Learning as extraction of low-dimensional representations

Shimon Edelman
Dept. of Applied Mathematics and Computer Science
The Weizmann Institute of Science
Rehovot 76100, Israel
edelman@wisdom.weizmann.ac.il

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

November 18, 1996

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. In this chapter, we argue that, whereas many perceptual problems are tractable precisely because their intrinsic dimensionality is low, the raw dimensionality of the sensory data is normally high, and must be reduced by a nontrivial computational process, which, in itself, may involve learning. Following a survey of computational techniques for dimensionality reduction, we show that it is possible to learn a low-dimensional representation that captures the intrinsic low-dimensional nature of certain classes of visual objects, thereby facilitating further learning of tasks involving those objects.

1 Introduction

It is widely assumed that the sophisticated behavior of biological cognitive systems is due to their ability to learn from the environment, and, furthermore, that a direct consequence of learning is 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, representation is a central concept in the study of learning, as it is in other fields of cognitive science.

The very universality of the idea of representation, which makes it equally useful in connectionist and in symbolic accounts of cognition, may suggest that a general theory of representation is likely to be a multifaceted, loosely coupled collection of domain-specific subtheories. We contend, however, that it is possible to identify certain core properties that any representation of the world must possess to be able to support efficient learning and learning-related behavior. Specifically, we believe that representations aimed at capturing similarity — itself the basis for generalization in learning (Shepard, 1987) — must be *low-dimensional*.

The link between the issues of similarity and of low-dimensional representations (LDRs) becomes apparent when one considers problems that arise in visual psychophysics. By definition, such problems involve a relationship between the physical characteristics of a stimulus and the perceptual event it evokes. Now, in many situations, 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. In the remainder of this section, we examine the extent to which this expectation is fulfilled in a typical perceptual task (namely, color perception), and discuss the computational implications of these findings.

1.1 A case study: color spaces

A paradigmatic perceptual task the understanding of which requires dealing with issues of dimensionality is the perception of color. Consider the problem of computing the reflectance of a surface patch from measurements performed on its retinal image. The central feature of this problem is that the expected solution (i.e., the reflectance function of the surface) resides, in principle, in an infinite-dimensional space, because a potentially different (in the worst case, random) value of reflectance may have to be specified for each of the infinite number of wavelengths of the incident light (D'Zmura and Iverson, 1996). Furthermore, the spectral content of the illumination (which is confounded with the reflectance function multiplicatively, and which must be discounted to allow the computation of the reflectance) is also potentially infinite-dimensional, for the same reason.

In human vision, the recovery of surface reflectance in the face of possible variations in the illumination is known as color constancy. Computationally, the achievement of color constancy 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 set of measurements (short of an infinite and therefore an infeasible one) 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]

1.1.1 Low-dimensional physiological color space

In human vision, there are three kinds of different retinal cone types (R, G, B; in addition, there are the rods, whose spectral selectivity resembles that of the R cones). The hue information contained in the cone responses is further transformed on its way to the brain by combining these into two channels: $R - G$ and $B - Y$, where Y denotes $R + G$. The question arises, therefore, how is it possible to recover the potentially infinite-dimensional spectral quantities using this measurement mechanism.[2]

The solution to this paradox is made possible by the finite (in fact, low) dimensionality of the space of the *actual* surface reflectances. This observation has been quantified by Cohen (1964), who showed that over 99% of the variance in the set of Munsell chip reflectance functions could be accounted for using just three basis functions (corresponding roughly to variations in intensity and in color-opponent $R-G$ and $B-Y$ channels). The space of illuminations likely to be encountered in nature appears to be equally low-dimensional: a principal component analysis of 622 measurements of daylight illumination (carried out at different times of day) showed that over 99% of the variance can be accounted for by as few as three principal components (Judd et al., 1964).

The findings of Cohen and of Judd et al. help one understand why a small number of independent color-selective channels suffice to represent internally most of the richness of the world of color.[3] 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 that of luminance. It should be noted that color constancy requires simultaneous processing of more than one spatial location, so that the effective dimensionality of the input to the constancy mechanism is slightly higher than two.

