Lietuvos matematikos rinkinys. LMD darbai
51 tomas, 2010, 227–231

ISSN 0132-2818
www.mii.lt/LMR/

Comparison of linear discriminant functions in image classification

Lijana Stabingienė, Giedrius Stabingis, Kęstutis Dučinskas

Klaipeda University, Department of Statistics
H. Manto 84, LT-92294 Klaipėda

E-mail: lijana.stabingiene@gmail.com; gstabingis@gmail.com; kestutis.ducinskas@ku.lt

Abstract. In statistical image classification it is usually assumed that feature observations given labels are independently distributed. We have retracted this assumption by proposing stationary Gaussian random field (GRF) model for features observations. Conditional distribution of label of observation to be classified is assumed to be dependent on its spatial adjacency with training sample spatial framework.
Perfomance of the Bayes discriminant function (BDF) and performance of plug-in BDF are tested and are compared with ones ignoring spatial correlation among feature observations.For illustration image of figure corrupted by additive GRF is analyzed. Advantage of proposed BDF against competing ones is shown visually and numerically.

Keywords: training sample, Markov Random Fields, spatial correlation.

Introduction

Image classification is a problem of dividing an observed image into several homogeneous regions by labeling pixels based on feature information and information about spatial adjacency relationships with training sample. Switzer (1980) [3] was the first to treat classification of spatial data. It is usually assumed that feature observations conditional on labels are independent (conditional independence) and normally distributed and the labels follow the Markov Random Field (MRF) model. This approach is widely used in image classification [2].

In this paper we propose classification rules by retracting the requirement of conditional independence, i.e., between feature observation to be classified and feature observations in training sample.

The stationary Gaussian Random Fields (GRF) model for features and MRF model for class labels are considered. In the case of partial parametric uncertainty, the plug-in BDF is proposed. This is the generalization of the discriminant function derived in the case of training sample with fixed training sample and fixed prior probabilities for labels [1]. The numerical analysis of proposed discriminant function is performed in the case of isotropic exponential spatial correlation among features observations. For the MRF based on NN(4) neighborhood system, the the performance of proposed discriminant functions are evaluated numerically and visually.

1 The main concepts and definitions

Suppose that the feature is modeled by Gaussian random field $\{Z(s)\colon s \in \mathcal{D} \subset \mathcal{R}^2\}$ In the cotext of image analysis index s means pixel.

The marginal model of observation $Z(s)$ in class Ω_l is

$$Z(s) = \mu_l + \varepsilon(s)$$

where μ_l is constant mean and the error term is generated by zero − mean stationary Gaussian random field $\{\varepsilon(s)\colon s \in \mathcal{D}\}$ with covariance function defined by model for all $s, u \in \mathcal{D}$

$$\operatorname{cov}\bigl\{\varepsilon(s), \varepsilon(u)\bigr\} = \sigma^2 r(s-u)$$

where $r(s-u)$ is the spatial correlation function and σ^2 is variance as a scale parameter.

Denote the marginal Mahalanobis distance by $\Delta_0 = |\mu_1 - \mu_2|/\sigma$

Let $L = \{1, 2\}$ be a label set. A label of pixel $s \in \mathcal{D}$ associated with $Z(s)$ is a random variable $Y(s)$ taking values in L. Let $S_n = \{s_i \in \mathcal{D};\ i = 1, \dots, n\}$ be a set of training pixels. Set $Y = (Y(s_1), \dots, Y(s_n))'$ and $Z = (Z(s_1), \dots, Z(s_n))'$ and call them labels vector and features vector, respectively.

Thus, the vector $T' = (Z', Y')$ constitutes the training sample.

Suppose that the event $\{T = t\}$ is equivalent to the event $\{Z = z\} \cap \{Y = y\}$, where t, z, y are the realizations of the corresponding random vectors. Suppose that the set of training pixels S_n is fixed. Denote by n_l the number of observations in $T = t$ with labels equal l, $l = 1, 2$.

Assume that the model of Z_n for given $Y_n = y_n$ is

$$Z = X_y \mu + E_n \tag{1}$$

where $X_y = 1_{n1} \oplus 1_{n2}$ is a design matrix, $\mu' = (\mu_1, \mu_2)$ and E is the n-vector of random errors that has multivariate Gaussian distribution $N_n(0, \sigma^2 R)$.

