

# Nonlinear Dimensional Reduction Via Nearest Neighbor Reconstruction

Daniel Yarlett

## 1 Introduction

This report describes a novel model of nonlinear dimensional reduction. The model applies to objects that can be represented as high-dimensional discrete probability distributions, and operates by embedding these high-dimensional objects in a low-dimensional space of selected dimensionality such that each high-dimensional object can be optimally reconstructed based on a weighted sum over the nearest neighbors of the object in the low-dimensional space. The weight function decays exponentially with increasing distance in the low-dimensional space which encourages the algorithm to locate ‘similar’ objects near to one another in the low-dimensional space.

We demonstrate the effectiveness of the model on the MNIST Database of Handwritten Digits (LeCun, Bottou, Bengio and Haffner, 1998). This data set consists of 70,000 digitized images of handwritten numeric digits which have been size-normalized and frame-centered, and where each pixel is represented by a monochromatic intensity level. The model proves effective in locating digits belonging to the same class in tight clusters in a 2-dimensional space despite being presented with unlabeled data, indicating that the model could assist in applications such as the visualization and classification of high-dimensional data.

## 2 The Model

We assume that each object to be modeled can be represented by a discrete probability distribution  $P^{(m)}$ , consisting of elements  $p_i^{(m)}$  denoting the probability of feature  $i$  under the object,  $m$ , where  $\sum_i p_i^{(m)} = 1$ .

The criterion the model optimizes is the error arising from reconstructing the probability distribution associated with a given object based on a weighted sum over the probability

distributions associated with the nearest neighbors of the object in a postulated low dimensional manifold. We measure the reconstruction error on feature  $i$  of object  $m$ ,  $E_i^{(m)}$ , using the Kullback-Leibler metric

$$E_i^{(m)} \equiv p_i^{(m)} \log \frac{p_i^{(m)}}{q_i^{(m)}} = p_i^{(m)} \log p_i^{(m)} - p_i^{(m)} \log q_i^{(m)} \quad (1)$$

where  $q_i^{(m)}$  denotes the reconstructed probability estimate of  $p_i^{(m)}$  under the model. This ‘reconstructed’ probability estimate is generated by a weighted sum over the probabilities assigned to feature  $i$  by the nearest neighbors of object  $m$  in the low dimensional manifold, defined as follows

$$q_i^{(m)} \equiv \alpha p_i^{(m)} + (1 - \alpha) \frac{\sum_{n \in N} w_{nm} p_i^{(n)}}{\sum_{n \in N} w_{nm}} \quad (2)$$

where  $0 < \alpha < 1$ , and  $N$  is the set of all objects that are the nearest neighbors of object  $m$ . The nearest neighbor set for an object is defined by measuring the Euclidean distance of all the objects from the target object, and then taking the  $k$  nearest objects. Note that an object cannot be a nearest neighbor of itself.

The  $\alpha$  parameter employed above is intended to prevent the reconstructed probability from equaling 0 in the event that none of the nearest neighbors of  $m$  assign a non-zero probability to feature  $i$  (which would cause the logarithm in Equation (1) to become degenerate); in most applications we have found that  $\alpha \approx 0.01$  leads to good performance.

$w_{nm}$  denotes the weight assigned to object  $n$  relative to object  $m$  in the sum, and is a decreasing function of the distance between the two objects in the low dimensional manifold as follows

$$w_{nm} \equiv \exp(-d_{nm}^{\frac{1}{2}}) \quad (3)$$

where

$$d_{nm} \equiv (x^{(m)} - x^{(n)})^2 + (y^{(m)} - y^{(n)})^2 + \dots \quad (4)$$

and  $x^{(m)}$  represents the position of object  $m$  on dimension  $x$ ,  $y^{(m)}$  represents the position of object  $m$  on dimension  $y$ , and so on, for each dimension of the low-dimensional space.

### 3 Learning Low-Dimensional Locations

The locations assigned to the objects in the postulated low-dimensional space are learned by setting them such that they minimize the sum of the reconstruction error in Equation (1) for each feature of each object in the data set,  $\sum_{m,i} E_i^{(m)}$ .

