Learning as formation of low-dimensional representation spaces

Shimon Edelman
Center for Biological and Computational Learning
MIT E25-201, Cambridge, MA 02142, USA
edelman@ai.mit.edu

Nathan Intrator
School of Mathematical Sciences
Tel Aviv University, Tel Aviv 69978, Israel
nin@math.tau.ac.il

Abstract

Psychophysical findings accumulated over the past several decades indicate that perceptual tasks such as similarity judgment tend to be performed on a low-dimensional representation of the sensory data. Low dimensionality is especially important for learning, as the number of examples required for attaining a given level of performance grows exponentially with the dimensionality of the underlying representation space. Because of this *curse of dimensionality*, in shape categorization the high initial dimensionality of the sensory data must be reduced by a nontrivial computational process, which, ideally, should capture the intrinsic low-dimensional nature of families of visual shapes. We show how to make a connectionist system use class labels to learn a representation that fulfills this requirement, thereby facilitating shape categorization. Our results indicate that low-dimensional representations are best extracted in a learning task that combines discrimination and generalization constraints.

Introduction

The sophisticated behavior of biological cognitive systems is widely assumed to stem from their ability to learn from the environment, which leads to the formation of an internal representation of information pertinent to the task. Because learned representations can be employed in shaping the behavior in similar, and thus potentially related situations in the future (Shepard, 1987), similarity (Nosofsky, 1992; Medin et al., 1993) and its representation (Edelman, 1997) are central concerns in cognitive science.

In this note, we consider a core property of any representation of similarities among objects whose goal is efficient learning and generalization: *low dimensionality*. Because the link between the issues of similarity and of low-dimensional representation (LDR) is readily apparent in visual psychophysics, we concentrate on this area, rather than on other potential applications in cognition, ranging from vision to language and reasoning. Psychophysics, by definition, involves a relationship between the physical characteristics of a stimulus and the perceptual event it evokes. Now, for specific discrimination tasks, a natural framework for a physical description of various relationships — among them similarities — between the different possible stimuli is a low-dimensional metric space. In those cases, it is reasonable to expect that the representational system reflect the dimensional structure, the topology, and maybe even the metrics, of the stimulus space. We start, therefore, by examining the extent to which this expectation is fulfilled in a typical perceptual task: color perception.

A case study: color space

The central feature of the problem of computing the reflectance of a surface patch from measurements performed on its retinal image is that the expected solution (i.e., the reflectance function of the surface) resides, in principle, in an infinite-dimensional space: a potentially different value of reflectance may have to be specified for each of the infinite number of wavelengths of the incident light (D'Zmura and Iverson, 1997). Computationally, the recovery of surface reflectance in the face of possible variations in the illumination (itself a nominally infinite-dimensional quantity) is difficult enough because of the need to pry apart two multiplicatively combined functions, reflectance and illumination. The infinite dimensionality of these functions seems to suggest, further, that no finite set of measurements would suffice to support the recovery of surface reflectance. Nevertheless, human vision exhibits color constancy under a wide range of conditions (Beck, 1972), despite the small dimensionality of the neural color coding space (De Valois and De Valois, 1978); moreover, the dimensionality of the psychological (perceived) color space is also small (Boynton, 1978). In fact, both these color spaces are two-dimensional.[1]

Low-dimensional physiological color space. In human vision, there are only three kinds of different retinal cone types (R, G, B, in addition to the rods, whose spectral selectivity resembles that of the R cones). The question arises, therefore, how is it possible to recover the potentially infinite-dimensional spectral quantities using this measurement mechanism.

The solution to this paradox lies in the low dimensionality of the space of the *actual* surface reflectances and daylight illumination spectra.[2] This finding (Cohen, 1964; Judd et al., 1964) helps one understand why a small number of independent color-selective channels suffice to represent internally the rich world of color. The reason is simple: *the internal representation space can be low-dimensional, because the distal space happens to be low-dimensional.*

[1] An additional dimension in both cases is luminance. Color constancy requires simultaneous processing of several spatial locations, making the effective dimensionality for this task somewhat higher.

