Nearest Linear Manifold Classification

Yuhui Yao, Lihui Chen and Yan Qiu Chen
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore, 639798
Tel:65-7904484 Fax:65-7912687
Email: elhchen@ntu.edu.sg

July 14, 2001

Abstract

A novel classifier, named Nearest Linear Manifold uses a small number of prototypes to represent a class and extend their representational capacity by using the linear manifold of the prototypes to provide more sufficient feature information for classification.

1 Introduction

In the age of the Internet, Intranets and Data Mining, the paradigms of classification are needed for improving. Multiple-Prototype Classifier is one of important techniques using a small number vectors to represent a class and then for further classification. It makes use of some clustering methods [5, 6] to generate prototypes and design the classifier with 1-nearest prototype (NP) rule. In the NP-based classification, since the number of prototypes is small, it is desirable to have a sufficiently large number of feature points stored to account for as many variations as possible. In this letter, we propose a novel classification method namely Nearest Linear Manifold Classification (NLM) to extend the representational capacity of available prototype samples.

2 Nearest Linear Manifold Classification

The NLM assumes that at least two prototypes are available for each class, which is usually satisfied. It attempts to extend the representational capacity of available prototypes by linearly connecting each pair vectors within the same class to form the linear manifolds of that class. In the following, we will discuss a multiple-prototype generation scheme using fuzzy c-means, a new distance from a pattern vector to a manifold and describe Nearest Linear Manifold based classification.

2.1 Multiple-Prototype generation

The selection of the prototypes from a class \( \omega \) is achieved through the minimization of fuzzy c-means objective function

\[
Q(U, V^\omega, X^\omega) = \sum_{j=1}^{k^\omega} \sum_{i=1}^{N^\omega} w_{ij}^m \|x_i^\omega - v_j^\omega\|^2
\]

where \( V^\omega = \{v^\omega_1, \ldots, v^\omega_k\} \) are prototypes of class \( \omega \). \( U \) is a fuzzy \( k^\omega \)-partition matrix of the class \( \omega \) data set \( X^\omega = \{x_i^\omega\} \). \( m \in [1, \infty) \) represents the amount of increased sharing of points among all clusters. \( m = 1 \) corresponds to the crisp case and \( m = \infty \) corresponds to the maximally fuzzy case. It is normally chosen.
$m \approx 2$ which is known to give good results for a physical interpretation of fuzzy c-means [7]. $\|\|$ is a distance measurement and is usually applied with Euclidean Distance. We select $k^\omega$ according to experiment results in the case of the least classification error or to be the integer closest to $\sqrt{N^\omega}$, that is $k^\omega = \text{round}(\sqrt{N^\omega})$, to balance the numbers of the representatives and of the points each representative is accountable to, where $N^\omega$ is the number of pattern vectors in class $\omega$. The fuzzy $c$-means algorithm can be highlighted as follows. The class $\omega$ data are cycled in the computing, during the $t$th cycle the vectors of the output are updated as:

$$v^\omega_{j,t} = \frac{\sum_{i=1}^{N^\omega} (u_{ij})^m x_i^\omega}{\sum_{i=1}^{N^\omega} (u_{ij})^m} \quad \text{for} \quad 1 \leq j \leq k^\omega \quad (2)$$

$$u_{ij,t} = \frac{1}{\sum_{h=1}^{k} \left( \frac{\|x_i^\omega - v_j^\omega\|^2}{\|x_h^\omega - v_j^\omega\|^2} \right)^{(\frac{1}{\mu} - 1)}} \quad \text{for} \quad 1 \leq j \leq k^\omega, \quad (3)$$

The iteration is terminated when

$$E^\omega_t = \|v^\omega_t - v^\omega_{t-1}\|_A = \sum_{j=1}^{k^\omega} \|v^\omega_{j,t} - v^\omega_{j,t-1}\|^2 \leq \varepsilon \quad (4)$$

where $\varepsilon$ is a small positive constant.

### 2.2 The linear manifold distance

In a feature space of any dimension, we linearly connect two prototypes $x_1, x_2$ of the same class to a linear manifold, denoted $\mathbf{x_1 x_2}$. The query pattern vector $x$ is projected onto this linear manifold as point $p$ (Fig. 1).

