

Weighted metric scaling and distance-based regression in genomic data

Carles M Cuadras¹ and Daniel Cuadras¹

¹ Department of Statistics, University of Barcelona, Spain

The so-called distance-based regression (Cuadras, 1989; Cuadras, Arenas, 1990; Cuadras et al, 1996), is a method which regresses a continuous variable on the principal dimensions obtained via metric scaling. This method is quite flexible and can deal with explanatory mixed variables. It has been used in non-linear regression (Cuadras and Fortiana, 1993), in predicting short-term solar flares (Jakimiec and Bartkowiak, 1994), in relation with a new methodology to construct a tuned QSAR model (Amat et al, 1998), in extending MANOVA (Gower and Krzanowski, 1999), in molecular quantum similarity measurements (Robert et al, 2000), in the rate making processes (Boj et al, 2007). This approach can also be applied in relating genotype (e.g., DNA data) to phenotypes of interest (Wessel and Schork, 2006).

This contribution presents weighted MDS (WMDS) and its relation to classic or unweighted MDS. Next, WMDS is applied to describe parametric correspondence analysis (Cuadras and Cuadras, 2006). Finally, we consider distances between the allelic composition of a set of individuals, obtain principal dimensions as explanatory variables by WMDS, and perform a distance based-regression taking into account a quantitative variable representing a phenotype of interest.