Inge Koch
The University of Adelaide
Adelaide, Australia

Optimal Partitioning of Quadratic Forms for Multivariate Data

Quadratic forms can be represented as $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$, where $\mathbf{x}$ is a d-dimensional vector and A a symmetric matrix. Quadratic forms capture multivariate information in a single number $Q(\mathbf{x})$. This property makes them useful in statistics, for example in hypothesis testing. Other quadratic forms that are commonly used in statistics include the Mahalanobis distance and Fisher's discriminant function.

If the number of variables of the multivariate vector or data is large, or if the value $Q(\mathbf{x})$ obtained from the quadratic form is large, it will be informative to partition the quadratic form into contributions of individual variables. A partition of $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ is a transformation $\mathbf{x} \to \mathbf{w}$, with $\mathbf{x}, \mathbf{w} \in \mathbb{R}^d$ such that $Q(\mathbf{x}) = \mathbf{w}^T\mathbf{w}$. There are many different ways such partitions can be formed. We are interested in a solution or partition $\mathbf{w}$ of $Q(\mathbf{x})$ such that pairs of corresponding variables (x_j, w_j) are highly correlated and such that $\mathbf{w}$ is simpler than the given $\mathbf{x}$.

We review natural and meaningful partitions, determine selection criteria and show how to construct optimal partitions. The optimal partitions are based on transformations of the random vector(s) or data that maximise the correlations between individual variables and the new transformed variables under appropriate constraints. I present properties of the partitions including some optimality results. In addition, I show how these transformations work in practice in a partitioning of the Mahalanobis distance, and of Hotelling's one- and two-sample T^2 statistics, and I explain how we can extend these ideas to discriminant analysis.