Multivariate Behavioral Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/hmbr20

Using Support Vector Machines for Facet Partitioning in Multidimensional Scaling

Patrick Mair, Joshua S. Cetron & Ingwer Borg

MULTIVARIATE BEHAVIORAL RESEARCH
https://doi.org/10.1080/00273171.2022.2035207

Using Support Vector Machines for Facet Partitioning in Multidimensional Scaling

Patrick Mair[a], Joshua S. Cetron[a], and Ingwer Borg[b]

[a]Department of Psychology, Harvard University; [b]Institute of Psychology, WWU Muenster

ABSTRACT

In this article we focus on interpreting multidimensional scaling (MDS) configurations using facet theory. The facet theory approach is attempting to partition a representational space, facet by facet, into regions with certain simplifying constraints on the regions' boundaries (e.g., concentric circular sub-spaces). A long-standing problem has been the lack of computational methods for optimal facet-based partitioning. We propose using support vector machines (SVM) to perform this task. SVM is highly attractive for this purpose as they allow for linear as well as nonlinear classification boundaries in any dimensionality. Using various classical examples from the facet theory literature we elaborate on the combined use of MDS and SVM for facet-based partitioning. Different types of MDS are discussed, and options for SVM kernel specification, tuning, and performance evaluation are illustrated.

KEYWORDS
Multidimensional scaling; support vector machines; facet theory

Introduction

Multidimensional scaling (MDS; Borg & Groenen, 2005; Borg et al., 2018; Kruskal, 1964a, 1964b) is a statistical technique that represents proximity data (e.g., inter-correlations of test items) as distances among points in a low-dimensional space. MDS allows researchers to describe, explore, or test the structure of the data (Mair et al., 2016). The main graphical output in MDS is the *configuration plot*. The most obvious way of interpreting this plot is to judge the distances among the points: which objects are close to each other, and which objects are far apart from each other? But there are several other ways to interpret an MDS configuration. Often researchers almost automatically ask for the meaning of the plot's "dimensions" which are typically the principal axes of the plot. While in some cases it is possible to meaningfully label these axes, it is important to keep in mind that these axes can be rotated or reflected arbitrarily, and oblique dimensions would also span the MDS space. Therefore, as Borg et al. (2018) elaborate on, interpretation of these axes should not always be the primary objective. Dimensional interpretations can be supported, however, by using the dimensions' coordinates as predictors of external variables on which the objects have also been measured (Mair, 2018).

Sometimes researchers apply cluster analysis (e.g., hierarchical clustering, *k*-means, etc.) on the fitted MDS distances, or they use more sophisticated algorithms that perform scaling and clustering simultaneously (DeSarbo et al., 1991; Rusch et al., 2021; van der Maaten & Hinton, 2008). In any case, in a scaling configuration clusters are lumps or chains of points surrounded by empty space. These separability requirements are rarely, if ever, derivable from content theory. Cluster analysis is therefore not a method for validating an MDS interpretation (Borg et al., 2018).

A more refined and more general method of interpreting a configuration is based on *regions*, of which clusters and dimensions are but special cases. Regions are sub-spaces of the MDS space. They result from *partitioning* the space into connected, non-overlapping, and exhaustive parts, just as sub-dividing a map of a country by inserting regional boundaries, or as cutting a pie into wedges. The overall pattern of the regions may then show a law of formation for the observed proximities such as, for example, a circumplex or a grid of crossing axes.

In MDS, regions are typically based on *facets*. Facets are attributes of the objects of interest, or

CONTACT Patrick Mair mair@fas.harvard.edu Department of Psychology, Harvard University, Cambridge, MA, USA
Supplemental data for this article is available online at https://doi.org/10.1080/00273171.2022.2035207.

conceptual notions that sort these objects into classes. They are assigned to the objects by the researcher as a way to structure the research domain. This is possible in any research design, but it is not always done explicitly. Intelligence tests, for example, often use tasks presented in numerical, geometric, or verbal language. They also require the test person to apply, find, or learn a rule that solves the task. These two facets, "language" and "rule," can guide the construction of test items, or allow coding given items into different classes of test items. Given such a coding scheme, one can take it to test results on the items, and ask if and how it is mirrored, facet-by-facet, in the data. One relevant correspondence hypothesis for an MDS representation of the items' inter-correlations is that items that belong to the same item class are all represented by points contained in the same region, and items of different classes fall into different regions. If so, and if there are sufficiently many points in a given dimensionality and if the facets are clear and not overly simple, it shows that classifying items of a particular universe of items by these facets is not only conceptually but also empirically non-trivial and, therefore, useful. The facets, then, succeed to structure the data. They may even suggest a law of formation explaining why the various item classes are related to each other in the observed way. The general hypothesis that the facets "cut" the data space into homogeneous regions can be strengthened by requiring that the regional boundaries satisfy certain additional criteria. For example, they could be predicted to be ordered, linear, and parallel, and two facets can even form a checkerboard pattern in the plane. Such hypotheses are developed systematically in *facet theory* (Borg, 1994, 2005; Borg & Shye, 1995; Canter, 1985; Guttman, 1959; Hackett, 2014; Hackett & Fisher, 2019; Shye, 1998).

In practice, the effects of facets are often difficult to identify in an MDS configuration, because the point classes representing different categories of a facet cannot be easily separated in space. Rather, they form fuzzy and overlapping clusters, or exhibit outliers suggesting curvy regional boundaries that are hard to describe and that may also show many classification errors. Such boundaries, in turn, prevent interpretations of the MDS space in the sense of an overall law of formation that is likely to hold across replications. Finding optimal compromise boundaries is, however, rarely easy in practice, in particular if there are many points and if there is little previous experience with the particular content domain. In the facet theory literature, the partitioning and visualization nowadays is most often done "by hand," using pencil and eraser in a seemingly endless trial-and-error fashion, guided by experience, and limited to 2-dimensional solutions (Borg et al., 2020).

This paper tackles the problem how facets can be optimally identified in an MDS space, and how their effect on the MDS configuration can be visualized. We propose using *support vector machines* (SVM; Cortes & Vapnik, 1995; Steinwart & Christmann, 2008; Vapnik, 2000) from the statistical/machine learning area for facet identification. We coin this approach *MDS-SVM*. Using linear, polynomial, and radial kernels we are able to partition an MDS space of any dimensionality. MDS-SVM uses a two-step approach where we first fit the MDS solution in an unsupervised manner, followed by finding separating hyperplanes using SVM with the point coordinates as predictors and a class vector (i.e., facet category codings) as response.

The paper is organized as follows. In the theory section we start with a brief introduction to MDS, elaborate on classification using SVM, and show how these two methods can be combined, leading to the MDS-SVM approach. We then present a sequence of real-life datasets, some of them are classic datasets from the facet theory literature, and illustrate how SVM-MDS is able to detect facets. R (R Core Team, 2021) code to reproduce the results is given in the supplemental materials.

MDS-SVM theory

Multidimensional scaling

MDS takes an input dissimilarity matrix $\mathbf{\Delta}$ of order $n \times n$ (n as the number of objects to be scaled) with elements δ_{ij}. These dissimilarities can be either directly observed (e.g., a participant in an experiment is asked to directly rate proximities among objects) or derived (e.g., by computing a correlation matrix on a multivariate dataset, subsequently converted into a dissimilarity matrix). For a given dimensionality p the most popular target function to be minimized is called *stress* (Kruskal, 1964a), and defined as

$$\sigma(\mathbf{X}) = \sum_{i<j} w_{ij}(\hat{d}_{ij} - d_{ij}(\mathbf{X}))^2. \quad (1)$$

with the constraint $\sum_{i<j} w_{ij}\hat{d}_{ij}^2 = n(n-1)/2$. $\mathbf{X}$ is an $n \times p$ matrix containing the object coordinates in the p-dimensional space (*configuration*). The corresponding fitted (Euclidean) distances are

$$d_{ij}(\mathbf{X}) = \sqrt{\sum_{s=1}^{p} (x_{is} - x_{js})^2}.$$

The weights w_{ij} in Equation (1) are non-negative a priori weights (most often simply 1's or 0's for missing data), whereas $\hat{d}_{ij}$ are the *disparities* (or *d-hats*) resulting from an optimal scaling transformation of the input dissimilarities, $\hat{d}_{ij} = f(\delta_{ij})$. Popular transformations are ratio, interval, and ordinal. Stress minimization can be tackled using an iterative approach based on majorization (De Leeuw, 1977; De Leeuw & Mair, 2009). Note that Equation (1) is typically normalized to *stress-1* by dividing it by $\sum_{i<j} w_{ij}\hat{d}_{ij}^2$ and then taking the square root. This makes the stress value scale-free (Mair et al., 2022).