We can compute the projection point with

$$p = x_1 + \mu(x_2 - x_1) \quad (5)$$

where $\mu \in \mathbb{R}$ is a position parameter which has

$$\mu = \frac{(x - x_1) \cdot (x_2 - x_1)}{(x_2 - x_1) \cdot (x_2 - x_1)} \quad (6)$$

when “.” stands for dot product. The parameter $\mu$ describes the position of $p$ relative to $x_1$ and $x_2$. When $\mu = 0$, $p = x_1$. When $\mu = 1$, $p = x_2$. When $0 < \mu < 1$, $p$ is an interpolating point between $x_1$ and $x_2$. When $\mu > 1$, $p$ is forward extrapolating point on the $x_2$ side. When $\mu < 0$, $p$ is backward extrapolating point on the $x_1$ side.

The distance of linear manifold is defined as

$$d(x, \mathbf{x_1 x_2}) = \begin{cases} \|x - x_1\|, & \mu < 0 \\ \|x - p\|, & 0 \leq \mu \leq 1 \\ |x - x_2|, & \mu > 1 \end{cases} \quad (7)$$

where $\mu$ is calculated as same as Equ. 6 and $\| \cdot \|$ is some norm. A visual explanation of such distance is shown in Fig. 2. If the query vector $x$ falls into the light shadow region of the linear manifold $\mathbf{x_1 x_2}$, the distance is calculated the distance between $x$ and $p$. When the query vector $x$ falls into the dark shadow region of the linear manifold $\mathbf{x_1 x_2}$, with different values of $\mu$ the distance is calculated by the nearest prototype of the pair with some norm.

### 2.3 Nearest-Linear-Manifold based classification

Nearest-Linear-Manifold (NLM) based classification make uses of all prototypes and the straight segment passing through any two prototypes within the same class forms a linear manifold of that class. The NLM
distance is the first rank distance

\[ d_{nlm}(x, v_{i^*}^\omega v_{j^*}^\omega) = \min_{1 \leq \omega \leq c} \min_{1 \leq i < j \leq N_\omega} d_{nlm}(x, v_i^\omega v_j^\omega) \]  

where here \( c \) is the number of classes, \( \omega \) denotes class \( \omega \). The first rank gives the NLM classification composed of the best matched class \( \omega^* \) and the two best matched prototypes \( v_{i^*} \) and \( v_{j^*} \) of the class.

3 Experimental Results

3.1 Iris data

The Iris data (150 four-dimensional patterns; 50 patterns per class) is one of the important benchmarks in pattern recognition. In the experiment, we study the best-case results of NLM with the comparison with other Multiple-Prototype Classifiers \([1, 3, 4]\) with same number of prototypes, which has been illustrated in Table 1. It is clear that NLM provides super results.

3.2 NIPS bacteria classification database

This data set is used for competition\(^1\). This is a small problem based on the classification of bacteria into two categories. In the experiment we use the first 128 instances for training and the remind 100 instances for testing. After 10 times training and testing, we present the average testing accuracy of NLM with comparison 1-nearest neighbor rule shown in Table 2. We can conclude that NLM uses less number of prototypes to present a class and has achieved better accurate than 1-nearest neighbor rule in the data set.

4 Conclusion

In this letter we have presented a novel classifier that uses a small number of prototypes to represent a class and extend their resentational capacity by using linear manifold between the prototypes and provide more sufficient feature information for classification. All experimental results have demonstrated that NLM has great potentials into real application.

---

\( ^1 \) The database can be obtained at http://q.cis.uoguelph.ca/ skremer/NIPS2000/data1/.

---

Table 1: Number of errors comparison: best-case

<table>
<thead>
<tr>
<th>( c )</th>
<th>6</th>
<th>9</th>
<th>15</th>
<th>30</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLM</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>DR</td>
<td>3</td>
<td>3</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>LVQ</td>
<td>14</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>GLVQ</td>
<td>14</td>
<td>4</td>
<td>4</td>
<td>4</td>
</tr>
</tbody>
</table>

Table 2: The average testing accuracy comparison in the NIPS bacteria classification database.

<table>
<thead>
<tr>
<th></th>
<th>accuracy</th>
<th>number of prototypes</th>
</tr>
</thead>
<tbody>
<tr>
<td>NLM</td>
<td>79.4%</td>
<td>11</td>
</tr>
<tr>
<td>NN</td>
<td>71%</td>
<td>128</td>
</tr>
</tbody>
</table>

---
References


