PARAMETER ESTIMATION OF LINEAR CLASSIFIERS UNDER HIGH-DIMENSIONAL ASYMPTOTICS

Mengyi Zhang

I. Robustness to the Gaussian distribution

In practice, there are situations where the tails of the data may be heavier than those of the normal distribution and usually Student's t distribution can be used to characterize the distribution with heavy tails [1]. The synthetic data are generated according to (A1)-(A3), except that the entries of $\mathbf{Z}_1$ and $\mathbf{Z}_2$ follow t distribution with mean 0, variance 1, and degree of freedom ν. (When $\nu \to \infty$, t distribution reduces to the case of Gaussian distribution.) The other settings are the same as Section 4.1 of the paper.

In the first set of experiments, we fix $\nu = 5$, and the other settings are the same as that for Fig.1 in the paper. Let $p = 60$, $n_1 = n_2 = 40$, $\Delta^2 = 6$, $\rho \in [0,\ 3]$, the performance of true error, proposed error estimator, and conventional plug-in error estimator is plotted in Figure 1a. It can be seen that the conventional plug-in error estimator with three types of $\mathbf{R}_0$ underestimate the error seriously for all values of ρ. With each value of ρ, the proposed error estimator is close to the corresponding true error for all the three types of $\mathbf{R}_0$. Let $p = 60$, $\Delta^2 = 6$, $\rho = 1$, $n_1 = n_2$, $n \in [50,\ 120]$, the performance of true error, proposed error estimator, and conventional plug-in error estimator is plotted in Figure 1b. It can be seen that the conventional plug-in error estimator with three types of $\mathbf{R}_0$ underestimate the error seriously for all values of n. For each type of $\mathbf{R}_0$, the proposed error estimator is close to the corresponding true error, and the gaps become even smaller as n increases.

In the second set of experiments, we first let ν vary in $[5,\ 30]$ and then fix $\nu = 5$. The different discriminant approaches with different covariance matrix estimators and optimal shrinkage parameter are then compared. It can be seen in Figure 1c that the misclassification error of all the classifiers are not sensitive to the change of ν. In Figure 1d with $\nu = 5$, the proposed GLDA classifier outperforms the other classifiers and its performance is close to that the CLV classifier. Therefore, the proposed GLDA classifier with covariance matrix estimator is robust to the Gaussian distribution.