While in most applications MDS is used as an exploratory tool, it is possible to fit MDS in a confirmatory manner. The idea of confirmatory MDS involving external constraints goes back to Bentler and Weeks (1978), Borg and Lingoes (1980), De Leeuw and Heiser (1980), and Heiser and Meulman (1983). In case of external constraints on dimensions of the MDS solution $\mathbf{X}$, one first formulates an $n \times q$ matrix $\mathbf{Z}$ ($q \geq p$) of dimensional scale values based on theoretical considerations, and then allows for a rescaling transformation of the form

$$\mathbf{X} = \hat{\mathbf{Z}}\mathbf{C}, \tag{2}$$

which is directly incorporated into Equation (1). $\hat{\mathbf{Z}}$ is a transformed version of predictor matrix $\mathbf{Z}$ achieved through an optimal scaling step within each iteration of the majorization algorithm. $\mathbf{C}$ is a $q \times p$ matrix of regression weights to be estimated, subject to potential further restrictions (Mair et al., 2022). These further restrictions allow users to, for instance, impose an axial structure on the configuration.

Both exploratory and confirmatory MDS can be applied within the context of the MDS-SVM approach. For further technical details on MDS and related methods we refer the reader to Borg and Groenen (2005).

Support vector machines

The basic objective of a facet-theoretical approach to interpreting an MDS solution is to partition the point configuration into regions that correspond to the content classes of the observations represented by this space. Although this partitioning can be done by hand, SVM is a family of classifiers that provide an effective, data-driven method for separating a set of points into classes.

The SVM computes a boundary separating the entire space of observations into classes using *support vectors* to determine the location of the boundary. Support vectors are the points of each class that lie closest to a separating boundary and that, if moved, would change the boundary. The support vectors thus determine the position of the boundary; other points are not relevant. For two-dimensional data, the simplest boundary an SVM can find is a straight line that separates the classes. For higher-dimensional data, the boundary found by an SVM is a multidimensional surface (*hyperplane*). Generally speaking, SVM classifiers divide a p-dimensional point space so that points belonging to different classes lie on different sides of a separating hyperplane of dimension $p-1$.

To elaborate in more detail on the SVM classifier within an MDS context, we will use the following data on PTSD (post-traumatic stress disorder) symptoms reported by 362 survivors of the Wenchuan earthquake in China in 2008, collected by McNally et al. (2015). There are 17 PTSD symptom items scaled on a rating scale from 1 ("not at all") to 5 ("extremely"), indicating how seriously they have been bothered by the symptom during the past month. As described in the previous section, we can use MDS to visualize the similarity structure of the symptom observations in two dimensions, as shown in Figure 1. To generate this MDS configuration, we first obtain $\boldsymbol{\Delta}$ by computing the Euclidean distances among the 17 items. We then fit a two-dimensional, ordinal MDS on these distances, leading to a stress-1 value of 0.133. This result reproduces the analysis from Mair et al. (2016) and Mair (2018).

Now suppose that these PTSD symptoms each belong to one of two categories of a hypothetical facet: Class A, containing the symptoms "future," "numb," "distant," "lossint," "amnesia," "avoidth," and "avoidact"; and Class B, containing the remaining symptoms. We can apply the simplest form of an SVM to partition the MDS space into two regions according to these classes. This implementation is known as a linear SVM, because it finds a linear boundary (in this case, a straight line) that separates the points into two regions representing the two facet categories perfectly.[1] Figure 2 shows the result of this linear SVM classification. This example provides the simplest case of an SVM classification where points belong to two classes that can be perfectly separated in space by a linear hyperplane.

Let us look at the theory behind this partitioning.[2] In two dimensions, a hyperplane is a line; in three dimensions, it is a plane; and in p dimensions, it is a $(p-1)$-dimensional hyperplane. A hyperplane with parameters $\beta_0, \ldots, \beta_p$ is the set of all points with coordinates $x_1, \ldots, x_p$ for which it holds that

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p = 0. \tag{3}$$

We now define $\mathbf{X}$ as the $n \times p$ configuration resulting from an MDS fit. Let $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{ip})^\top$ be the

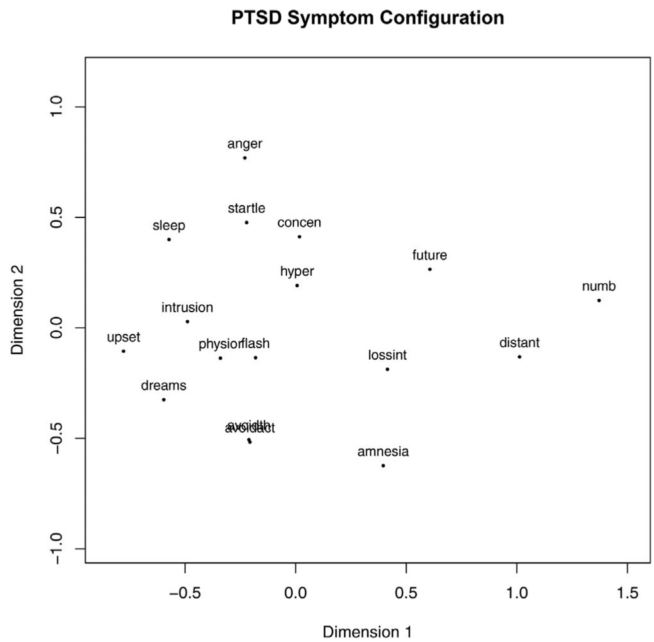

Figure 1. MDS configuration of PTSD symptoms (Wenchuan data).

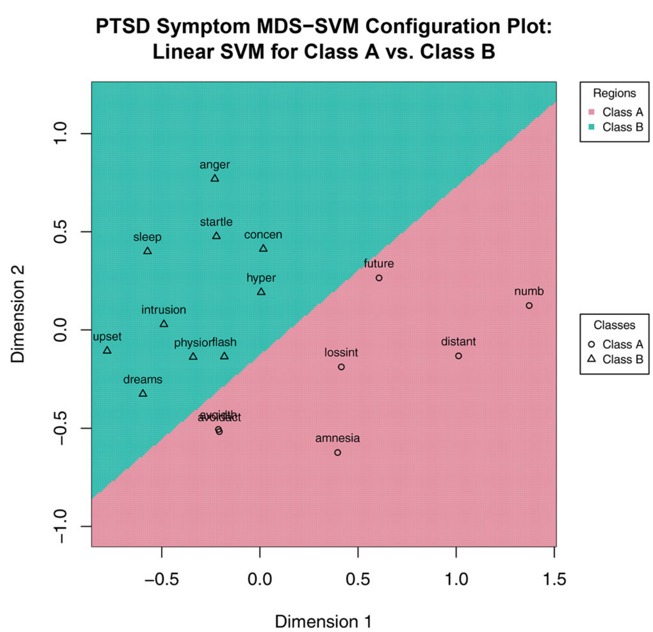

Figure 2. MDS configuration of PTSD symptoms facet with linear SVM partitioning into two regions for hypothetical classes A and B, which are perfectly separable.

p-dimensional coordinate vector for object/observation i (with $i = 1, \ldots, n$). In addition, for mathematical convenience we encode the binary response y_i (two facet categories) using $\{-1, 1\}$. This implies that the separating hyperplane has the property that

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}) > 0, \tag{4}$$

which gives a decision boundary where each observation i is assigned to a class depending on which side of the hyperplane it is located.

If the classes are perfectly separable, as in Figure 2, there usually exist many boundaries that split the point classes. However, one can find an optimal boundary by rewriting the solution in terms of the *margin* M between the support vectors. Recall that the support vectors are the points from each class that are closest to the boundary line. M is the distance between the boundary and any support vector. Thus, the margin defines the width of the "street" between the point classes that the boundary splits. If we have

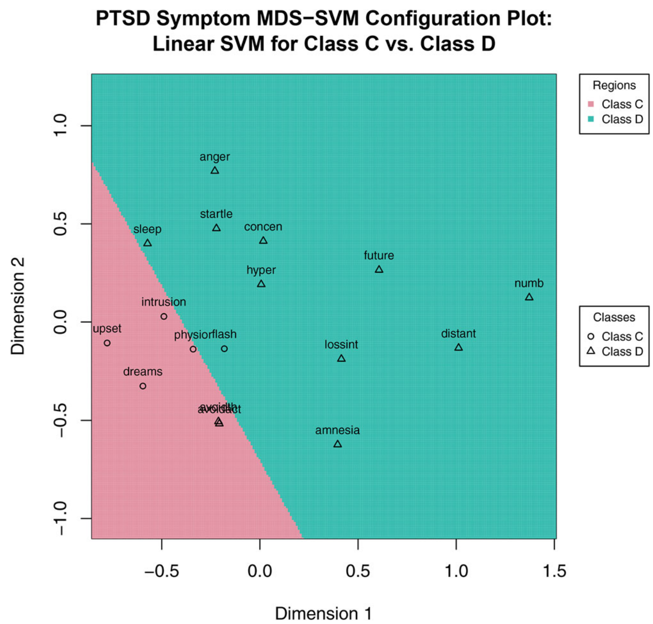

Figure 3. MDS configuration of PTSD symptoms facet with linear SVM partitioning into two regions for hypothetical classes C and D. Linear SVM classification for these classes results in several misclassified observations.

perfect linear separability, we can pick the optimal SVM boundary as the one that has the widest margin. This boundary is maximally robust in the sense that the points are still correctly classified even if slightly moved in space (e.g., by random noise). A boundary that allows for the greatest amount of random noise is known as the *maximal margin classifier*. It represents the solution to the following maximization problem, modifying Equation (4): Maximize M over $\beta_0, \ldots, \beta_p$ subject to

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}) \geq M \qquad (5)$$

with the constraint $\sum_{j=1}^{p} \beta_j^2 = 1$ (with $j = 1, \ldots, p$). This implies that $y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip})$ represents the distance from the i-th object to the decision boundary.