[2] Over 99% of the variance of Munsell chip reflectance functions can be accounted for using three basis functions, corresponding roughly to variations in intensity and in color-opponent $R-G$ and $B-Y$ channels (Cohen, 1964). Likewise, over 99% of the variance of daylight spectra can be accounted for by three principal components (Judd et al., 1964).

Low-dimensional psychological color space. As the dimensionality of the physiological coding space for color matches that of the universe of stimuli it is geared to respond to, it should not be surprising that the representation space fed by the color pathway is equally low-dimensional. A data processing tool that has proved to be exceptionally useful in the characterization of internal representation spaces is *multidimensional scaling*, or MDS. This technique is derived from the observation that the knowledge of distances among several points constrains the possible locations of the points (relative to each other) to a sufficient degree as to allow the recovery of the locations (i.e., the coordinates of the points) by a numerical procedure (see Shepard, 1980, for a review). In the processing of color perception data, the configuration derived by MDS is invariably found to be approximately circular (placing violet close to red), and to reside in two dimensions, one of which corresponds to the hue, and the other – to the saturation of the color (Boynton, 1978).

The exploration of the metric and the dimensional structure of psychological spaces has been boosted by the improvement of the metric scaling techniques and by the development of non-metric multidimensional scaling (Shepard, 1966; Kruskal, 1964). By 1980, a general pattern was emerging from a large variety of perceptual scaling experiments: the subject's performance in tasks involving similarity judgment or perception can be accounted for to a substantial degree by postulating that the perceived similarity reflects the metric structure of an underlying perceptual space, in which the various stimuli are represented as points (Shepard, 1980).[3]

Low-dimensional shape representation space

A series of psychophysical studies, originating with (Shepard and Cermak, 1973), suggest that the low-dimensional similarity space framework can be extended from the representation of basic perceptual qualities (such as colors) to that of complex shapes. In these studies, low-dimensional similarity patterns were imposed on families of stimuli, by exerting parametric control over the shape of each object. The low-dimensional similarity space has been recovered in each experiment by applying MDS to the response data of human subjects (Shepard and Cermak, 1973; Cortese and Dyre, 1996; Edelman, 1995; Cutzu and Edelman, 1996). Moreover, the locations of the stimuli in the MDS-derived space closely reflected their arrangement in the parametrically defined pattern imposed in each experiment. These two properties of the internal shape representation space — low dimensionality and preservation of distal similarity relationships — indicate that the human visual system routinely solves a formidable computational problem: massive dimensionality reduction.

Constraints on dimensionality reduction

Although empirical evidence for the low dimensionality of the psychological representation spaces has been accumulating steadily for decades, there is still a widespread tendency in psychology to overlook the computational problem presented by the derivation of low-dimensional representations from perceptual data. The main reason behind this is the mistaken assumption that the raw data available to the cognitive system reside in an immediately accessible low-dimensional space. For example, textbooks typically describe visual perception as the extraction of information from the *two-dimensional* retinal image, completely ignoring the fact that the immediate successor of the retinal space in the processing hierarchy is, in primates, a million-dimensional space spanned by the activities of the individual axons in the optic nerve.

Assuming that the information required by the system is present in this raw measurement space, one may wonder why the human visual system bothers to reduce dimensionality at all. A crucial theoretical consideration here has to do with *learning*. Specifically, learning from examples is computationally infeasible if it has to rely on high-dimensional representations. The reason for this is known as the *curse of dimensionality*: the number of examples necessary for reliable generalization grows exponentially with the number of dimensions (Bellman, 1961; Stone, 1982). Learnability thus necessitates dimensionality reduction.

The choice of the computational approach to dimensionality reduction is guided by two considerations. The first is the scale of the problem: shape representation in human vision requires reduction from tens and hundreds of thousands to just a few dimensions. The second consideration is preservation of a certain order of points corresponding to different objects, as they are mapped from the high-dimensional measurement space into the low-dimensional representation space: objects that are geometrically similar should be mapped to nearby locations (Edelman, 1997). Intuitively, then, the process of dimensionality reduction must preserve the topology of the set of stimuli (measurement-space points) pertinent to the task at hand.