II. Experiments on Real Data

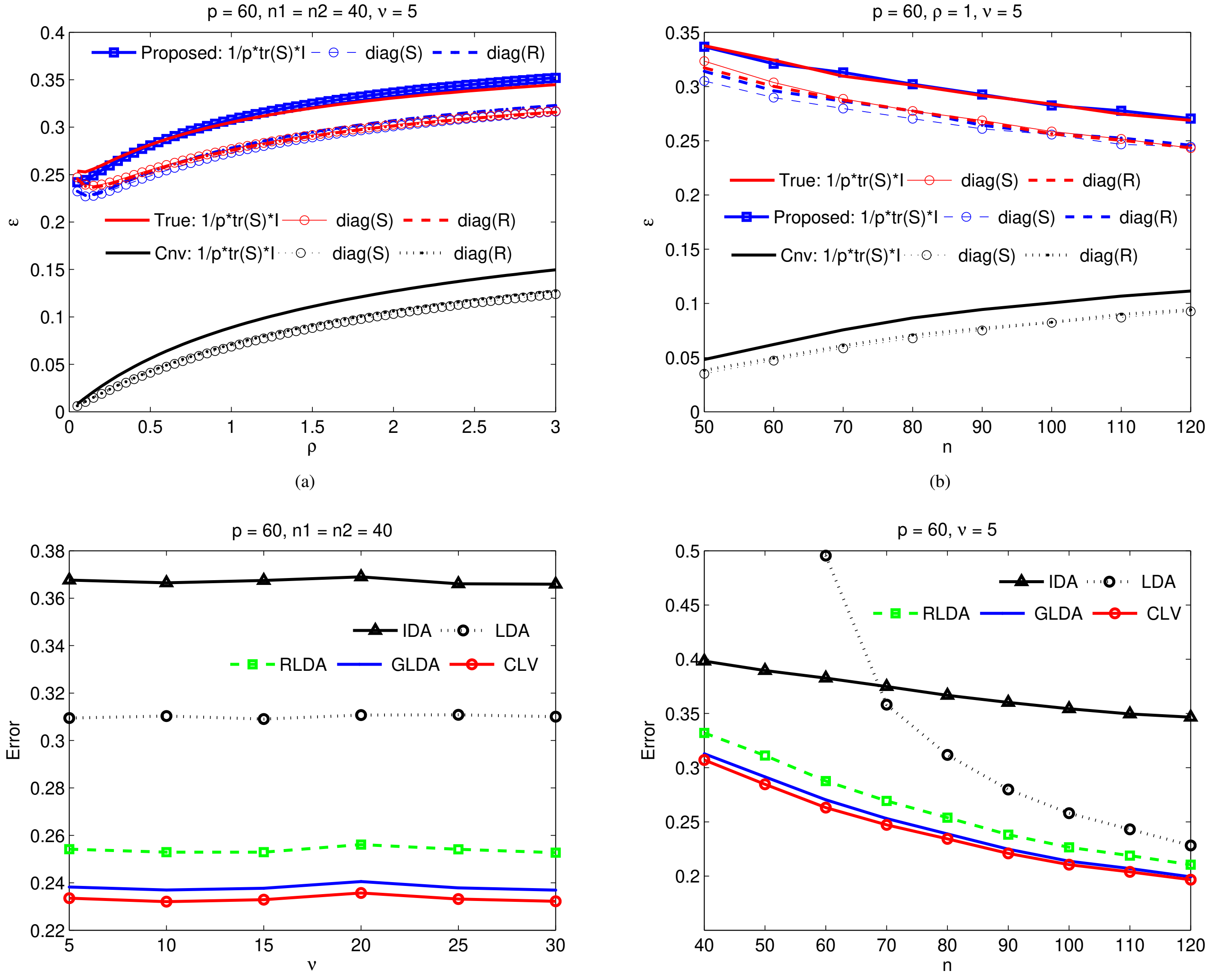
The real-data experiments employ the data collected in [2], which are available in the UCI public dataset. The data are sonar returns from a metal cylinder and a rock positioned on a sandy ocean floor. A set of 97 rock return patterns and 111 cylinder return patterns have been collected, and each pattern has 60 features representing the energy within a particular frequency band. The goal is to classify sonar return patterns for these two targets.

The dataset is divided into a training set and a testing set. Let r be the ratio of the total number of cylinder returns to the total number of rock returns, then $r = 111/97 = 1.14$. In the training set, the training samples are randomly taken out of the whole dataset and let n be the number of training samples. Moreover, the number of training samples in the class of rock returns is $n_1 = \lfloor \frac{n}{1+r} \rfloor$ where $\lfloor \cdot \rfloor$ is the floor function, and the number of training samples in the class of cylinder returns is $n_2 = n - n_1$. The rest of the dataset are considered to be the testing set. Since the sample covariance matrix is not invertible for $n < p$ ($p = 60$), and the conventional LDA classifier cannot be obtained in this case. Therefore, the number of training samples are considered in two intervals, i.e., $n \in [30,\ 60]$ and $n \in [70,\ 100]$, such that the LDA classifier exist in the later case. The proposed GLDA classifier exploits

$$\hat{\mathbf{R}}_{GLDA} = \mathbf{S}_{pooled} + \rho \mathsf{diag}\left(\mathbf{S}_{pooled}\right) \tag{1}$$

. Despite LDA and IDA, its performance of testing error is compared with RLDA and the clairvoyant classifier CLV.