The case of perfect separation is rare in actual classification scenarios which limits the applicability of the maximal margin classifier. SVM classifiers can, however, utilize additional parameters that allow for generalization beyond the maximal margin classifier to cases where the points are not perfectly separable. Consider an alternative hypothetical facet for the PTSD data: Class C and Class D, shown in Figure 3. For these classes the linear SVM does not perfectly separate all the observations; it makes several misclassifications: "flash" (flashbacks) is misclassified into Class D, and "avoidth" (avoid thinking) and "avoidact" (avoid activities) are misclassified into Class C.

SVM classifiers are able to tolerate some misclassifications by introducing two modifications to the maximal margin classifier formula: slack variables ϵ_i which allow for observations to fall inside the margin or to be misclassified altogether (i.e, they fall on the wrong side of the separating hyperplane), and a hyperparameter (or tuning parameter) C which determines the total amount of slack that will be afforded during classification. These parameters modify Equation (5) as follows: Maximize M over $\beta_0, \ldots, \beta_p$ subject to

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip}) \geq M(1 - \epsilon_i) \qquad (6)$$

where $\sum_{j=1}^{p} \beta_j^2 = 1$, $\epsilon_i \geq 0$, and $\sum_{i=1}^{n} \epsilon_i \leq C$. This defines the *linear support vector classifier* (also called *soft margin classifier*).

By setting $C = 0$, the support vector classifier becomes again equivalent to the maximal margin classifier specified in Equation (5). By increasing the value of C, however, we increase the tolerance for violations of the margin. Crucially, no matter how the tuning parameter is adjusted, only the support vectors influence the solution obtained by the classifier. The support vectors now consist of any points that fall on, within, or on the wrong side of the margin.

The hyperparameter C is most commonly tuned through cross-validation (CV), either using subsets of a single set of observations (e.g., split-half CV, k-fold CV, or leave-one-out CV), or using two separately collected data sets (out-of-sample prediction). In either method the user establishes a grid of values for C, (repeatedly) train the classifier on one set of data, and then evaluate the classification on the holdout portion. The optimal value for C is then extracted

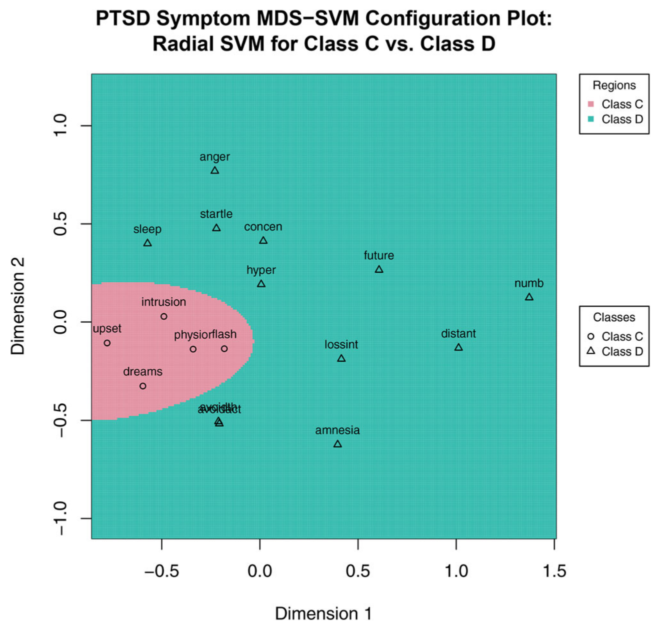

Figure 4. MDS configuration of PTSD symptoms facet with radial SVM partitioning into two regions for hypothetical classes C and D. Radial SVM classification for these classes results in perfect classification of the points.

from the classifier that performs best at predicting the category labels for the holdout data.

In addition to tuning the classifier by adjusting C, its flexibility can be extended in yet another way. By utilizing a mathematical technique known as the "kernel trick" we allow for nonlinear decision boundaries, leading to what is then actually called a *support vector machine* (SVM). The use of nonlinear kernel functions such as polynomials or radial basis functions (which can give rise to additional hyperparameters subject to tuning) enables SVM classification of points to be organized into more complicated spatial arrays. For example, if we attempt to separate classes C and D from Figure 3 using an SVM with a radial kernel instead of a linear kernel, we can capture the nonlinear category boundary between the classes without any misclassifications, as shown in Figure 4.

The ability to specify a particular kernel makes SVM a versatile tool for identifying the categorical separability of diverse point spaces. If the points are not linearly separable, the kernel idea makes it possible to find linear boundaries in an enlarged (higher-dimensional) feature space. Figure 5 presents a toy example similar to the illustration in Boehmke and Greenwell (2020, p. 278). We transform the original 2D feature space to an enlarged 3D feature space by $x_3 = x_1^2 + x_2^2$ (corresponds to a polynomial kernel function with degree 2), allowing for a linear separation of the points in 3D.

The kernel "trick" implies that we actually do not need to transform our data into this enlarged feature space. All we need is the inner product of the data (see below) in this space to compute the decision boundary. In this toy example, projecting back into a 2D space leads to a circular decision boundary. In practice, we need to find kernel K that allows for sufficient separation of the point classes. Within the context of facet theory the most relevant kernels are:

- linear: $K(\mathbf{x}_i, \mathbf{x}_{i'}) = \sum_{j=1}^{p} x_{ij}x_{i'j}$ (resulting in a support vector classifier);
- polynomial: $K(\mathbf{x}_i, \mathbf{x}_{i'}) = \left(1 + \sum_{j=1}^{p} x_{ij}x_{i'j}\right)^d$ with d as the polynomial degree;
- radial: $K(\mathbf{x}_i, \mathbf{x}_{i'}) = \exp\left(-\gamma \sum_{j=1}^{p} (x_{ij}x_{i'j})^2\right)$ with γ as the inverse of the radius of influence of the support vectors.

These kernels, all based on inner products, allow researchers to encode popular facet structures presented in Borg and Shye (1995, p. 131) such as a multiplex, a radex, or a spherex.

When there are more than two categories to classify, the SVM solution becomes more complex to derive and to evaluate. Multi-class SVM (Hsu & Lin, 2002) are commonly performed using either a one-versus-one approach where each pairwise classification option is tested (e.g., for a three-category dataset: A vs. B, A vs. C, B vs. C), or a one-versus-all approach where every category is tested against all the remaining categories (e.g., A vs. B and C, B vs. A and C, C vs. A and B). In both approaches, the final category label for each point is selected through a voting

