**Group learning for high-dimensional class-imbalanced data**

1. *Group learning for digit matrix data*

**Experiment 1:**

- *positive class:* 800-digit matrix composed of a certain percentage (0.25, 0.10, 0.025, and 0.01) of digits ‘1’ and the others are even digits (‘0, 2, 4, 6, 8’) (see Fig. 1);

- *negative class:* 800-digit matrix composed of 800 even digits (see Fig. 2);

- number of training inputs/matrices: 5 examples from positive class and 40 from negative class;

- number of validation matrices: 80 (40 per class);

- number of test matrices: 1000 (500 per class)

**Results:**

**Table 1.** Prediction performance for digit matrix data (Experiment 1)

|  |  |  |
| --- | --- | --- |
| Percentage of digits ‘1’ in positive samples | Sensitivity (SS) | Specificity (SP) |
| 0.25 | 1.00 | 1.00 |
| 0.10 | 1.00 | 0.98 |
| 0.025 | 0.86 | 0.94 |
| 0.01 | 0.62 | 0.78 |



**Fig. 1.** Example of the positive class matrix with 800 images which include 720 even digits (‘0, 2, 4, 6, 8’) and 80 digits ‘1’.



**Fig. 2.** Example of the negative class matrix with 800 images which are all even digits (‘0, 2, 4, 6, 8’).

**Experiment 2:**

In Experiment 2, all digits in the digit matrix are removed the left-half portion (see Figs. 3 and 4). Other experimental settings are similar to Experiment 1 (see details below).

- *positive class:* 800-digit matrix composed of 25% digits ‘1’ and the others are even digits (‘0, 2, 4, 6, 8’) (see Fig. 3);

- *negative class:* 800-digit matrix composed of 800 even digits (see Fig. 4);

- number of training inputs/matrices: 5 examples from positive class and 40 from negative class;

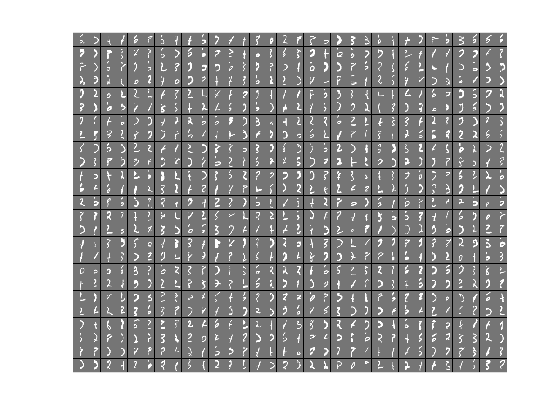
- number of validation matrices: 80 (40 per class);

- number of test matrices: 1000 (500 per class)

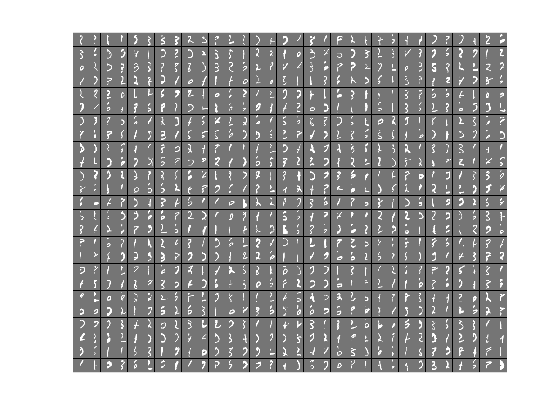
**Results:**

**Table 2.** Prediction performance for half-images digit matrix data (Experiment 2)

|  |  |
| --- | --- |
| Sensitivity (SS) | Specificity (SP) |
| 1.00 | 0.95 |



**Fig. 3.** Example of the positive class matrix with 800 half images which include 720 even digits (‘0, 2, 4, 6, 8’) and 80 digits ‘1’.