The experiments are repeated 500 times with different random sampling of training data, and the average of the testing errors are evaluated. The results are plotted in Fig.2. It can be seen in Fig.2a that when $n \in [30,\ 60]$, the testing error of RLDA, GLDA, and CLV decrease as the number of training samples increases, because estimation error of the sample covariance matrix decreases. The IDA classifier only estimates the diagonal elements of the covariance matrix, hence the estimation error does not decrease much as the number of training samples increases. The proposed GLDA classifier outperforms the IDA, RLDA classifiers, and its performance becomes more close to the lower bound of CLV classifier as n increases. In Fig. 2b with $n \in [70,\ 100]$, the testing error of the conventional LDA classifier decrease as the number of training samples increases, and the testing error of other classifiers do not decrease much. The performance of IDA, RLDA, and GLDA are more stable when $n > p$. The proposed GLDA classifier outperforms the LDA, IDA, RLDA classifiers, and difference of testing error between GLDA and the CLV lower bound is $1\% - 2\%$. Therefore, the proposed GLDA classifier is more effective than the IDA, RLDA, and RLDA classifiers and its error performance is close to its lower bound.



(a)

(b)

(c)

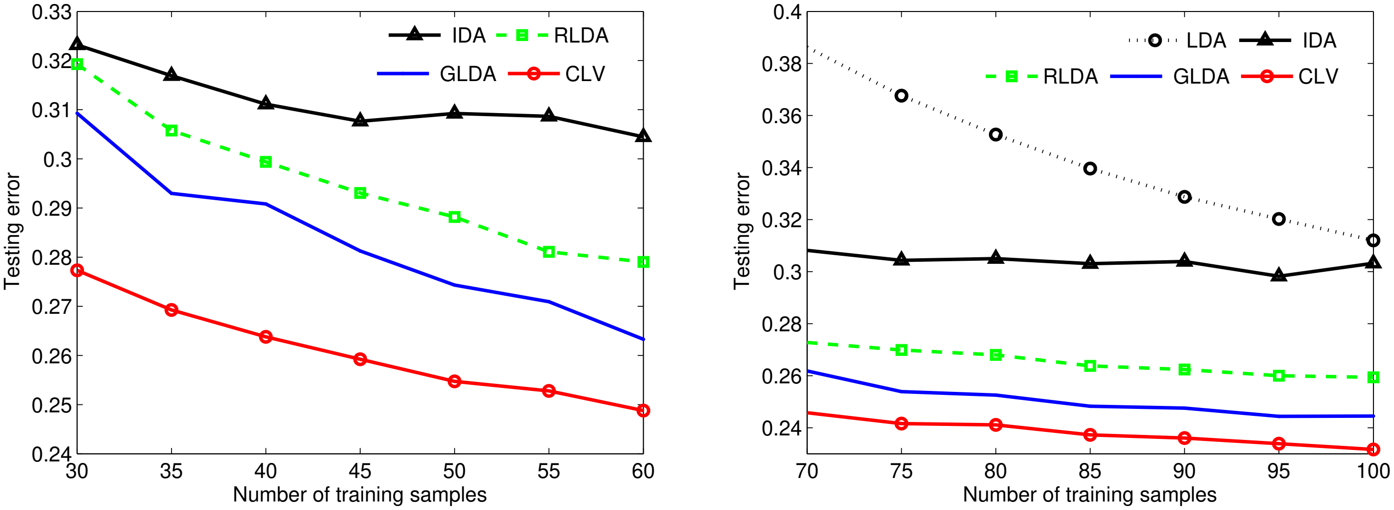
(d)

Figure 1. The error with non-Gaussian distributions versus: (a) $\rho \in [0,\ 3]$; (b) $n \in [50,\ 120]$; (c) optimal ρ, $\nu \in [5,\ 30]$; (d) optimal ρ, $\nu = 5$, $p \in [40,\ 120]$

References

[1] C. W. Dunnett and M. Sobel, "A bivariate generalization of student's t-distribution, with tables for certain special cases," *Biometrika*, vol. 41, no. 1-2, pp. 153–169, 1954.

[2] R. P. Gorman and T. J. Sejnowski, "Analysis of hidden units in a layered network trained to classify sonar targets," *Neural Networks*, vol. 1, no. 1, pp. 75–89, 1988.



(a)

(b)

Figure 2. Testing error versus the number of training samples for the sonar data, $p = 60$, (a) $n \in [30,\ 60]$ (b) $n \in [70,\ 100]$