In order to facilitate the optimization process we here derive the rate of change of  $E_i^{(m)}$  with respect to its location on an arbitrary dimension,  $x$ , of the low dimensional manifold.

The derived gradient equation can be used in order to minimize the overall reconstruction error through a process of iterative gradient-descent.

Substituting Equation (2) into Equation (1) we get

$$E_i^{(m)} \equiv p_i^{(m)} \log p_i^{(m)} - p_i^{(m)} \log \left( \alpha p_i^{(m)} + (1 - \alpha) \frac{\sum_{n \in N} w_{nm} p_i^{(n)}}{\sum_{n \in N} w_{nm}} \right)$$

And differentiating this equation yields

$$\begin{aligned} \frac{dE_i^{(m)}}{dx^{(m)}} &= (\alpha - 1) p_i^{(m)} \left( \alpha p_i^{(m)} + (1 - \alpha) \frac{\sum_{n \in N} w_{nm} p_i^{(n)}}{\sum_{n \in N} w_{nm}} \right)^{-1} \frac{d}{dx^{(m)}} \left( \frac{\sum_{n \in N} w_{nm} p_i^{(n)}}{\sum_{n \in N} w_{nm}} \right) \\ &= \frac{(\alpha - 1) p_i^{(m)}}{q_i^{(m)}} \frac{d}{dx^{(m)}} \left( \frac{\sum_{n \in N} w_{nm} p_i^{(n)}}{\sum_{n \in N} w_{nm}} \right) \end{aligned} \quad (5)$$

We now differentiate the final term in Equation (5), by replacing the  $w_{nm}$  terms with their definition in Equation (3) and then applying the quotient rule

$$\begin{aligned} \frac{d}{dx_i^{(m)}} \left( \frac{\sum_{n \in N} w_{nm} p_i^{(n)}}{\sum_{n \in N} w_{nm}} \right) &= \frac{d}{dx_i^{(m)}} \left( \frac{\sum_{n \in N} \exp(-d_{nm}^{\frac{1}{2}}) p_i^{(n)}}{\sum_{n \in N} \exp(-d_{nm}^{\frac{1}{2}})} \right) \\ &= \frac{\left( \sum_{n \in N} (x^{(n)} - x^{(m)}) \exp(-d_{nm}^{\frac{1}{2}}) d_{nm}^{-\frac{1}{2}} p_i^{(n)} \right) \left( \sum_{n \in N} \exp(-d_{nm}^{\frac{1}{2}}) \right)}{\left( \sum_{n \in N} \exp(-d_{nm}^{\frac{1}{2}}) \right)^2} \\ &\quad - \frac{\left( \sum_{n \in N} \exp(-d_{nm}^{\frac{1}{2}}) p_i^{(n)} \right) \left( \sum_{n \in N} (x^{(n)} - x^{(m)}) \exp(-d_{nm}^{\frac{1}{2}}) d_{nm}^{-\frac{1}{2}} \right)}{\left( \sum_{n \in N} \exp(-d_{nm}^{\frac{1}{2}}) \right)^2} \end{aligned} \quad (6)$$

By substituting Equation (6) into Equation (5) we get the final expression for the gradient we have been seeking

$$\frac{dE_i^{(m)}}{dx^{(m)}} = \frac{(\alpha - 1) p_i^{(m)}}{q_i^{(m)}} \left( \frac{\left( \sum_n v_{nm} w_{nm} d_{nm}^{-\frac{1}{2}} p_i^{(n)} \right) \left( \sum_n w_{nm} \right) - \left( \sum_n w_{nm} p_i^{(n)} \right) \left( \sum_n v_{nm} w_{nm} d_{nm}^{-\frac{1}{2}} \right)}{\left( \sum_n w_{nm} \right)^2} \right) \quad (7)$$

where  $v_{nm} \equiv x^{(n)} - x^{(m)}$ .