Topology-preserving methods are especially useful for representing data known to reside in an intrinsically low-dimensional space (embedded in a high-dimensional measurement space). For instance, for color stimuli, there is a natural low-dimensional pattern of similarities that must be preserved (pink should be represented as closer to red than to green); furthermore, the objective (distal) reflectance and illuminant spaces are low-dimensional, as we have seen in the introduction. Likewise, in shape representation, the relevant distal spaces are low-dimensional (in any smooth measurement space, views of an object undergoing a transformation such as rotation, or a deformation such as morphing into another object, form low-dimensional manifolds).

In all these cases, objects to be represented may be visualized as points drawn on a sheet of rubber, which is then crumpled into a (high-dimensional) ball, as illustrated in Figure 1. The objective of a dimensionality-reducing mapping is to unfold the sheet and to make its low-dimensional structure explicit. If the sheet is not torn in the process, the mapping is topology-preserving; if, moreover, the rubber is not stretched or compressed, the mapping preserves the metric structure of the original space, and, hence, the original configuration (similarity pattern) of points.

The requirement that the mapping be an isometry is very restrictive: if it is to hold globally, the mapping must be linear. For local approximate isometry, any smooth and regular map-

[3]Some qualifications to this view are discussed in (Gregson, 1975). In particular, the metric model must be modified (Krumhansl, 1978; Edelman et al., 1996) to account for asymmetry and lack of transitivity of similarity judgments in some tasks (Tversky, 1977; Tversky and Gati, 1982).

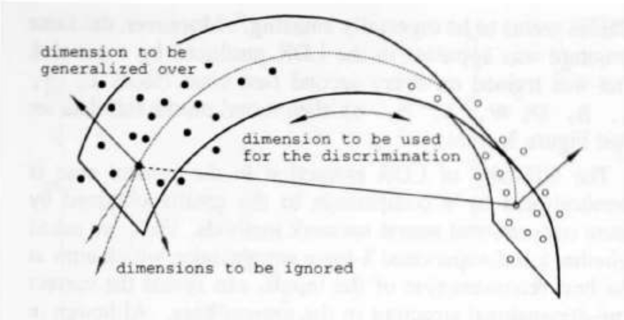

Figure 1: A schematic illustration of a problem space whose efficient representation requires nonlinear dimensionality reduction. The instances of the two classes cling to a low-dimensional manifold, embedded in a measurement space, whose dimensionality may run in the tens of thousands. As in discriminant analysis, some dimensions are crucial for distinguishing between the categories, while other dimensions must be downplayed, or collapsed. In the context of object recognition, the former may be the dimensions of object identity, and the latter – of object orientation. Standard discriminant analysis methods in multidimensional spaces are plagued by the presence of irrelevant dimensions; in this paper, we show that training with a combined objective of (1) discrimination among labeled categories known to reside within the manifold, and (2) explicit collapse of dimensions over which discrimination is to be generalized, leads to a reliable recovery of the target manifold, even when it is significantly curved (i.e., when the problem is highly nonlinear), and is embedded in a measurement space of nearly a thousand dimensions.

ping is sufficient.[4] Moreover, near linearity and smoothness are also *necessary* for topology preservation. This is good news, as far as the learnability of the mapping is concerned: a smooth mapping implies a small number of parameters to be learned. This, in turn, reduces the likelihood of overfitting and of poor generalization, which plague learning algorithms in high-dimensional spaces.

