Forbidden Magnification? I.

Abha Jain and Erzsébet Merényi*
Rice University, Electrical and Computer Engineering Department, MS-380

6100 Main Street, Houston, TX 77005, U.S.A.
e-mail: {abha,erzsebet}@rice.edu

Abstract. This paper presents some interesting results obtained by the algorithm by Bauer, Der and Hermann (BDH) [1] for magnification control in Self-Organizing Maps. *Magnification control* in SOMs refers to the modification of the relationship between the probability density functions of the input samples and their prototypes (SOM weights). The above mentioned algorithm enables explicit control of the magnification properties of a SOM, however, the available theory restricts its validity to 1-D data or 2-D data when the stimulus density separates. This discourages the use of the BDH algorithm for practical applications. In this paper we present results of careful simulations that show the scope of this algorithm when applied to more general, "forbidden" data. We also demonstrate the application of negative magnification to magnify rare classes in the data to enhance their detectability.

1 Introduction

One theoretically interesting and powerful data analysis aspect of SOMs is the so called *map magnification*. The following power law relates the density of weights in the input space $Q(\underline{w})$ to the *pdf* $P(\underline{w})$ of the input samples,

$$Q(\underline{w}) = cP(\underline{w})^{\alpha} \tag{1}$$

where α is the *magnification exponent* and c is a constant [1]. Some values of α are associated with particular quantization or information theoretical properties [1]. A map with $\alpha = 1$ maximizes information theoretic entropy. $\alpha = 1/3$ for 1-D data corresponds to the minimum mean squared error quantization case. $\alpha < 0$ enables better categorization by enlarging response areas for low-frequency inputs. This is potentially useful for making discoveries as it would enhance the detectability of rare classes. Kohonen's SOM algorithm (KSOM) [2] achieves $\alpha = 2/3$ (under certain conditions) [3]. This value of α is optimal in neither minimum distortion nor maximum entropy sense. A SOM variant called Conscience algorithm [4] can induce $\alpha = 1$, but not any other value. BDH enables explicit control of magnification by using adaptively adjusted local learning rates. The available theory suggests that the algorithm will successfully induce the intended value of α only for:

Request color reprint by email.

*Authors are partially supported by the Applied Information Systems Research Program of NASA, Office of Space Science, NAG9-10432.

- 1-D input data
- 2-D data, $\underline{v} = (v_1, v_2)$, if and only if
  $p_{\underline{v}}(\underline{v}) = p_{v_1}(v_1)p_{v_2}(v_2)$ (i.e., the *pdf* factorizes into the marginals)

Careful simulations were carried out to observe the performance of the BDH algorithm on data with and without the properties listed above. In the next two sections, results for 1-, 2- and higher dimensional data are discussed.

2 Simulations for 1- and 2-D data

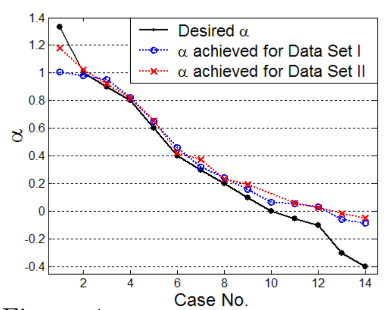

Figure 1: *Results of magnification control on 1-D data. Original figure is in color. Download paper from http://www.ece.rice.edu/~erzsebet/papers/esann04-1.pdf*

First we confirmed that BDH works well for 1-D data, by simulations similar to those in [1] for two 1-D data sets (Figure 1). Data Set I is $p(x) = 2x, x\varepsilon[0, 1]$ and Data Set II is $p(x) = 3x^2, x\varepsilon[0, 1]$. Then we investigated the performance of BDH on data for which the supporting theory does not guarantee success. Data with different amounts of correlations (ρ) between the two dimensions were generated. Several values of α were induced. In each case, α achieved by the map was calculated by a histogram based method used in [5] and compared to the desired α.