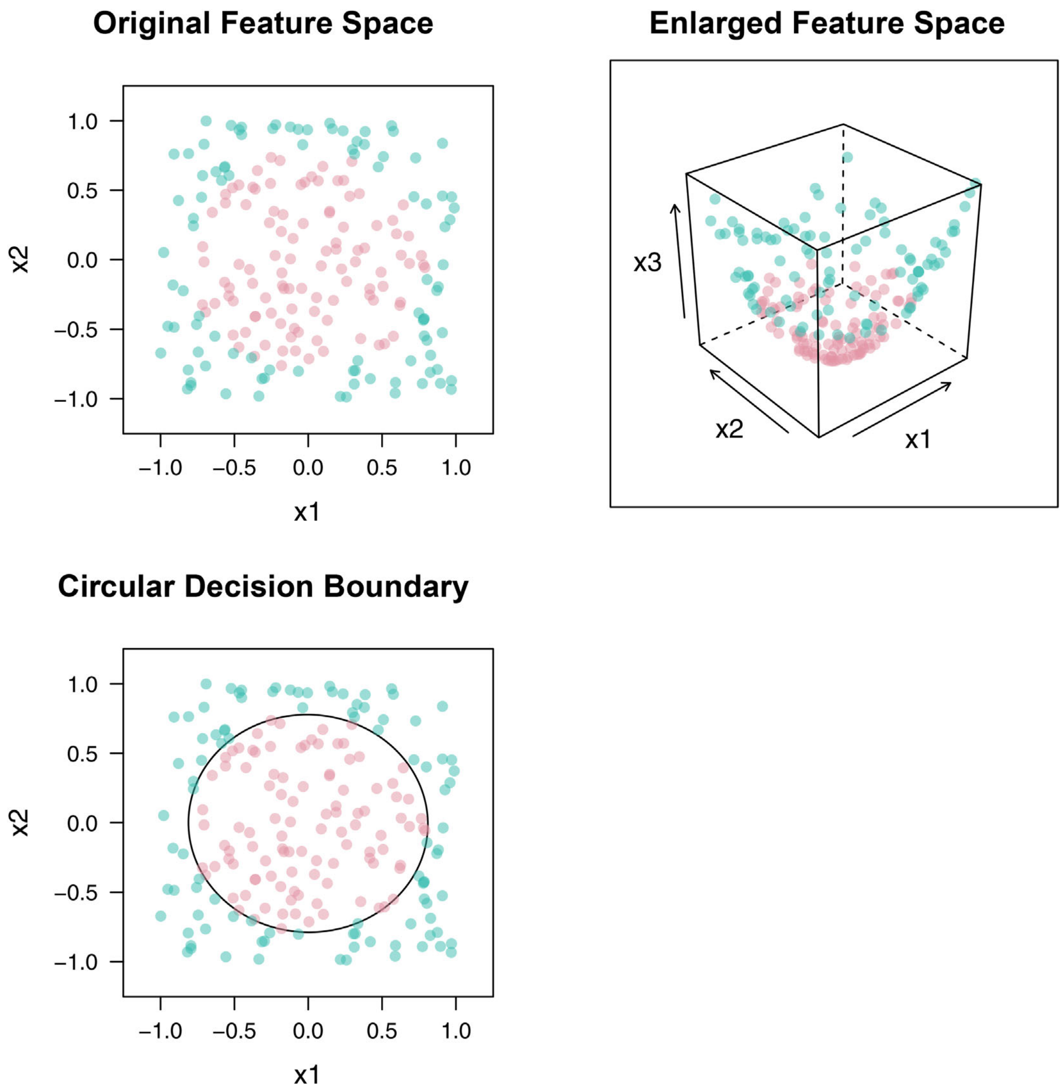

Figure 5. Enlarging the feature space for nonlinear classification. Top left: original data in the 2D space. Top right: transformed data in the enlarged 3D space, linearly separable. Bottom: decision boundary projecting back into the original 2D space.

procedure based on the aggregated category predictions. One consequence of SVM classifiers converting multi-class problems into a series of binary classifications in this manner is that the category boundaries that partition the point space in a multi-class linear SVM may appear piecewise linear or nonlinear, despite the fact that each classification is performed using a linear procedure, as we will see in some of the applications below.

In facet theory data analysis, the number of categories under consideration is frequently greater than two, making multi-class SVM necessary for facet partitioning. In fact, the symptoms measured in our PTSD example data are actually partitioned into three classes according to the DSM-IV (American Psychiatric Association, 1994): intrusive recollection, avoidance/numbing, and arousal. These classes define the PTSD facet, which can be written as PTSD = {intrusive recollection, avoidance/numbing, arousal}. We can therefore use multi-class SVM to partition the PTSD symptoms into their conceptual classes, as shown in Figure 6. With this three-category linear SVM, which uses a one-versus-one voting scheme to determine category membership, we can partition the space perfectly: all symptoms are correctly assigned to either the intrusive recollection, the avoidance/numbing, or the arousal region. In the facet theory literature, a linear facet that cuts a plane into wedges is called a *polar facet*.

Now that the MDS-SVM approach is fully introduced, let us briefly comment on algorithmic and computational challenges. In facet theory applications researchers typically aim to scale a considerably small number of points, say $n < 50$. This does not pose any runtime challenges on MDS algorithms; a solution can be found within a few seconds. Note that we usually

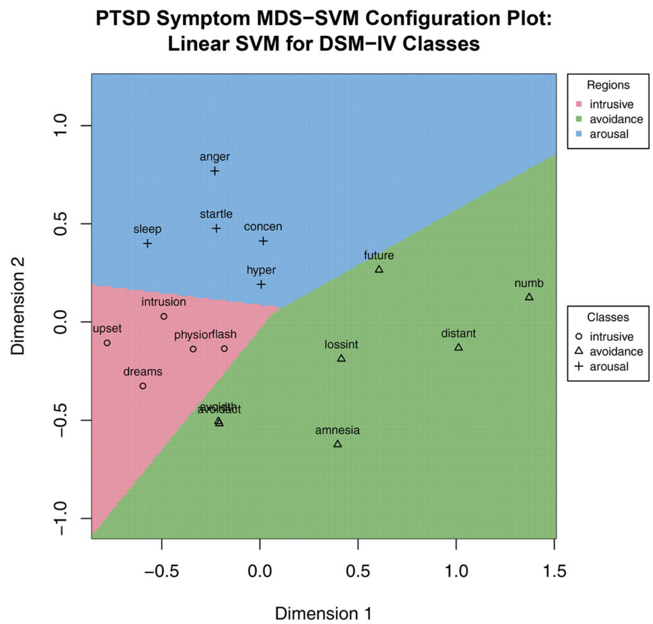

Figure 6. MDS configuration of PTSD symptoms facet with linear SVM partitioning into three regions for conceptual classes intrusive recollection, avoidance/numbing, and arousal, based on specifications from the DSM-IV. Linear multi-class SVM classification for these classes results in perfect classification of the points.

fit an MDS solution in a space of dimension $p = 2$ or $p = 3$. As a consequence, in the subsequent classification task the SVM has to deal with a small input data matrix of dimension $n \times p$. Even with extensive parameter tuning the SVM solution can be obtained within a few seconds. Since MDS-SVM is a two-step approach, it does not create any additional algorithmic challenges that are not already inherent to individual MDS and SVM computations.[3] For instance, MDS solutions suffer from a local minimum problem, whereas SVM can be sensitive to noisy data. For further details we refer the reader to more technical literature such as Borg and Groenen (2005) and Vapnik (2000).

Performance evaluation

For a general multi-class problem the performance of the SVM classification can be evaluated using standard predictive performance tools as presented in Kuhn (2008) and implemented in the caret package. One the basis of the confusion matrix, where we cross-classify the predicted values with the observed facet labels, the classification accuracy can be computed (including a 95% confidence interval, if desired). Reviewing confusion matrices in addition to overall classification accuracy can help to determine, for example, whether particular categories are especially confusable with one another. Confusion matrices can also be used to track classification performance over multiple SVM instances, such as during cross-validation, providing insight into the classification stability of particular categories across SVM with different tuning parameters. In addition, users can consider the following measures for predictive performance examination (Kuhn, 2008):

- Cohen's κ agreement coefficient;
- no-information rate: the largest proportion of the observed classes;
- one-sided hypothesis test to evaluate whether the overall accuracy rate is greater than the rate of the largest class;
- sensitivity, specificity, positive/negative predictive value, prevalence and other measures based on a one-versus-all variant of the confusion matrix;
- multi-class version of the area under the receiver operating characteristic (ROC) curve (Hand & Till, 2001).

If desired, one can also use the McNemar test to assess whether two different SVM classifications (e.g., by using different kernels or tunings) have the same error rate or not (Dietterich, 1998).

For the purpose of our facet partitioning task we consider the confusion matrix and accuracy as the most important tools for performance evaluation. They will be used in the application section.

Related techniques

MDS-SVM combines an unsupervised learning technique (MDS for scaling the objects) with a supervised learning technique (SVM for object classification). In this section we discuss the rationale for using these techniques for facet partitioning within a larger context of classical multivariate and statistical learning methods.

One could think of using other multivariate techniques such as principal component analysis (PCA) or exploratory factor analysis (EFA) for the unsupervised part. However, in quantitative facet theory research, MDS has been the scaling method of choice since its beginnings. The reason for this is that facet theory formulates hypotheses in terms of the structure of the *distances* among the variables as represented in a geometric space of low dimensionality. MDS, therefore, has also been called *smallest space analysis* (SSA) in this context (Guttman, 1968), or *multidimensional similarity structure analysis* (Borg & Lingoes, 1987). Stress-based MDS defines an explicit functional connection between the input dissimilarities and the resulting distances in a geometric space. While there are similarities between PCA and MDS (see, e.g., Borg & Groenen, 2005), PCA does not tackle this proximity scaling problem directly and usually also leads to higher-dimensional solutions. Non-metric versions of PCA such as Princals that use optimal scaling transformations are again different from MDS as they optimize the criterion of "homogeneity" (Gifi, 1990). Similarly, EFA plays no role in facet theory either as it is not a proximity scaling method.

Clustering algorithms such as k-means do not result in a scaling representation of the data. Still, MDS and clustering stand in a complementary relationship and can be used together in several ways (Kruskal, 1977). But, as mentioned in the introduction, clustering applied on top of an MDS configuration gives us clumps of points surrounded by empty space as opposed to partitions. One could consider computing Voronoi regions (see, e.g., Aurenhammer, 1991) from a prototype-based clustering solution in an MDS space. However, aside from the fact that this approach does not allow us to include any facet information and is therefore fully unsupervised, there is generally no facet-based hypothesis that would ask for Voronoi cells.