The oldest nonlinear method for topology-preserving dimensionality reduction is multidimensional scaling (MDS), originally developed in psychometrics as a method for the recovery of the coordinates of a set of points from measurements of the pairwise distances between those points. MDS can serve to reduce dimensionality if the points are embedded into a space of fewer dimensions than the original space in which interpoint distances were measured. The main problem with MDS, considered as a method for massive dimensionality reduction rather than for exploration of experimental data in applied sciences, is its poor scaling with dimensionality (Intrator and Edelman, 1996). The same problem arises in the various attempts to extend a popular tool for linear dimensionality reduction, principal component analysis (PCA), to handle nonlinear spaces.[5]

A number of learning methods for topology-preserving dimensionality reduction have been derived from the idea of a self-supervised auto-associative network (Elman and Zipser, 1988; DeMers and Cottrell, 1993; Demartines and Hérault, 1996). Because these methods are unsupervised, they extract representations that are not orthogonal to the irrelevant dimensions of the input space. As a result, these methods are less likely to find the target manifold (Intrator and Edelman, 1997), which is defined, to a large extent, by the measurement-space directions to which it is orthogonal; see Figure 1. Supervised approaches, based on joint optimization of discriminability and of topology preservation, are described in (Koontz and Fukunaga, 1972; Webb, 1995); these methods, which resemble MDS, suffer from the same poor scaling with the dimensionality.

A scheme for the extraction of low-dimensional representations

We now proceed to show that training with a combined objective of (1) discrimination among labeled objects known to reside within the target manifold, and (2) explicit collapse of dimensions, orthogonal to the manifold, over which discrimination is to be generalized, leads to a reliable recovery of the target manifold.

Solving the problem we chose to address — learning to recognize visual objects from examples — requires the ability to find meaningful patterns in series of images, or, in other words, in spaces of very high dimensionality. As mentioned above, dimensionality reduction in this task is greatly assisted by the realization that a low-dimensional solution, in fact, exists. The mere knowledge of its existence does not, however, automatically provide a method for computing a low-dimensional solution. To do that, the learning system must be biased towards solutions that possess the desirable properties — a task that is highly nontrivial in a high-dimensional space, because of the curse of dimensionality. Our method for dimensionality reduction effectively biases the learning system by combining multiple constraints via the use of an extensive set of class labels. The use of multiple class labels steers the low-dimensional representation to become invariant to those directions of variation in the input space that are irrelevant to classification; this is done merely by making class labels independent of these directions.

The extraction of a low-dimensional representation

As in the "bottleneck" approaches to dimensionality reduction (Cottrell et al., 1987; Leen and Kambhatla, 1994), we forced a classifier[6] to learn a set of class labels for input objects, while constraining the dimensionality of the representation — e.g., the number of hidden units in a 3-layer network — used by the classifier. Unlike in the standard methods, however, the classifier had to produce only the labels, rather than reconstruct the input patterns. This approach, therefore, constitutes a compromise between completely unsupervised and totally supervised methods in that it uses a label that individuates a given data item, but does not require information regarding the relationship between the different items, let alone the

[4] A discussion of such *quasiconformal* mappings in the context of shape representation can be found in (Edelman and Duvdevani-Bar, 1997).

[5] An example of such an approach is clustering followed by (local) PCA (Leen and Kambhatla, 1994).

[6] Experimentation with various architectures, including multilayer perceptrons and radial basis function networks, yielded similarly encouraging results; see (Intrator and Edelman, 1997) for details.

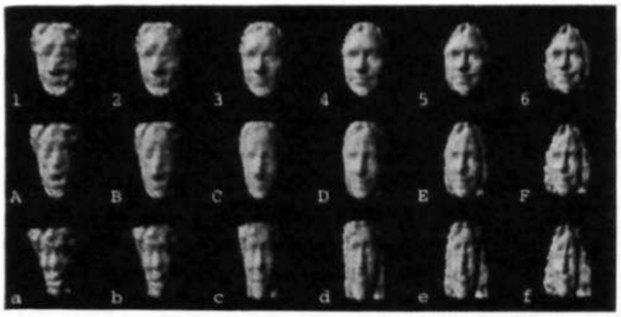

Figure 2: Some of the images from the FACES data set. *Top:* the 18 heads obtained by placing a 3×6 grid in the space of the two leading principal components of the original nine heads. *Bottom:* the 7 views of the rightmost head in the top row above; the views differ by 3° steps of rotation in depth, summing up to a total difference of 18°. Prior to classification, the images, originally of size 400×400, were reduced to 784 dimensions by cropping the background and by correlation with a 49×16 bank of filters (the exact spatial profile of these filters turned out to be unimportant; Gaussian filters did just as well as opponent center-surround ones).

complete reconstruction of the data as in the bottleneck autoencoder systems.