I Data independent in the two dimensions: $p_{\underline{v}}(\underline{v}) = p_{v_1}(v_1)p_{v_1}(v_2)$. This data set was generated according to the following *pdf*:

$$p_{v_1} = 2v_1, \quad v_1 \in [0, 1] \quad \texttt{and} \quad p_{v_2} = 2v_2, \quad v_2 \in [0, 1]$$
$$p_{\underline{v}}(\underline{v}) = 2v_1v_2, \quad v_1, v_2 \in [0, 1] \tag{2}$$

BDH on this data set achieved α values close to the desired α (Figure 2).

II Weakly correlated data in two dimensions: $\rho \ll 1$. The data consists of 2-D samples, $\underline{v} = (v_1, v_2)$ of two kinds. One is such that $v_2 = v_1 + n$ and the other is $v_2 = -v_1 + n$, where $n = \mathcal{N}(0, 0.0625)$. v_1 and v_2 are weakly correlated with the correlation coefficient $\rho_{v_1v_2} = 0.0044$. From Figure 3 it can be seen that α achieved and desired α are almost equal at $\alpha = 1$ and the two values differ increasingly (but in a predictable manner) as α decreases. This is a stronger result than available from theory, as the theory only guaranteed success if and only if v_1 and v_2 were independent.

III Data highly correlated in two dimensions: $\rho \approx 1$. This data consists of 2-D samples, $\underline{v} = (v_2, v_1)$, such that $v_2 = v_1 + n$, where $n = \mathcal{N}(0, 0.25)$. The correlation coefficient is $\rho_{v_1v_2} = 0.9026$. In this strongly correlated case, even though α achieved by the map differs from the desired α, there is a clear observable trend that the achieved values of α are systematically

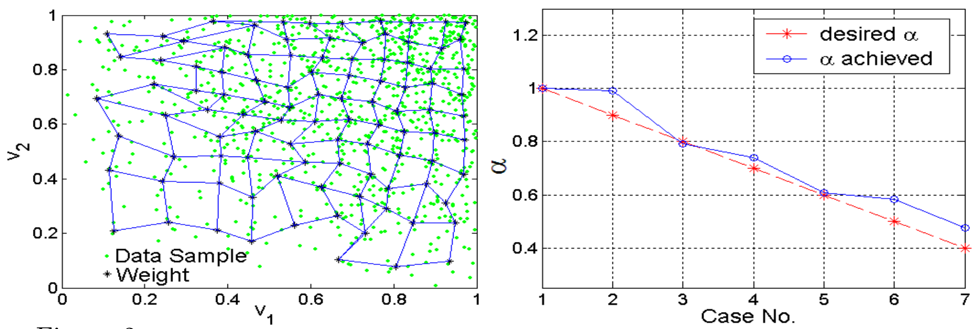

Figure 2: ***Left:*** *2-D input data samples (small dots) and distribution of SOM weights (larger dots), resulting from BDH with* $\alpha = 0.6$ *after* $2,000,000$ *steps. Weights adjacent in the SOM are connected. 2-D input samples* $\underline{v} = (v_1, v_2)$ *are such that* v_1 *and* v_2 *are independent.* ***Right:*** *Comparing* α *achieved to* α *desired, the discrepancies are largely due to the fact that the theoretical results are asymptotic and we only have a finite number of PEs (100).*

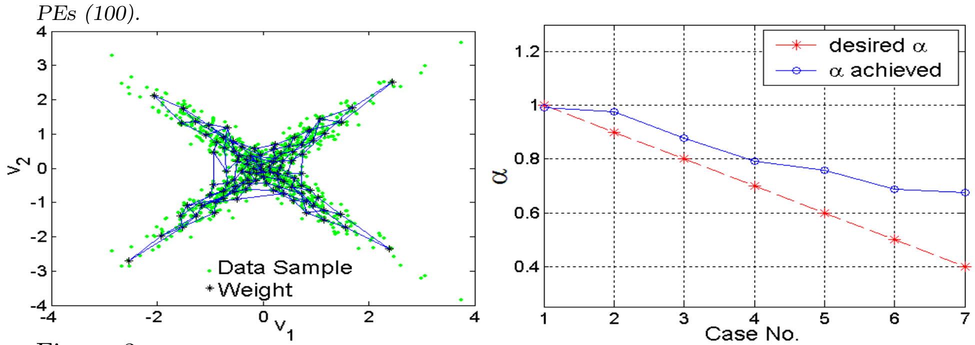

Figure 3: ***Left:*** *2-D input data samples (small dots) and distribution of SOM weights (larger dots), resulting from BDH with* $\alpha = 1.0$ *after* $2,000,000$ *steps. Weights adjacent in the SOM are connected. 2-D input samples* $\underline{v} = (v_1, v_2)$ *are such that,* $\rho_{v_1 v_2} \ll 1$*.* ***Right:*** *The difference between* α *achieved and* α *desired increases in a predictable manner as* α *decreases from 1.*