Let us now focus on the supervised part of the modeling task and elaborate on the rationale for using SVM. Depending on the application one may frequently find a different classifier that predicts better than an SVM. An important argument for using SVM over other classifiers is a substantive one: Our main goal is to partition the MDS space in a flexible manner based on theory-driven facet labels. Of key importance is that the classifier is able to fit linear and nonlinear boundaries in the MDS space which can be visualized, as this partitioned configuration plot is the main output of MDS-SVM, subject to interpretation and potentially further theoretical refinements of the facets. Decision trees, for instance, are not flexible enough as they would partition the space into axes-parallel partitions (see James et al., 2013). Linear discriminant analysis (LDA) and corresponding nonlinear extensions (see Kuhn & Johnson, 2013) could be an alternative as they are able to compute optimal hyperplanes as well. However, LDA approaches are unnecessarily restrictive as they are based on a multivariate normal assumption and generally require equal variance-covariance matrices across the classes in order to find the optimal hyperplane. In an MDS configuration we cannot readily assume that these assumptions are fulfilled. In addition, LDA methods use every point to determine the classification boundary, whereas SVM only uses the points closest to the boundary. This is an attractive feature for facet partitioning as these points are the most critical ones when developing theories, and the boundary should be sensitive to these points. Other shortcomings of LDA compared to SVM are discussed extensively in Gokcen and Peng (2002) who also found that SVM outperforms LDA in terms of predictive accuracy for various benchmark datasets they used. This said, efforts have been made to establish various relationships between SVM and other statistical techniques (e.g., logistic regression). We refer to James et al. (2013) for an overview.

Finally, another technique worthwhile mentioning here is supervised MDS (Witten & Tibshirani, 2011). It integrates a binary class vector into the MDS fit and computes the configuration according to this classification problem. Apart from the fact that it has only been developed for a two-class problem, it computes the configuration in relation to this supervised task. For our purposes we want to keep the MDS configuration unsupervised, and perform the classification in a second step without changing the configuration.

Applications

In this section we present various MDS-SVM examples of increasing complexity, some of them using classic datasets from facet theory. All computations are performed in R using the smacof package for fitting MDS (De Leeuw & Mair, 2009; Mair et al., 2022) in conjunction with the SVM implementation provided in the e1071 package (Meyer et al., 2021).

Portrait Value Questionnaire

In this example we present an application from psychological research on personal values. It uses the PVQ40 (Portrait Value Questionnaire) data from Borg et al. (2017), containing 40 items to be scaled (sample size 151). The PVQ is based on the theory of universal values (Schwartz, 1992). It distinguishes 10 "basic" values: power (PO), achievement (AC), hedonism (HE), stimulation (ST), self-direction (SD), universalism (UN), benevolence (BE), tradition (TR), conformity (CO), and security (SE). Theoretically, these values are structured in a circle of psychological conflicts and compatibilities such that adjacent values on the circle are compatible and values on opposite sides of the circle are conflicting. The rationale is that striving for conflicting values with high priority is difficult because pursuing one set of values often leads to consequences that violate conflicting values. In contrast, holding adjacent values with high priority is likely to be easy partly because adjacent values can often be pursued simultaneously by the same action. Items that measure the basic values, therefore, are expected to lead to a circumplex in MDS space, that is, to a configuration of regions organized like ten wedges of a pie, with each basic value corresponding to one wedge. Moreover, the wedges are predicted to be ordered as PO-AC-HE-…-SE-PO, with the subgroups of "higher-order values" CO/TR (conservation) opposite to HE/ST/SD (openness to change), and PO/AC (self-enhancement) opposite to BE/UN (self-transcendence).

We compute the correlation matrix of the items and convert these correlations to dissimilarities. Then, we run a 2D ordinal MDS. It finds a solution with a low stress-1 value of 0.168. We use a linear SVM to partition the MDS space into basic values regions. We establish a grid for C within a range of 10 and 100 with a step size of 2. Note that establishing a tuning grid is often a trail-and-error task. If, after a first CV round, it turns out that the optimal value results in one of the grid extremes, the user may define a refined grid around this extreme value in both directions, and repeat the CV for this new grid. Similarly, if the initial grid is coarsely defined, the user may redefine a finer grid around the optimal value in the first CV round, followed by another round of CV on this finer grid. However, for this example the CV leads to an optimal value of $C = 12$.

The linear SVM partitioning leads to an accuracy of 90%. Table 1 shows the corresponding confusion matrix. The columns present the observed value classes, and the rows the predicted value classes. Four items get assigned to wrong (but neighboring) categories of the value facet. For instance, one PO item is classified as AC, one HE item as ST, and so on. The corresponding graphical illustration is given in Figure 7.

Table 1. Confusion matrix for value facet classification.

	Observed									
Predicted	AC	BE	CO	HE	PO	SD	SE	ST	TR	UN
AC	4	0	0	0	1	0	0	0	0	0
BE	0	4	0	0	0	0	0	0	0	0
CO	0	0	4	0	0	0	0	0	1	0
HE	0	0	0	2	0	0	0	0	0	0
PO	0	0	0	0	2	0	0	0	0	0
SD	0	0	0	0	0	4	0	0	0	0
SE	0	0	0	0	0	0	5	0	1	0
ST	0	0	0	1	0	0	0	3	0	0
TR	0	0	0	0	0	0	0	0	2	0
UN	0	0	0	0	0	0	0	0	0	6

The plot shows that the SVM partitioning succeeds to cut the plane almost perfectly into regions representing the basic values. The regionality approximates a circumplex, as predicted. It can be taken as a starting point for generating a perfect circumplex: Choose an origin just south of the point labeled se4 in Figure 7, and then draw straight radial lines emanating from this origin that separate the basic values fan-like into wedges. Apart from a slight overlap of the PO and AC regions, and some overlap of the CO and the TR regions, this yields a closely fitting, idealized circumplex of regions (Borg et al., 2018, p. 22). What this shows is that the SVM approach can provide a good first partitioning of the MDS space. The hand-made final adjustment is based here on series of similar studies using personal values.

Intelligence I

Next we use a classic dataset from Guttman (1965), showing inter-correlations among 21 intelligence test tasks. One facet of such items is that they ask the testee to either find a "rule" or to apply a rule. For example, "You exhibit paired elements, such as (dog, puppy), (cow,?). The subject, in answering this properly by setting calf in place of the question mark, shows that he has deduced the rule. Elements were exhibited here which obey a rule, but it was not said explicitly what the rule is; the tester infers from the response that the subject has a correct perception of the rule" (Guttman, 1965, p. 27). When the rule itself is exhibited and assumed to be understood by testee, and she has to operate according to the rule, the item would be an "achievement" item. This theorizing led to the facet requirement = {analytical, complex, achievement-1, achievement-2} which distinguishes

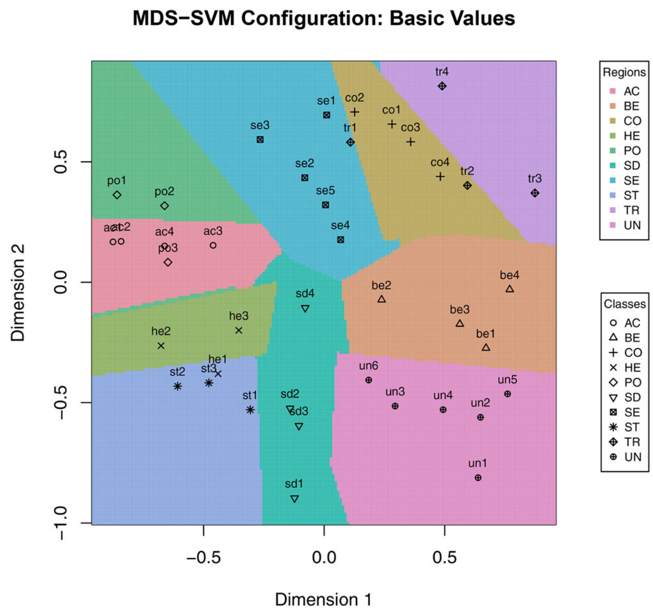

Figure 7. Ordinal MDS configuration with linear SVM partitioning of the basic values facet into its 10 elements.

four abilities required to solve the test items (see also Guttman and Levy (1991) for further details).

We begin the analysis with fitting an ordinal MDS resulting in a stress-1 value of 0.128. Guttman (1965) hypothesized that the facet "requirement" leads to a "modular" (i.e., circular bands around a common origin) partitioning of the MDS space of the tasks' inter-correlations, with analytical tasks in the center, then a band of complex tasks, and then the achievement-1 and achievement-2 tasks. This hypothesis translates into using a radial SVM kernel. We set up a parameter grid for the tuning parameter C (ranging from 0.0625 to 16) to specify the cost of margin violation. A 10-fold CV determines $C = 8$ as the optimal choice. The radial parameter γ is kept at its default value of $1/p = 0.5$. The resulting SVM fit leads to an overall predictive accuracy of 95.24%.