Consider the problem of classification (estimation of $Y(s_0)$) of the feature observation $Z_0 = Z(s_0)$, $s_0 \in \mathcal{D}$, $s_0 \notin S_n$ with given training sample $T = t$.

Denote by r_0 the vector of spatial correlations between Z_0 and Z_n and the matrix of spatial correlations among components of Z_n, respectively. Since Z_0 is correlated with training sample, we have to deal with conditional Gaussian distribution of Z_0 given $T = t$ $(Z = z, Y = y)$ with means μ_{lt}^0 and variance σ_{0t}^2 that are defined by

$$\mu_{lt}^0 = E\bigl(Z_0 \bigm| T = t,\ Y(s_0) = l\bigr) = \mu_l + \alpha_0'(Z_0 - X_y\mu), \quad l = 1, 2$$

and variance

$$\sigma_{0t}^2 = V\bigl(Z_0 \bigm| T = t,\ Y(s_0) = l\bigr) = \sigma^2 R_{0n}$$

where

$$\alpha_0' = r_0' R^{-1}, \qquad R_{0n} = 1 - r_0' R^{-1} r_0.$$

The analogous notations will be valid for the realizations of random variables mentioned above. Let N_0 be a neighborhood of s_0 among locations from S_n.

Assumption 1 *The conditional distribution of $Y(s_0)$ given $T = t$ depends only on $Y = y$, i.e.,*

$$\pi_l(y) = P\bigl(Y(s_0) = l \bigm| T = t\bigr), \quad l = 1, 2.$$

Under the assumption that the classes are completely specified the Bayes discriminant function (BDF) [2] minimizing the probability of misclassification is formed by the logarithm of ratio of conditional densities described above.

Then BDF for classification of Z_0 given $T = t$ (with $Y = y$) under the Assumption 1 is

$$W_t(Z_0) = \left(Z_0 - \frac{1}{2}\left(\mu_{1t}^0 + \mu_{2t}^0\right)\right)' \left(\mu_{1t}^0 - \mu_{2t}^0\right)/\sigma_{0t}^2 + \gamma(y) \tag{2}$$

where $\gamma(y) = \ln(\pi_1(y)/\pi_2(y))$.

If Z_0 is assumed to be independent to T, then BDF is

$$W(Z_0) = \left(Z_0 - \frac{1}{2}(\mu_1 + \mu_2)\right)' (\mu_1 - \mu_2)/\sigma^2 + \gamma(y) \tag{3}$$

Denote it by BDFI.

Suppose that means $\{\mu_l\}$ and σ^2 are unknown and need to be estimated from training sample T.

The plug – in BDF (PBDF)is obtained by replacing the parameters in BDF with their estimates based on $T = t$. Then PBDF to the classification problem specified above is

$$\hat{W}_t(Z_0) = \left(Z_0 - \frac{1}{2}\left(\hat{\mu}_{1t}^0 + \hat{\mu}_{2t}^0\right)\right)' \left(\hat{\mu}_{1t}^0 - \hat{\mu}_{2t}^0\right)/\hat{\sigma}_{0t}^2 + \gamma(y) \tag{4}$$

where for $l = 1, 2$, $\hat{\mu}_{lt}^0 = \hat{\mu}_l + \alpha_0'(z_n - X_y\hat{\mu})$, and $\hat{\sigma}_{0t}^2 = \hat{\sigma}^2 R_{0n}$.

Analogously the PBDF for $W_k(Z_0)$ is

$$\hat{W}(Z_0; \hat{\Psi}) = \left(Z_0 - \frac{1}{2}(\hat{\mu}_1 + \hat{\mu}_2)\right)' (\hat{\mu}_1 - \hat{\mu}_2)/\hat{\sigma}^2 + \gamma(y) \tag{5}$$

Denote it by PBDFI.

2 Numerical example and conclusions

The performance of classification rules associated with DF specified in (2)–(5) are compared visually and numerically. The example of classification (restoration) of figure image corrupted by additive stationary GRF with isotropic exponential covariance is considered. Assume that conditional distribution of $Y(s_0)$ given $Y = y$ depends only on labels for locations from neighborhood $N_0 = NN(4)$, i.e.,

$$\pi_1(y) = 1/\bigl(1 + \exp\left(\rho(1 - 2j/4)\right)\bigr), \quad j = 0, 1, \ldots, 4,$$

where ρ is non negative constant called a clustering parameter, and j is the number of locations from N_0 with labels equal 1.