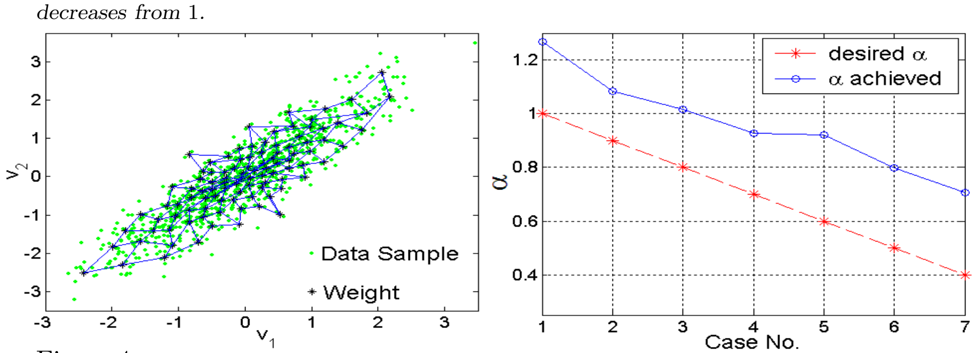

Figure 4: ***Left:*** *2-D input data samples (small dots) and distribution of SOM weights (larger dots), resulting from BDH with* $\alpha = 1.0$ *after* $1,000,000$ *steps. Weights adjacent in the SOM are connected. 2-D input samples* $\underline{v} = (v_1, v_2)$ *are such that* $\rho_{v_1 v_2} \approx 1$*.* ***Right:*** *Comparing* α *achieved to* α *desired, there is a clear observable trend in the differences between desired and achieved values of* α*. The differences are more or less constant. Original figure is in color. Download paper from http://www.ece.rice.edu/∼erzsebet/papers/esann04-1.pdf*

decreasing, following the desired values with a more or less constant shift (Figure 4). This is again a stronger result than the theory provides and encourages further investigation of BDH for real data, including higher dimensional data.

In the next section, some interesting results of applying BDH to higher dimensional data are discussed.

3 Simulations for higher dimensional data

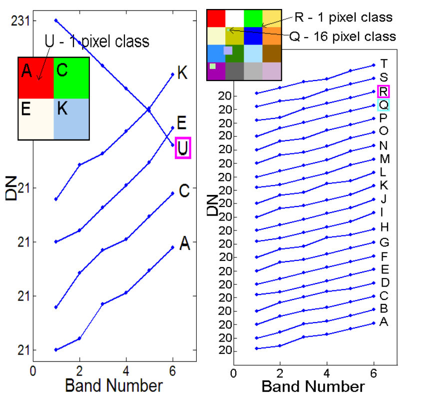

Figure 5: *Two 6-D Data Sets - Images and spectral signatures.* ***Left:*** *5-class Data Set.* ***Right:*** *20-class Data Set. Original figure is in color. Download paper from http://www.ece.rice.edu/∼erzsebet/papers/esann04-1.pdf*

Our methodology to examine the performance of the BDH algorithm on any 1-and 2-D data sets was to induce a certain value of α, evaluate the α achieved by the map and compare the two. However, evaluation of α is not an easy task in general, especially, if the input *pdf* is unknown (as is most commonly the case). The evaluation of α involves the estimation of the *pdf* of the data and of the weights. So far, we were using a histogram based method for α evaluation. This method becomes inapplicable for high-D data as the number of samples required for *pdf* estimation increases exponentially with dimensionality. So for higher dimensional cases, we evaluate the performance of the algorithm indirectly: by observing the resulting map.

