

# Comparison of feature extraction methods with microarray gene-expression data

Takanobu Miyamoto<sup>1</sup>  
t-miy@yamaguchi-u.ac.jp

Yoshihiko Hamamoto<sup>1</sup>  
hamamoto@yamaguchi-u.ac.jp

<sup>1</sup> Yamaguchi University, 2-16-1 Tokiwadai Ube, Yamaguchi 755-8611, Japan

## Abstract

With microarray gene-expression data, we compare supervised feature extraction methods with the unsupervised feature extraction methods. From experimental results, it is shown that the supervised feature extraction methods are more powerful than the unsupervised feature extraction methods in terms of class separability.

**Keywords**: microarray, feature extraction, orthonormal discriminant vector analysis

## 1 Introduction

Visualization of microarray data is extremely important for biological knowledge discovery. The use of unsupervised and supervised feature extraction methods are known for visualization techniques. In this study, we compare the supervised feature extraction method with the unsupervised feature extraction method.

Unsupervised feature extraction methods refer to situations in which available samples are not labeled even if category label for each sample is known. Typical unsupervised feature extraction methods are the Sammon's nonlinear mapping and the principal component analysis (PCA) [1]. In this study, we use PCA as the representative unsupervised feature extraction method. Supervised feature extraction methods refer to situations in which available samples are labeled. In supervised feature extraction methods, a teacher provides a category label for each sample. Typical supervised feature extraction methods are discriminant analysis and the orthonormal discriminant vector analysis (ODV) method [2]. In this study, we use the ODV method as representative supervised feature extraction methods.

## 2 Feature Extraction Methods

### 2.1 Principal Component Analysis (PCA)

The PCA reduce dimensionality by forming linear combinations of the features. The object of the PCA is to find a lower-dimensional representation that accounts for the variance of the features. First, the  $d$ -dimensional covariance matrix is estimated for the full data set. Next, the eigenvectors and eigenvalues are computed, and sorted according to decreasing eigenvalue. The two eigenvectors corresponding to the top two eigenvalues are chosen as features.

### 2.2 Orthonormal Discriminant Vector Analysis (ODV)

The ODV method extracts features  $\mathbf{t}$  which maximize the Fisher's criterion  $J(\mathbf{t})$  subject to the orthonormality of features. Fisher's criterion  $J(\mathbf{t})$  on the feature  $\mathbf{t}$  is estimated by:

$$J(\mathbf{t}) = \frac{\mathbf{t}^T B_d \mathbf{t}}{\mathbf{t}^T W_d \mathbf{t}}. \quad (1)$$

The between-class covariance matrix  $B_d$  and the within-class covariance matrix  $W_d$  are estimated by:

$$B_d = \sum_{i=1}^{m-1} \sum_{j=i+1}^m P_i P_j (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)^T, \quad (2)$$

$$W_d = \sum_{i=1}^m \Sigma_i \quad (3)$$

where  $m$  is the number of classes,  $P_i$  is the a prior probability of class  $w_i$ ,  $\mu_i$  is the mean vector of class  $w_i$ , and  $\Sigma_i$  is the covariance matrix of class  $w_i$ . The Fisher's criterion measures the difference between means normalized by the averaged variance on the feature  $t$ . The first feature  $t_1$  is obtained by computing the eigenvector corresponding to the largest eigenvalue of  $W^{-1}B$ . The first feature of the ODV method and that of discriminant analysis are same. The second feature  $t_2$  is obtained by computing the eigenvector corresponding to the largest eigenvalue of  $W_{d-1}^{-1}B_{d-1}$  on the orthogonal complementary space of the first feature  $t_1$ . Discriminant analysis has a serious shortcoming in that the maximum number of feature to be extracted is  $m-1$ . When  $m=2$ , discriminant analysis always results in the extraction of a single feature. In contrast with discriminant analysis, the ODV method has no limitation on the maximum number of features to be extracted. Thus, in this study, we extract features by using the ODV method.