Figure 8 presents the radial regions on top of the MDS configuration. We see that the test item number 15 is wrongly assigned to the "complex" region. All other test items are correctly classified. The partitioning pattern detected by SVM suggests the concentric bands pattern predicted by Guttman. It could be fully developed by hand if one is willing to complement the achievement-1 and achievement-2 regions so that they become complete circular bands with sections that do not contain any points. This is done by Guttman (1965), but it goes beyond the given data by speculating about the structure of the universe of such intelligence tasks.

Intelligence II

In this second example from the area of intelligence research we use data from Guttman and Levy (1991). Each participant is required to answer subtests that can be classified by the facet "mode of expression" into

- oral tests with subtests information, similarities, arithmetic, vocabulary, comprehension, digit span, picture completion;
- manual tests with subtests picture arrangement, block design, object assembly;
- paper & pencil tests with subtests coding, mazes.

This facet distinguishes three different ways in which the test items are answered by the testee: either orally, or by manual manipulation, or by means of paper and pencil. The facet is simply a property of how the test items are constructed. Using MDS-SVM we explore a theoretical facet structure in a 3D-space, with linear separation of the three expression types of tests. We fit a 3D ordinal MDS leading to a stress-1 value of 0.043. The 3D configuration is given in Figure 9 and shows a clear linear separability of the points.

To find the optimal separating hyperplanes we fit a linear SVM. We use 10-fold CV to tune C leading to an optimal value of 2. As expected, the linear SVM classifies all objects perfectly. Yet, it is tricky to visualize the hyperplanes in 3D space. What one can do is to produce pairwise 2D plots (D1 vs. D2, D1 vs. D3, D2 vs. D3). What we learn from the 3D configuration in Figure 9 is that if we look at it from top (i.e., D1 vs. D2) we see the most obvious separation. Therefore, Figure 10 shows the corresponding dimension slice through the origin of the third dimension.

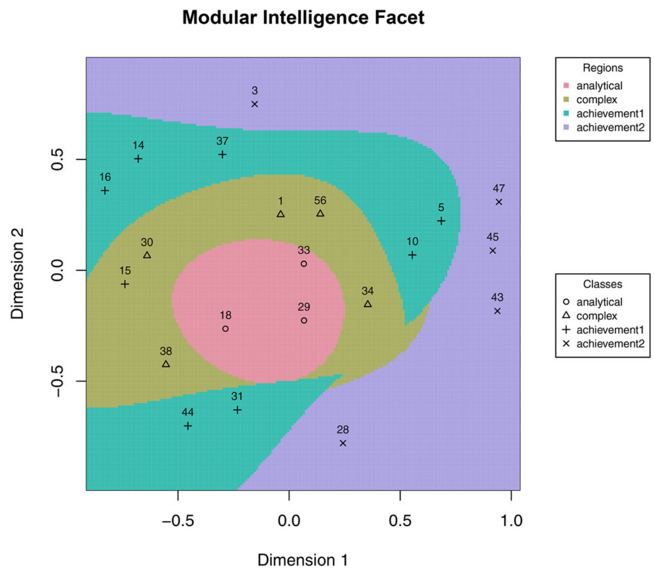

Figure 8. Ordinal MDS configuration with radial SVM partitioning.

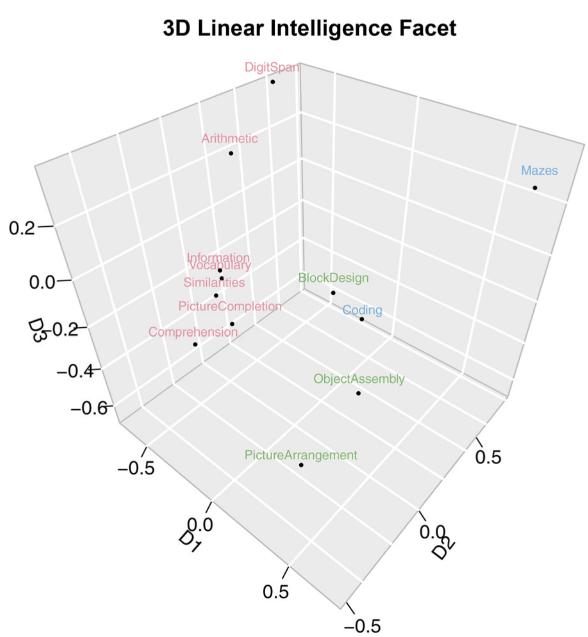

Figure 9. Ordinal 3D MDS configuration with intelligence facet: oral (red), manual (green), and paper (blue).

Morse codes

The next example uses the Morse code data by Rothkopf (1957), a classic dataset in the MDS literature. 598 participants judged whether two Morse code signals, presented acoustically one after another, were the same or not. The responses were aggregated to confusion rates and converted into dissimilarities. We use the facet definitions from Shepard (1963) and Wish (1967).

We start with fitting a 2D ordinal MDS (unconstrained). The resulting configuration is used as the starting configuration in a subsequent confirmatory MDS with the facets "signal type" and "signal length" as coordinates in **Z**. We apply an additional ordinal optimal scaling transformation with the primary approach to ties (i.e., ties in the data do not have to be mapped into the same distance). This leads to an MDS model that restricts its solution to a (metric) duplex, that is, a point configuration that can be partitioned by two (not necessarily orthogonal) sets of parallel and linear boundaries that correspond to the two facets. The solution has a stress-1 value of 0.246. By construction, the points are already separated into regions. This should make it easy for linear SVM to fit the separating hyperplanes for the following point classes: all short beeps, more short than long beeps, same short and long beeps, more long than short beeps, and all long beeps.

We will now explore whether a linear SVM is able to detect the linearly separated regions (see Figure 11). This time, instead of tuning C directly whose theoretical range is from 0 to ∞, we use a different parameterization of the tuning parameter. It is called ν-parameterization (Schölkopf et al., 2000) and normalized to the $[0, 1]$ interval. This parameterization is sometimes more convenient for tuning than the default C-classification. Using a grid within the $[0.01, 0.3]$ interval (step size of 0.01) the optimal ν is 0.18, as found by 10-fold CV. Evaluation of the SVM results suggests that 100% of the signals are correctly classified.

The plot in Figure 11 shows that the effect of the (ordered) facet "signal type" on the similarity of the Morse code signals is sorting the signals along a dimension into regions ranging from "all long (beeps)" over "more long than short," "equal," "more

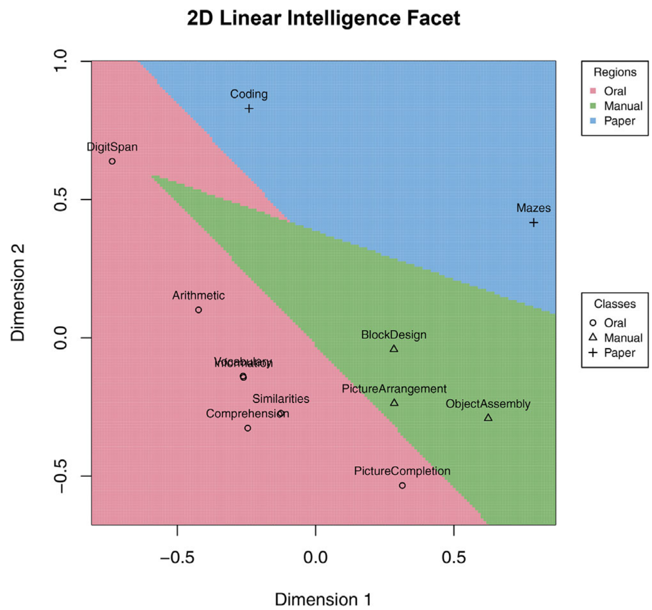

Figure 10. Linear SVM partitioning in a 3D MDS space: span plot (D1 vs. D2) showing the linear hyperplanes.