The ability of this method to discover simple structure embedded in a high-dimensional measurement space was demonstrated on a face data set, in which the extraction of the LDR (low-dimensional representation) requires a highly nonlinear transformation on the measurement space.[7] At the basis of this data set lies a two-dimensional parametric representation space, in which 18 classes of faces are placed on a regular 3×6 grid; an additional parametric dimension, orthogonal to the first two, models the within-class variation (see Figure 2). To impose a distinctive low-dimensional structure on the set of faces, we followed the simple approach of common parameterization by principal component analysis (PCA). This was done by starting with a set of nine 3D laser scans of human heads, and by embedding the 3×6 grid in the 2D space spanned by the two leading "eigenheads" obtained from the data by PCA. Each of the 18 heads derived by PCA from the original scanned head data was piped through a graphics program, which rendered the head from seven viewpoints, obtained by stepping the (simulated) camera in 3° rotation steps around the midsagittal axis.

Results

The application of the label-based method led to a good recovery of the relevant low-dimensional description of the FACES data set (see Figure 3). The performance of this method in recovering the row/column parametric structure of the 18 classes seems to be especially amazing.[8] Moreover, the same structure was apparent in the LDR produced by a network that was trained on every second face class (faces `1, 3, 5, B, D, F, a, c, e`), then tested on the full data set (see Figure 3, right).

The difficulty of LDR extraction in the present case is demonstrated by a comparison to the results obtained by more conventional neural network methods. First, we asked whether a self-supervised 3-layer autoencoder, which aims at the best reconstruction of the inputs, can reveal the correct low-dimensional structure in the present case. Although in the linear case such networks do quite well, essentially by extracting the principal components of the data (Elman and Zipser, 1988), the performance on the FACES data was poor (the network consistently converged to the mean of the data), presumably due to the nonlinearity introduced by the imaging step. Second, we experimented with a 5-layer nonlinear bottleneck autoencoder (Leen and Kambhatla, 1994), which, likewise, performed poorly. The outcome of this experiment showed that self-supervised dimensionality reduction cannot recover a good LDR in the present case, illustrating the importance of guidance provided by the class labels. Third, we tested a modified version of our method, in which the classifier was not trained to ignore the direction orthogonal to the target manifold (cf. Figure 1; this was done by training on the 72 face view labels, instead of the 18 face identity labels). Here, too, the LDR was poor, underscoring the importance of guidance provided by an explicit specification of the dimension to be collapsed.

Discussion

The research program which led to the results outlined above is motivated by the notion that a good representation of the visual world is, first and foremost, a low-dimensional representation. We described a family of methods that can map a high-dimensional data set into a low-dimensional space, which is topologically a good approximation of a nonlinear manifold present in the original measurements. The property of topology preservation, shared by all the methods we considered, appears to be due to the smooth nature of the mapping they realize (Intrator and Edelman, 1997; Edelman and Duvdevani-Bar, 1997). This property alone, however, is insufficient to ensure the extraction of the correct manifold: control experiments indicate the importance both of the use of class labels (which help define the tangent space to the manifold, as it is illustrated in Figure 1), and of the stipulation of the generalization set for the stimuli (which defines the normal to the manifold). Moreover, the separate definition of the normal for each class (i.e., the specification of the viewpoint-induced variation for each of the 18 faces) is important for the recovery of nonlinear (curving) manifolds, in which the direction of the normal changes from point to point.

[7] We tested this method also on another data set, consisting of parameterized fractal images (Intrator and Edelman, 1996).

[8] To assess the quality of the LDR, we defined a dichotomy task, in which nine of the 18 faces (labeled in Figure 2 as `2, 1, A, a, b, c, d, D, E`) were attributed to one class, and the other nine faces – to another class. In this task, the LDRs recovered by our method consistently supported better performance than the raw images.