### 3 Experimental Results

We used the microarray data consisting of 190 tumor and 90 normal tissue samples with 16,063 genes [3]. First, we selected 100 out of 16,063 genes by using one-dimensional Fisher criterion because many genes are redundant. Next, the level of expression of each gene was normalized to have zero mean and unit variance. Last, we mapped 270 samples down to the two-dimensional plane from 100-dimensional feature space with PCA and the ODV method. Results are respectively shown in Figure1 and Figure 2.

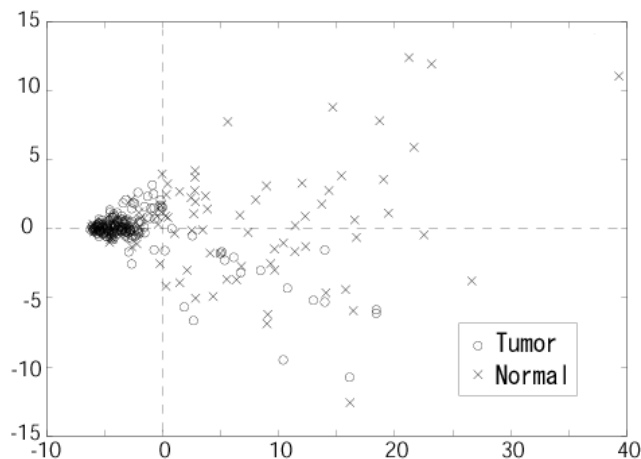


Figure1: Distribution of samples plotted with PCA

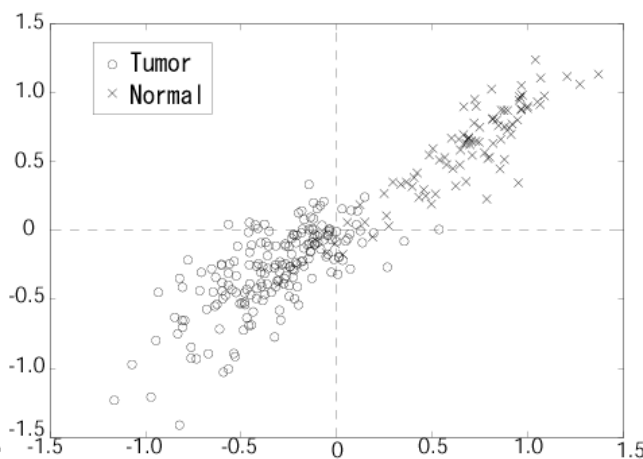


Figure2: Distribution of samples plotted with ODV

### 3 Discussions

We compared the ODV method with the PCA in terms of separability of two classes. In the PCA, distributions of two classes overlapped. In the ODV method, the distributions of two classes are clearly separated. We found that the ODV method is more suitable feature extraction method than the PCA for microarray data with category.

### References

- [1] S. Watanabe, "Karhunen-Loeve expansion and factor analysis: Theoretical remarks and applications", Trans. On 4th Pague Conf. Inf. Theory, pp. 635-660, 1965.
- [2] Y. Hamamoto, T. Kanaoka and S. Tomita, "On a theoretical comparison between the orthonormal discriminant vector method and discriminant analysis", Pattern Recognition, Vol.26, No.12, pp.1863-1867, 1993.
- [3] Sridhar Ramaswamy, Pablo Tamayo, Ryan Rifkin, Sayan Mukherjee, Chen-Hsiang Yeang, Michael Angelo, Christine Ladd, Michael Reich, Eva Latulippe, Jill P. Mesirov, Tomaso Poggio, William Gerald, Massimo Loda, Eric S. Lander,, and Todd R. Golub, "Multiclass cancer diagnosis using tumor gene expression signatures", PNAS, Vol.98, No. 26, pp.15149-15154, 2001.