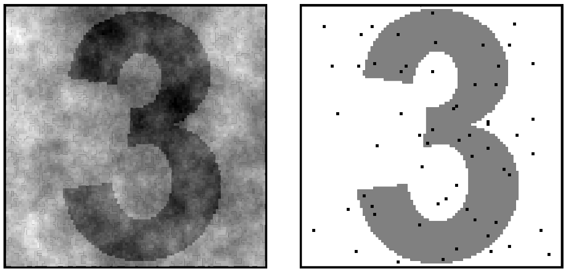
As a natural measure of performance for DF the empirical errors of misclassification $\hat{P}(1 \mid 2), \hat{P}(2 \mid 1)$ are used. They are calculated as frequencies of misclassification of scene pixels and presented in Table 1.

Table 1 shows the advantage of BDF against BDFI and advantage of PBDF against PBDFI in sence of minimum of empirical errors of misclassification.

We take as the true scene on area of 100×100 pixels in which there is a true B/W image of number 3. The objective is to reconstruct this scene on the basis of scene

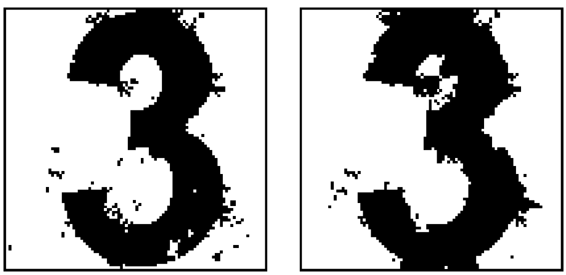
Table 1. Empirical errors of misclassification associated with different DF.

DF	BDF	BDFI	PBDF	PBDFI
$\hat{P}(2 \mid 1)$	0.0479	0.2165	0.0206	0.0236
$\hat{P}(1 \mid 2)$	0.1251	0.4208	0.0388	0.0771



(a) (b)

Fig. 1. (a) Scene corrupted by additive Gaussian Random field, (b) true image and training sample.



(a) (b)

Fig. 2. (a) Reconstructed scene with PBDF, (b) reconstructed scene with PBDFI.

corrupted by the additive stationary GRF with zero mean and isotropic exponential covariance (Fig. 1(a)). Training sample of size $n = 60$ with $n_1 = n_2 = 30$ is taken. The true scene and locations of training pixels are presented in (Fig. 2(b)). For image reconstruction all pixels of corrupted scene except training pixels are classified by using different DF.

Comparing visually the reconstructions based on PBDF (Fig. 2(a)) and PBDFI (Fig. 2(b)) with true scene (Fig. 1(b)) we see the advantage of the former one.

So the results of performed calculations give us the strong argument to encourage the users do not ignore the spatial dependence in image classification and reconstruction.

References

[1] K. Dučinskas. Approximation of the expected error rate in classificattion of the Gaussian random field observations. *Statistics and Probability Letters*, **79**:138–144, 2009.

[2] R. Nishii and S. Eguchi. Image classification based on Markov random field model with Jeffreys divergence. *Journal of Multivariate Analysis*, **97**(9):1997–2008, 2006.

[3] P. Switzer. Extensions of linear discriminant analysis for statistical classification of remotely sensed satellite imagery. *Math. Geol.*, **12**(4):367–376, 1980.

REZIUMĖ

Tiesinių diskriminantinių funkcijų taikymas vaizdų atpažinime

L. Stabingienė, G. Stabingis, K. Dučinskas

Straipsnyje nagrinėjamos tiesinės diskriminantinės funkcijos, kurių pagalba klasifikuojami Gauso atsitiktinio lauko stebiniai. Juodai balto vaizdo rekonstravimo pavyzdyje parodomas diskriminantinių funkcijų (DF), atsižvelgiančių į klasifikuojamo stebinio erdvinę priklausomę nuo mokymo imties, pranašumas prieš DF ignuoruojančias šią priklausomybę.

Raktiniai žodžiai: mokymo imtis, Markovo atsitiktiniai laukai, erdvinė koreliacija.