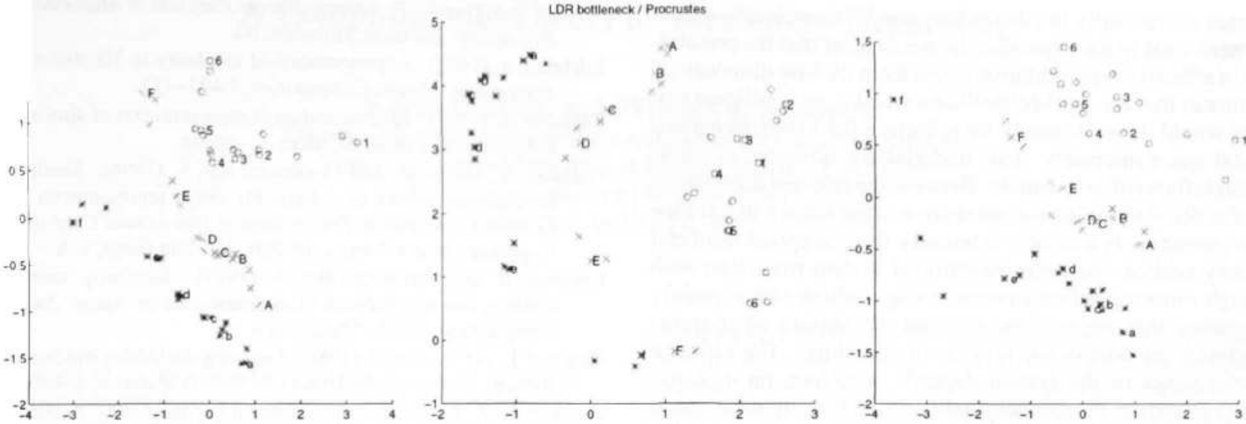

Figure 3: FACES data set, dimensionality reduction by a bottleneck multilayer perceptron (MLP); the plots show the locations of the 18×3 test stimuli in the space spanned by the activities of the units residing in a hidden layer (18 faces times 3 test orientations per face). *Left:* results obtained with a 3-layer MLP with 13 units in the middle hidden layer, trained for 20,000 epochs on the 18-way classification task. The low-dimensional representation proved to be a good substrate for solving classification tasks on which the system has not been trained: the error rate on a nonlinear dichotomy involving the 18 classes was 0.02, compared to 0.07 obtained by a system trained specifically on that dichotomy, but using the raw multidimensional representation; see (Intrator and Edelman, 1997) for details. *Middle:* results for a 5-layer bottleneck MLP with 2 hidden units in the middle hidden layer, trained on the 18-way classification task. The test dichotomy error rate was 0.1, compared to 0.29 on the raw data. *Right:* results obtained with a 3-layer MLP network with nine hidden units, trained on every second of the 18 classes that comprise the problem space (the nine classes used for training and the nine omitted classes formed a checkerboard pattern). Note that all 18 classes – both familiar ones and those not seen by the system – are in a topology-preserving formation. For this representation, the error rate on a nonlinear dichotomy involving the 18 classes was 0.14, compared to 0.28 obtained by a system trained specifically on that dichotomy, but using the raw multidimensional representation. Remark: in the left and the right plots, multidimensional scaling was used to visualize the representation spaces (which were, nominally, 13- and 9-dimensional, respectively); in the middle plot, the 2-dimensional hidden-unit space of the 5-layer network is plotted directly.

Implications

The computational feasibility of learning a representation that is both low-dimensional and similarity-preserving may be taken as further support for the attempts to make similarity a central explanatory concept in psychology. One such attempt is described in Shepard's (1987) paper, which appeared on the tri-centennial anniversary of the publication of Newton's *Principia*. In that paper, Shepard proposed a law of generalization that tied the likelihood of two stimuli evoking the same response to the proximity of the stimuli in a psychological representation space — the same space that so persistently turned out to be low-dimensional in the dozens of experiments surveyed in (Shepard, 1980; Shepard, 1987), as well as in the more recent works on shape representation (Edelman, 1995; Cortese and Dyre, 1996; Cutzu and Edelman, 1996).

