *between*-subject, between-run *univariate* d . In which parcels is this contrast significantly *less than zero*? (I.e., smaller within-subject distance than between-subject distance)?
2. Perform the same ‘fingerprinting’ procedure using the *multivariate* d (Equation 2); which parcels are identified as being significant?
3. Directly compare the magnitude of the contrast statistic between *univariate* and *multivariate* measures. This could be performed via a paired-sample t-test.