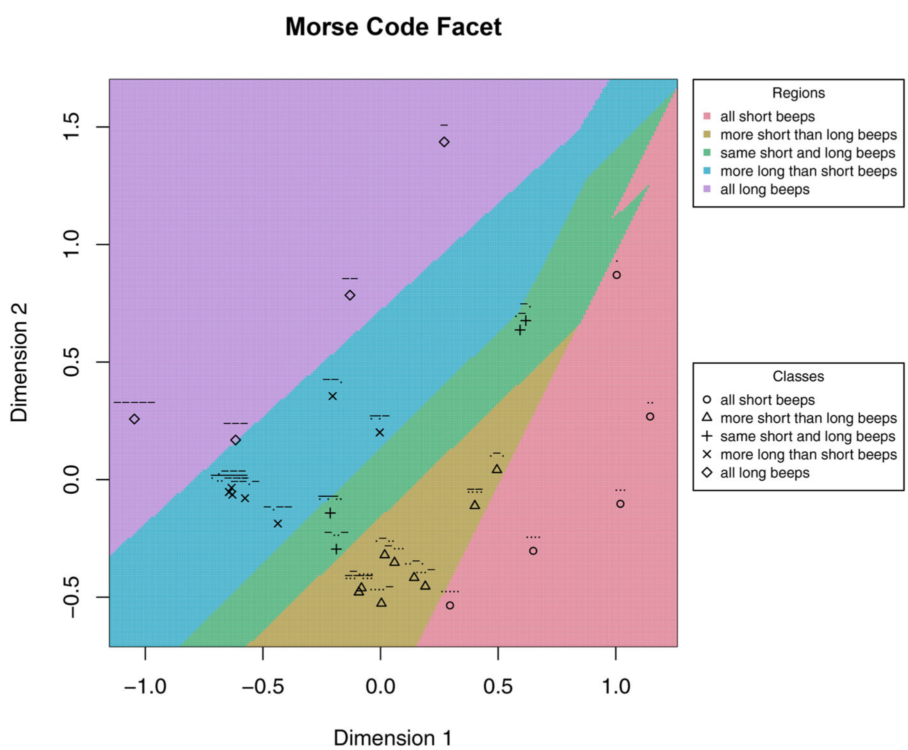

Figure 11. Restricted MDS configuration of Morse code data with linear SVM partitioning.

short than long" to "all short." This order of regions corresponds perfectly to the ordered facet. In this case we know, of course, that the MDS solution can be partitioned in the sense of a perfect duplex, because it was constrained to this type of configuration. This SVM partitioning succeeds to closely identify this regionality.

Discussion

In this article we propose to apply SVM for partitioning an MDS configuration space into non-overlapping, exhaustive regions within the domain of facet theory data analysis. SVM provide the necessary flexibility for finding potentially nonlinear regions in MDS spaces of arbitrary dimensions.

Let us now discuss some limitations of the MDS-SVM approach as presented in this paper, and show some options on how to tackle some of them in future research. Originally, SVM was designed for binary classification problems. The extension to multi-class problems is based on the one-versus-one or one-versus-all trick. In most facet theory applications we have a multi-class problem, and the application of SVM can lead to angular boundaries even though linear kernels are specified, as shown in some of the examples. Future research can focus on more recent variants of multi-class SVM as presented in Van den Burg and Groenen (2016), and Blanco et al. (2020). Note that multi-class SVM have also been extended to determine ordered partitions (Frank & Hall, 2001; Heredia-Gómez et al., 2020; Waegeman & Boullart, 2009). Using this type of SVM would allow facet researchers to partition the MDS space according to ordered facet information.

Another technical limitation of our approach is that available SVM kernels do not allow researchers to strictly enforce particular prototypical types of partitions such as a radex in 2D or a cylindrex in 3D. The user can solve such problems, in principle, by utilizing confirmatory MDS. However, so far confirmatory MDS models are only available for simple spherical structures and some mesh patterns (Browne, 1992; Cox & Cox, 1991; De Leeuw & Mair, 2009; Elad et al., 2005; Ling & Jacobs, 2007). But even where they exist, we suggest to always also run an exploratory MDS with subsequent facet-based partitions, and then compare it to the confirmatory solution. This often shows that the constrained solution moves only a few specific points to satisfy the external constraints. An example is the Morse code configuration in Figure 11, where enforcing strictly parallel regions for the facet "signal length" requires moving essentially only the point that represents the shortest beep in exploratory MDS space (Borg et al., 2018). This makes this particular signal somewhat special within the set of all Morse codes, suggesting that a simple additive dimension-based theory of similarity judgments for Morse codes is not sufficient. As the overall stress value of the confirmatory and the exploratory MDS solution is almost the same, such details can be overlooked easily if the solutions are not carefully compared.

From a practical point of view facet theory researchers sometimes want to explore whether some facets are "trivial" in the sense that they allow to perfectly partition almost any MDS configuration with n points in p dimensions. To address this question, a permutation test can be constructed for simple partitions such as the duplex using confirmatory MDS. The basic idea is to randomly permute the facet labels and then run confirmatory MDS enforcing the duplex (Borg et al., 2011, 2020). This gives a null distribution of stress values to which the stress value from the original (unconstrained) fit can be compared.

The data-driven MDS-SVM fit does not have to be the last stage of the facet analysis. The configuration plot can be used for theory formulation or revision, for instance, by simplifying partitions or by moving points, even if that leads to some misclassifications for a given set of data. An example is straightening the boundary lines in Figure 11 which, in this case, would not negatively affect the separability of the points, i.e., it would not lead to any misclassifications. Simplified boundaries often admit simplified interpretations in terms of underlying laws of formation. They also promise better replicability, because they avoid paying too much attention to noise.

The flexibility of the MDS-SVM approach can also be seen as a disadvantage, as it gives researchers the opportunity to come up with a facet solution by a trial-and-error strategy (different kernels, varying the tuning parameters, etc.) that is not anyhow theory based. We stress that facet theory researchers should aim for a solution that is on solid theoretical grounds and that is stable. For instance, for cases with derived proximities with a (moderately) large sample size, one could split the data into two sub-samples (say, a 50–50 split), derive two proximity matrices, fit two MDS solutions, and do a Procrustes transformation to remove meaningless differences in the configurations. Then, one could train an SVM on the first MDS configuration, and make the predictions based on the point locations in the second (test) configuration. If the facet elements can be predicted with high accuracy, one has reason to assume that the facets are stable across replications.

To conclude, the MDS-SVM approach proposed in this paper provides researchers with data-driven tools for partitioning MDS spaces. Even though we proposed the MDS-SVM approach within the context of facet theory, it can be used for any scaling problem where researchers have labeled objects and want to partition the space accordingly. An application from the neuroimaging area can be found in Cetron et al. (2019).

Article information

Conflict of interest disclosures: Each author signed a form for disclosure of potential conflicts of interest.

No authors reported any financial or other conflicts of interest in relation to the work described.

Ethical principles: The authors affirm having followed professional ethical guidelines in preparing this work. These guidelines include obtaining informed consent from human participants, maintaining ethical treatment and respect for the rights of human or animal participants, and ensuring the privacy of participants and their data, such as ensuring that individual participants cannot be identified in reported results or from publicly available original or archival data.

Funding: J. S. Cetron was supported by National Science Foundation Graduate Research Fellowship, grant number DGE 1745303.

Role of the funders/sponsors: None of the funders or sponsors of this research had any role in the design and conduct of the study; collection, management, analysis, and interpretation of data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

Acknowledgments

The ideas and opinions expressed herein are those of the authors alone, and endorsement by the authors' institutions is not intended and should not be inferred.

Notes

1. This solution was computed using the default tuning parameter setup from the e1071 package (Meyer et al., 2021). That is, no parameter tuning is involved. This is typically not recommended, but for this first example we keep the setup as simple as possible.

2. We introduce SVM in a non-technical way in the spirit of James et al. (2013) and Boehmke and Greenwell (2020). For a mathematically more rigorous SVM presentation see Vapnik (2000) and Steinwart and Christmann (2008).

3. The computational complexity of SMACOF is $\mathcal{O}(n^2)$; the complexity of SVM is between $\mathcal{O}(n^2)$ and $\mathcal{O}(n^3)$.

ORCID

Patrick Mair http://orcid.org/0000-0003-0100-6511

References

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders (DSM-IV)* (4th ed.). American Psychiatric Association.

Aurenhammer, F. (1991). Voronoi diagrams—A survey of a fundamental geometric data structure. *ACM Computing Surveys*, *23*(3), 345–405. https://doi.org/10.1145/116873.116880

Bentler, P. M., & Weeks, D. G. (1978). Restricted multidimensional scaling models. *Journal of Mathematical Psychology*, *17*(2), 138–151. https://doi.org/10.1016/0022-2496(78)90027-5

Blanco, V., Japón, A., & Puerto, J. (2020). Optimal arrangements of hyperplanes for SVM-based multiclass classification. *Advances in Data Analysis and Classification*, *14*(1), 175–199. https://doi.org/10.1007/s11634-019-00367-6

Boehmke, B., & Greenwell, B. (2020). *Hands-on machine learning with R*. Chapman & Hall/CRC.

Borg, I. (1994). Facet theory. In R. J. Sternberg (Ed.), *Encyclopedia of intelligence* (pp. 419–422). Macmillan.

Borg, I. (2005). Facet theory. In Brian S. Everitt & David C. Howell (Eds.), *Encyclopedia of statistics in behavioral science, volume 2* (pp. 595–599). Wiley.

Borg, I., Bardi, A., & Schwartz, S. H. (2017). Does the value circle exist within persons or only across persons? *Journal of Personality*, *85*(2), 151–162. https://doi.org/10.1111/jopy.12228

Borg, I., & Groenen, P. J. F. (2005). *Modern multidimensional scaling: Theory and applications* (2nd ed.). Springer.