The significance of having a similarity-preserving low-dimensional space as a substrate for representation is twofold. First, the introduction of the notion of a *similarity space* puts novel stimuli on an equal footing with familiar ones: a point corresponding to a novel stimulus is always located *somewhere* in the representation space; all one has to do is characterize its location with respect to the familiar points. The visual system literally never encounters the same stimulus twice: there are always variations in the viewing conditions such as illumination; objects look different from different viewpoints; articulated and flexible objects change their shape. Mere memory for past stimuli, faithful and extensive as it may be, is, therefore, a poor guide for behavior. In contrast, a suitable (i.e., similarity-preserving) representation space can help the system deal with objects for which no memory traces are available (cf. Figure 3, right). In such a space, proximity is a reliable guide for generalization.[9]

The second important trait of the representation space common to a range of stimuli in a given task — its low dimensionality — became gradually clear only recently, with the emergence of formal approaches to the quantification of complexity of learning problems. Whereas in some perceptual tasks (such as color vision) low dimensionality of the representation stems naturally from the corresponding low dimensionality of the stimulus space, in other tasks (notably, in object shape recognition) a computationally convenient basis for low-dimensional shape representation is yet to be developed. In the meanwhile, a useful common low-dimensional parameterization of shapes belonging to certain categories can be achieved via principal component analysis, as it was done here for the human heads; cf. (Atick et al., 1996).

Summary

To conclude, we propose that the development of systems capable of representing the world is governed by the following unifying principle: various aspects of the world are repre-

[9] Shepard's (1987) work shows that the validity of proximity as the basis for generalization is universal, and can be derived from first principles.

sented successfully insofar as they can be expressed in a low-dimensional space. Specifically, we suggest that the possibility of effective representation stems from the low-dimensional nature of the real-world classification tasks: an intelligent system would do well merely by reflecting the low-dimensional distal space internally. This undertaking, however, is not as straightforward as it sounds. Because the relevant dimensions of the distal stimulus variation are neither known in advance nor immediately available internally, the perceptual front end to any sophisticated representational system must start with a high-dimensional measurement stage, whose task is mainly to assure that none of the relevant dimensions of stimulus variation are lost in the process of encoding. The ultimate performance of the system depends, therefore, on its capability to reduce the dimensionality of the measurement space back to an acceptable level, which would be on par with that of the original, presumably low-dimensional, distal stimulus space.

Acknowledgments

Thanks to P. Dayan and J. Tenenbaum for useful suggestions, and to R. Goldstone, J. Hochberg, and P. Schyns for comments on this project.

References

Atick, J. J., Griffin, P. A., and Redlich, A. N. (1996). The vocabulary of shape: principal shapes for probing perception and neural response. *Network*, 7:1–5.

Beck, J. (1972). *Surface Color Perception*. Cornell University Press, Ithaca, NY.

Bellman, R. E. (1961). *Adaptive Control Processes*. Princeton University Press, Princeton, NJ.

Boynton, R. M. (1978). Color, hue, and wavelength. In Carterette, E. C. and Friedman, M. P., editors, *Handbook of Perception*, volume V, pp 301–347. Academic Press, New York, NY.

Cohen, J. (1964). Dependency of the spectral reflectance curves of the Munsell color chips. *Psychonomic Sciences*, 1:369–370.

Cortese, J. M. and Dyre, B. P. (1996). Perceptual similarity of shapes generated from Fourier Descriptors. *Journal of Experimental Psychology: Human Perception and Performance*, 22:133–143.

Cottrell, G. W., Munro, P., and Zipser, D. (1987). Learning internal representations from gray-scale images: An example of extensional programming. In *Ninth Annual Conference of the Cognitive Science Society*, pp 462–473, Hillsdale. Erlbaum.

Cutzu, F. and Edelman, S. (1996). Faithful representation of similarities among three-dimensional shapes in human vision. *Proceedings of the National Academy of Science*, 93:12046–12050.