[2] There are several dedicated color channels (as well as a luminance channel) *for each location* in the central visual field, so that the dimensionality of the measurement system is much higher; moreover, the number of dimensions per unit solid angle varies with retinal eccentricity, being the highest around the fovea. In the present discussion, however, we are only concerned with the applicability of this system to the estimation of color, not with its spatial resolution.

[3] All of it, for all we know; any pair of colors that are metameric with respect to our color vision are indistinguishable for us. Note that metamery (or like representations for unlike stimuli) may also occur when the illumination conditions are radically different from those which our visual system has evolved to tolerate (try to distinguish between US 1 cent and 10 cent coins under an orange sodium street lamp).

1.1.2 Low-dimensional psychological color space

In the preceding section we have seen that the physiological coding space for color is low-dimensional, and that its dimensionality matches that of the universe of stimuli it is geared to respond to. It should not be surprising, therefore, that the representation space fed by the color coding system is equally low-dimensional. Note, however, that the question of the dimensionality of the perceived color space belongs to psychology, not physiology. The only means for its resolution lies, therefore, in processing the responses of observers to color stimuli.

A data processing tool that proved to be exceptionally useful in the characterization of internal representation spaces, including that of color, 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 (Young and Householder, 1938). Assuming that the perceived similarities (that is, inverse distances, or proximities) among stimuli such as colors determine the responses made to those stimuli, one can process the responses by MDS, and examine the dimensionality of the resulting configuration of the points and the relative locations of the points. The assumption of the orderly relationship between the measured proximities and those derived from the resulting configuration is verified in the process, by the success of the MDS procedure, as manifested in the low stress (which is the cumulative residual discrepancy between those two quantities, computed over all the points). 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 (Shepard, 1962; Boynton, 1978).

1.2 Implications

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 in the early 1960's (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 directly reflects the metric structure of an underlying perceptual space, in which the various stimuli are represented as points (Shepard, 1980).[4]

[4]The metric model of the system of internal representations is not always directly applicable, as shown by asymmetry and lack of transitivity of similarity judgments that can be obtained under a range of conditions (Tversky, 1977). A recent proposal for a reconciliation of the feature contrast theory derived from these results with the metric perceptual scaling theory is described in (Edelman et al., 1996).

This pattern has not escaped the attention of theoretical psychologists. In a paper which appeared on the tri-centennial anniversary of the publication of Newton's *Philosophiae Naturalis Principia Mathematica*, and was motivated by a quest for psychological laws that would match those of mechanics, Shepard (1987) 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 experiments surveyed in (Shepard, 1980).

The significance of Shepard's insight is twofold. First, the introduction of the notion of a psychological 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 great importance of generalization stems from the fact that 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 representation space can help the system concentrate on the relevant features of the stimulus, which, presumably, remain invariant.[5] In such a space, proximity is a reliable guide for generalization. Shepard's (1987) work shows that the validity of proximity as the basis for generalization is universal, and can be derived from first principles.

The second reason behind the importance of having a common space for the representation of a range of perceptual qualities in any given task has to do with the low dimensionality of such a space. This point 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) the situation is less clear, although there are some indications that a useful common low-dimensional parameterization of diverse shapes can be achieved (see Figure 1).

In the case of object recognition, it is tempting to argue that one should use the multidimensional signal as is, because the shape information that the visual system needs is certainly present there: "The photoreceptors are [...] necessarily capable of coding, by their population response, any conceivable stimulus. Why are subsequent populations needed?" (Desimone and Ungerleider, 1989, p.268).[6] We now know that this approach to representation is untenable, as far as *learning* to recognize objects from examples is concerned. The reason for this is related to the notion of the *curse of dimensionality*: the number of examples necessary for reliable generalization grows exponen-

[5]The issue of invariant feature spaces is beyond the scope of the present discussion, which focuses on dimensionality.

[6]Desimone and Ungerleider meant this question to be rhetorical; representations of visual stimuli in the higher cortical areas are clearly different from those at the retinal level.