Borg, I., Groenen, P. J. F., Jehn, K., Bilsky, W., & Schwartz, S. (2011). Embedding the organizational culture profile into Schwartz's theory of universals in values. *Journal of Personnel Psychology*, *10*(1), 1–12. https://doi.org/10.1027/1866-5888/a000028

Borg, I., Groenen, P. J. F., & Mair, P. (2018). *Applied multidimensional scaling and unfolding* (2nd ed.). Springer.

Borg, I., & Lingoes, J. C. (1980). A model and algorithm for multidimensional scaling with external constraints on the distances. *Psychometrika*, *45*(1), 25–38. https://doi.org/10.1007/BF02293597

Borg, I., & Lingoes, J. C. (1987). *Multidimensional similarity structure analysis*. Springer.

Borg, I., Mair, P., & Cetron, J. S. (2020). Facet theory and multidimensional scaling. In S. Shye (Ed.), *Proceedings of the 17th International Facet Theory Conference* (pp. 18–28).

Borg, I., & Shye, S. (1995). *Facet theory: Form and content*. Sage.

Browne, M. W. (1992). Circumplex models for correlation matrices. *Psychometrika*, *57*(4), 469–497. https://doi.org/10.1007/BF02294416

Canter, D. (1985). *Facet theory: Approaches to social research*. Springer.

Cetron, J. S., Connolly, A. C., Diamond, S. G., May, V. V., Haxby, J. V., & Kraemer, D. J. M. (2019). Decoding individual differences in STEM learning from functional MRI data. *Nature Communications*, *10*(1), 1–10. https://doi.org/10.1038/s41467-019-10053-y

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, *20*(3), 273–297. [Database] https://doi.org/10.1007/BF00994018

Cox, T. F., & Cox, M. A. A. (1991). Multidimensional scaling on a sphere. *Communications in Statistics - Theory and Methods*, *20*(9), 2943–2953. https://doi.org/10.1080/03610929108830679

De Leeuw, J. (1977). Applications of convex analysis to multidimensional scaling. In J. Barra, F. Brodeau, G. Romier, & B. van Cutsem (Eds.), *Recent developments in statistics* (pp. 133–145). North Holland Publishing Company.

De Leeuw, J., & Heiser, W. J. (1980). Multidimensional scaling with restrictions on the configuration. In P. Krishnaiah (Ed.), *Multivariate analysis* (Vol. V, pp. 501–522). North Holland Publishing Company.

De Leeuw, J., & Mair, P. (2009). Multidimensional scaling using majorization: SMACOF in R. *Journal of Statistical Software*, *31*(3), 1–30. https://doi.org/10.18637/jss.v031.i03

DeSarbo, W. S., Howard, D. J., & Jedidi, K. (1991). MULTICLUS: A new method for simultaneously performing multidimensional scaling and cluster analysis. *Psychometrika*, *56*(1), 121–136. https://doi.org/10.1007/BF02294590

Dietterich, T. G. (1998). Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Computation*, *10*(7), 1895–1923. https://doi.org/10.1162/089976698300017197

Elad, A., Keller, Y., & Kimmel, R. (2005). Texture mapping via spherical multi-dimensional scaling. In R. Kimmel, N. Sochen, & J. Weikert (Eds.), *Scale-Space 2005, LNCS 3459* (pp. 443–455). Springer.

Frank, E., & Hall, M. (2001). A simple approach to ordinal classification. In L. De Raedt & P. Flach (Eds.), *European Conference on Machine Learning* (pp. 145–156). Springer.

Gifi, A. (1990). *Nonlinear multivariate analysis*. Wiley.

Gokcen, I., & Peng, J. (2002). Comparing linear discriminant analysis and support vector machines. In T. Yakhno (Ed.), *Advances in information systems* (pp. 104–113). Springer.

Guttman, L. (1959). Introduction to facet design and analysis. In *Proceedings of the 15th International Congress of Psychology* (pp. 130–132). North Holland.

Guttman, L. (1965). The structure of interrelations among intelligence tests. In C. W. Harris (Ed.), *Proceedings of the 1964 Invitational Conference on Testing Problems* (pp. 23–36). ETS.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, *33*(4), 469–506. https://doi.org/10.1007/BF02290164

Guttman, L., & Levy, S. (1991). Two structural laws for intelligence tests. *Intelligence*, *15*(1), 79–103. https://doi.org/10.1016/0160-2896(91)90023-7

Hackett, P. M. W. (2014). *Facet theory and the mapping sentence: Evolving philosophy, use and application*. Palgrave Macmillan.

Hackett, P. M. W., & Fisher, Y. (2019). Editorial: Advances in facet theory research: Developments in theory and application and competing approaches. *Frontiers in Psychology*, *10*, 1–3. https://doi.org/10.3389/fpsyg.2019.01642

Hand, D. J., & Till, R. J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning*, *45*(2), 171–186. https://doi.org/10.1023/A:1010920819831

Heiser, W. J., & Meulman, J. (1983). Constrained multidimensional scaling, including confirmation. *Applied Psychological Measurement*, *7*(4), 381–404. https://doi.org/10.1177/014662168300700402

Heredia-Gómez, M. C., García, S., Gutiérrez, P. A., Herrera, F. (2020). *OCAPIS: Ordinal classification and preprocessing in Scala* [Computer software manual]. R package version 1.0.0. https://cristinahg.github.io/OCAPIS/

Hsu, C.-W., & Lin, C.-J. (2002). A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, *13*, 415–425.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in R*. Springer.

Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, *29*(1), 1–27. https://doi.org/10.1007/BF02289565

Kruskal, J. B. (1964b). Nonmetric multidimensional scaling: A numerical method. *Psychometrika*, *29*(2), 115–129. https://doi.org/10.1007/BF02289694

Kruskal, J. B. (1977). The relationship between multidimensional scaling and clustering. In J. Van Ryzin (Ed.), *Classification and clustering* (pp. 17–44). Academic Press.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Ling, H., & Jacobs, D. W. (2007). Shape classification using the inner-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *29*(2), 286–299. https://doi.org/10.1109/TPAMI.2007.41

Mair, P. (2018). *Modern psychometrics with* R. Springer.

Mair, P., Borg, I., & Rusch, T. (2016). Goodness-of-fit assessment in multidimensional scaling and unfolding. *Multivariate Behavioral Research*, *51*(6), 772–789. https://doi.org/10.1080/00273171.2016.1235966

Mair, P., Groenen, P. J. F., & De Leeuw, J. (2022). More on multidimensional scaling and unfolding in R: smacof version 2. *Journal of Statistical Software*. (Forthcoming)

McNally, R. J., Robinaugh, D. J., Wu, G. W. Y., Wang, L., Deserno, M. K., & Borsboom, D. (2015). Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Clinical Psychological Science*, *3*(6), 836–849. https://doi.org/10.1177/2167702614553230

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F. (2021). *e1071: Misc functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien [Computer software manual]*. R package version 1.7-7. https://CRAN.R-project.org/package=e1071

R Core Team. (2021). *R: A language and environment for statistical computing [Computer software manual]*. https://www.R-project.org/

Rothkopf, E. Z. (1957). A measure of stimulus similarity and errors in some paired associate learning. *Journal of Experimental Psychology*, *53*(2), 94–101. https://doi.org/10.1037/h0041867

Rusch, T., Mair, P., & Hornik, K. (2021). Cluster optimized proximity scaling. *Journal of Computational and Graphical Statistics*, *30*(4), 1156–1167. https://doi.org/10.1080/10618600.2020.1869027

Schölkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, P. L. (2000). New support vector algorithms. *Neural Computation*, *12*(5), 1207–1245. https://doi.org/10.1162/089976600300015565

Schwartz, S. H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries. *Advances in Experimental Social Psychology*, *25*, 1–62.

Shepard, R. N. (1963). Analysis of proximities as a technique for the study of information processing in man. *Human Factors*, *5*(1), 33–48. https://doi.org/10.1177/001872086300500104

Shye, S. (1998). Modern facet theory: Content design and measurement in behavioral research. *European Journal of Psychological Assessment*, *14*(2), 160–171. https://doi.org/10.1027/1015-5759.14.2.160

Steinwart, I., & Christmann, A. (2008). *Support vector machines*. Springer.

Van den Burg, G. J. J., & Groenen, P. J. F. (2016). GenSVM: A generalized multiclass support vector machine. *Journal of Machine Learning Research*, *17*, 1–42.

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605.

Vapnik, V. N. (2000). *The nature of statistical learning theory* (2nd ed.). Springer.

Waegeman, W., & Boullart, L. (2009). An ensemble of weighted support vector machines for ordinal regression. *International Journal of Computer Systems Science and Engineering*, *3*, 47–51.

Wish, M. (1967). A model for the perception of Morse code-like signals. *Human Factors*, *9*(6), 529–539. https://doi.org/10.1177/001872086700900603

Witten, D. M., & Tibshirani, R. (2011). Supervised multidimensional scaling for visualization, classification, and bipartite ranking. *Computational Statistics & Data Analysis*, *55*(1), 789–801. https://doi.org/10.1016/j.csda.2010.07.001