De Valois, R. L. and De Valois, K. K. (1978). Neural coding of color. In Carterette, E. C. and Friedman, M. P., editors, *Handbook of Perception*, volume V, pp 117–166. Academic Press, New York, NY.

Demartines, P. and Hérault, J. (1996). Curvilinear component analysis: a self-organizing neural network for non linear mapping of data sets. Submitted to IEEE Transaction on Neural Networks.

DeMers, D. and Cottrell, G. (1993). Nonlinear dimensionality reduction. In Hanson, S. J., Cowan, J. D., and Giles, C. L., editors, *Advances in Neural Information Processing Systems 5*, pp 580–587. Morgan Kaufmann.

D'Zmura, M. and Iverson, G. (1997). A formal approach to color constancy: the recovery of surface and light source spectral properties using bilinear models. In Dowling, C., Roberts, F., and Theuns, P., editors, *Recent Progress in Mathematical Psychology*. Erlbaum, Hillsdale, NJ.

Edelman, S. (1995). Representation of similarity in 3D object discrimination. *Neural Computation*, 7:407–422.

Edelman, S. (1997). Representation is representation of similarity. Behavioral and Brain Sciences, to appear.

Edelman, S., Cutzu, F., and Duvdevani-Bar, S. (1996). Similarity to reference shapes as a basis for shape representation. In Cottrell, G. W., editor, *Proceedings of 18th Annual Conf. of the Cognitive Science Society*, pp 260–265, San Diego, CA.

Edelman, S. and Duvdevani-Bar, S. (1997). Similarity, connectionism, and the problem of representation in vision. *Neural Computation*, 9:701–720.

Elman, J. L. and Zipser, D. (1988). Learning the hidden structure of speech. *Journal of the Acoustical Society of America*, 4(83).

Gregson, R. A. M. (1975). *Psychometrics of similarity*. Academic Press, New York.

Intrator, N. and Edelman, S. (1996). How to make a low-dimensional representation suitable for diverse tasks. *Connection Science*, 8:205–224.

Intrator, N. and Edelman, S. (1997). Learning low dimensional representations of visual objects with extensive use of prior knowledge. in press.

Judd, D. B., MacAdam, D. L., and Wyszecki, G. (1964). Spectral distribution of typical daylight as a function of correlated color temperature. *Journal of the Optical Society of America*, 54:1031–1040.

Koontz, W. L. G. and Fukunaga, K. (1972). A nonlinear feature extraction algorithm using distance information. *IEEE Trans. Comput.*, 21:56–63.

Krumhansl, C. L. (1978). Concerning the applicability of geometric models to similarity data: the interrelationship between similarity and spatial density. *Psychological Review*, 85:445–463.

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

Leen, T. K. and Kambhatla, N. (1994). Fast non-linear dimension reduction. In Cowan, J. D., Tesauro, G., and Alspector, J., editors, *Advances in Neural Information Processing Systems*, volume 6, pp 152–159. Morgan Kauffman, San Francisco, CA.

Medin, D. L., Goldstone, R. L., and Gentner, D. (1993). Respects for similarity. *Psychological Review*, 100:254–278.

Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology*, 43:25–53.

Shepard, R. N. (1966). Metric structures in ordinal data. *J. Math. Psychology*, 3:287–315.

Shepard, R. N. (1980). Multidimensional scaling, tree-fitting, and clustering. *Science*, 210:390–397.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science*, 237:1317–1323.

Shepard, R. N. and Cermak, G. W. (1973). Perceptual-cognitive explorations of a toroidal set of free-form stimuli. *Cognitive Psychology*, 4:351–377.

Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Annals of statistics*, 10:1040–1053.

Tversky, A. (1977). Features of similarity. *Psychological Review*, 84:327–352.

Tversky, A. and Gati, I. (1982). Concerning the applicability of geometric models to similarity data: the interrelationship between similarity and spatial density. *Psychological Review*, 89:123–154.

Webb, A. R. (1995). Multidimensional-scaling by iterative majorization using radial basis functions. *Pattern Recognition*, 28:753–759.