Finding rare classes in a data set is a challenging task. Input classes with rare occurrence find little or no representation in the map when KSOM is used. Application of BDH with $\alpha < 0$ would result in negative magnification: classes which are rare in the input will be magnified in the map. This is a promising technique for detection of rare classes. In this section, we will look at two instructive cases in which BDH with negative α was applied on two 6-D synthetic data sets.

I Data set with 5 classes: This data set is a 128×128 pixel image where each pixel has a 6-D vector associated with it. It has 5 classes, whose signatures are shown in Figure 5 (left). Class 'U' is a rare class with only 1 data point of this kind. The rest of the classes have 4096 or 4095 data points each. In this data set, correlation coefficients between the different dimensions range from $0.004(\rho_{v_2 v_3})$ to $0.9924(\rho_{v_3 v_6})$, which renders this one of the "forbidden" cases for application of BDH. When KSOM is used, the rare class is represented by only one Processing Element (PE)(Figure 6, left). BDH with $\alpha = -0.8$, magnifies the rare class in the map: it is represented by 10 PEs (Figure 6, right).

II Data set with 20 classes: This data set is similar to the 5-class image except it has 20 classes, as shown in Figure 5 (right). Two of the classes, marked 'R' and 'Q' are relatively rare, with only 1 and 16 data points, respectively. Correlation coefficients between the different dimensions range from $0.0081(\rho_{v_5 v_4})$

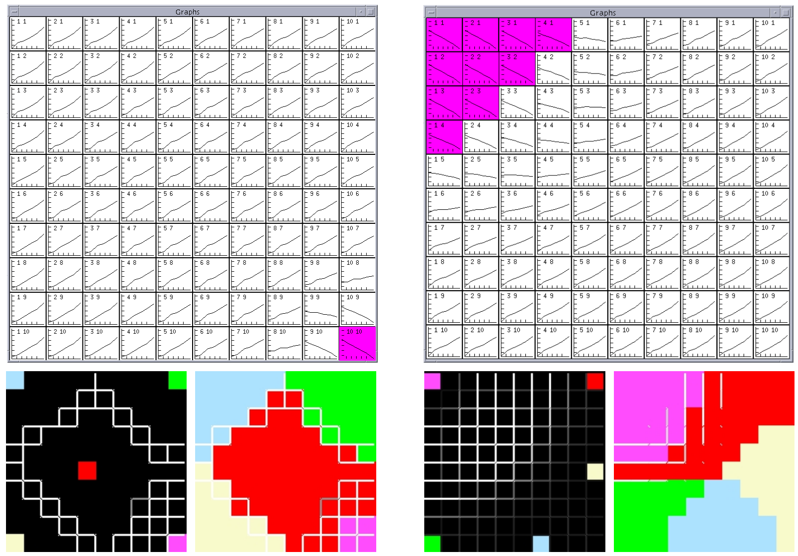

Figure 6: *Results of clustering of the* 5*-class data set.* ***Left:*** *Using KSOM. Top) Weight vectors in the* 10×10 *SOM. Only* 1 *PE represents the rare class 'U'. Bottom Left) Clusters identified in the map (the darker the fence between two PEs, the smaller is the difference between the corresponding weights and vice-versa). Bottom Right) This figure shows which class each weight vector is closest to, which complements the information on the left.* ***Right:*** *Using BDH with* $\alpha = -0.8$*. Top) Weight vectors in the* 10×10 *SOM. The rare class 'U' is now represented by* 10 *PEs! Bottom Left) Clusters identified in the SOM. Bottom Right) This figure shows which class each weight vector is closest to, which complements the information on the left. Original figure is in color. Download paper from http://www.ece.rice.edu/∼erzsebet/papers/esann04-1.pdf*

to $0.5641(\rho_{v_4 v_2})$, so this too is a "forbidden" case for BDH according to the available theory. Clustering this data set using KSOM is depicted in Figure 7 (left). The rare classes are detectible but each is poorly represented, by a single PE only. Also, the lack of strong fences (the lighter the color of the fence between two PEs, the larger the difference between the corresponding weights) separating them from surrounding PEs makes them less discernible. BDH with $\alpha = -0.8$, magnifies the rare classes in the map. Figure 7 (right) shows that class 'R' is now represented by 4 PEs as opposed to 1 in the map formed by KSOM, and class 'Q' by approximately 6 PEs. So once again, the rare classes have been magnified and also made more discernable by BDH with $\alpha < 0$.

The aim of negative magnification is to enhance rare classes to improve their detectability. Naturally the representation of the non-rare classes in such a map is somewhat distorted. The map obtained with $\alpha_{achieved} < 0$ and the one obtained with $\alpha_{achieved} \approx 1$ together provide a complete picture of the clustering of a data set.

4 Conclusions

Magnification control in the Self-Organizing Map can be achieved by using the BDH algorithm [1]. We showed by careful simulations that even though the available theory restricts the use of this algorithm to data with special prop-

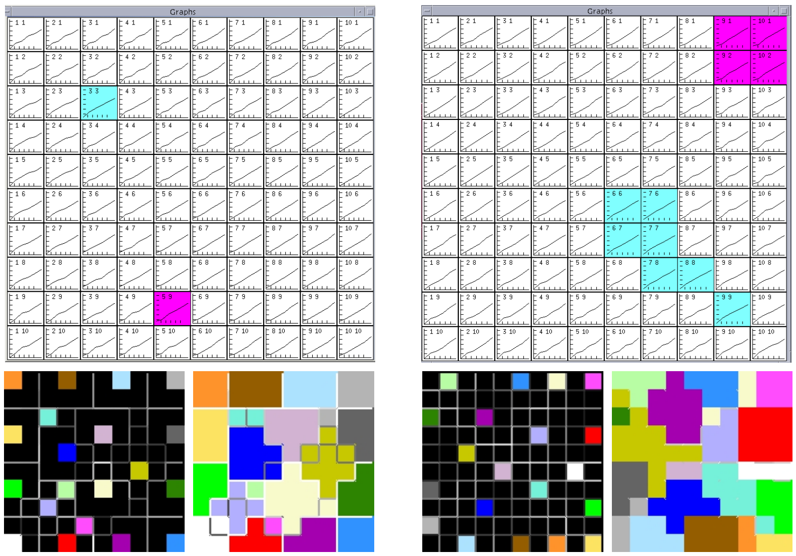

Figure 7: *Results of clustering of the 20-class data set.* ***Left:*** *Using KSOM. Top) Weight vectors in the* 10×10 *map. Only* 1 *PE for each of the rare classes 'R' and 'Q'. Bottom Left) Clusters identified in the SOM (the lighter the color of the fence between two PEs, the larger is the difference between the corresponding weights and vice-versa). Bottom Right) This figure shows which class each weight vector is closest to, which complements the information on the left.* ***Right:*** *Using BDH with* $\alpha = -0.8$*. Top) Weight vectors in the* 10×10 *SOM.* 4 *and* 6 *PEs now represent the rare classes 'R' and 'Q' respectively. Bottom Left) Clusters identified in the SOM. Bottom Right) This figure shows which class each weight vector is closest to, which complements the information on the left. Original figure is in color. Download paper from http://www.ece.rice.edu/∼erzsebet/papers/esann04-1.pdf*

erties, the definitive trend in the results for more general data is encouraging. We have demonstrated in particular that negative magnification enhances the detectibility of rare classes in "forbidden" data. Our preliminary sense is that the applicability of BDH may be justified for a broader range of data than is supported by present theories. The behavior of BDH magnification is worth more investigation, especially for $\alpha = 1$ and $\alpha < 0$ in order to utilize this powerful method for analysis of complex, high-D data.

References

[1] H.-U. Bauer, R. Der, and M. Herrmann. Controlling the magnification factor of self-organizing feature maps. *Neural Computation*, 8:757–771, 1996.

[2] T. Kohonen. The self-organizing map. *Proceedings of the IEEE*, volume 78, pages 1464–1480, 1990.

[3] H. Ritter and K. Schulten. On the stationary state of kohonen's self-organizing sensory mapping. *Biological cybernetics*, 54:99–106, 1986.

[4] D. DeSieno. Adding a conscience to competitive learning. *IEEE International conference on Neural Networks*, volume 1, pages I–117–I–124, 1988.

[5] H. Ritter. Asymptotic level density for a class of vector quantization processes. *IEEE Trans. on Neural Networks*, 2(1):173–175, January 1991.