**Fig. 4.** Example of the negative class matrix with 800 half images which are all even digits (‘0, 2, 4, 6, 8’).

1. *Statistical Invariance of Group Learning*

**Experiment 3:**

In Experiment 3, we investigate the statistical invariance of Group Learning using different descriptive statistics for making decision rule. Proposed Group Learning only uses mean value of SVM outputs of training data to perform decision rule. In Experiment 3, several descriptive statistics (including mean, std, median, 75th percentile, and 25th percentile) are used for making decision rule as shown in Fig. 5. The experimental settings are shown as below:

- *positive class:* 800-digit matrix composed of 25% digits ‘1’ and the others are even digits (‘0, 2, 4, 6, 8’) (see Fig. 1);

- *negative class:* 800-digit matrix composed of 800 even digits (see Fig. 2);

- number of training inputs/matrices: 5 examples from positive class and 40 from negative class;

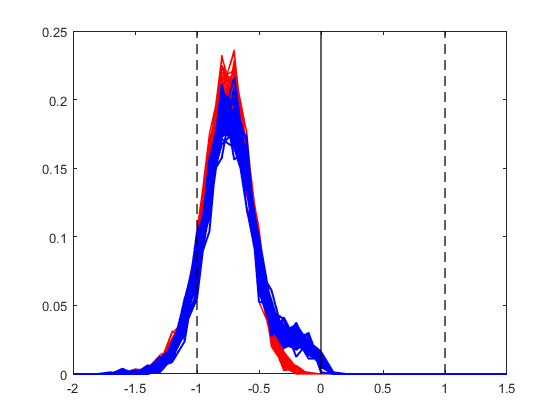
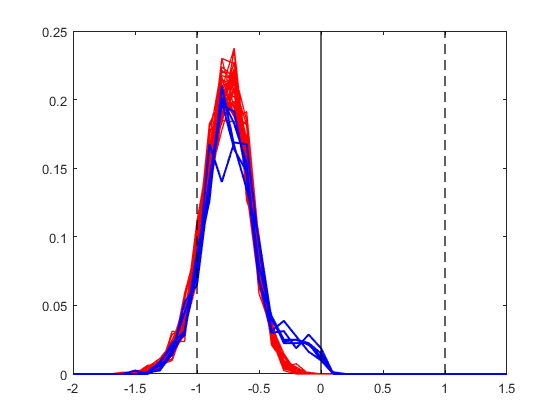
- number of validation matrices: 80 (40 per class);

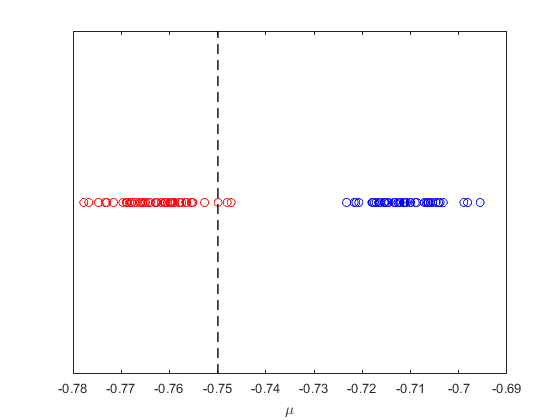
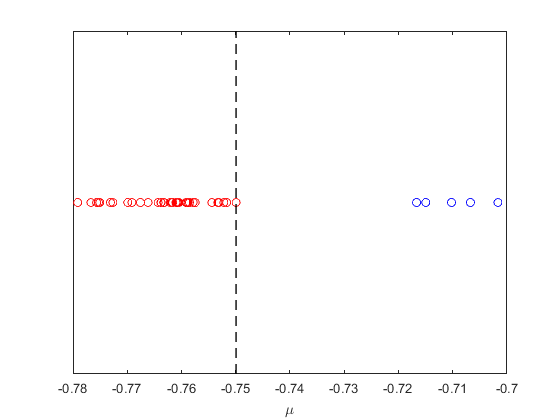
- number of test matrices: 1000 (500 per class)

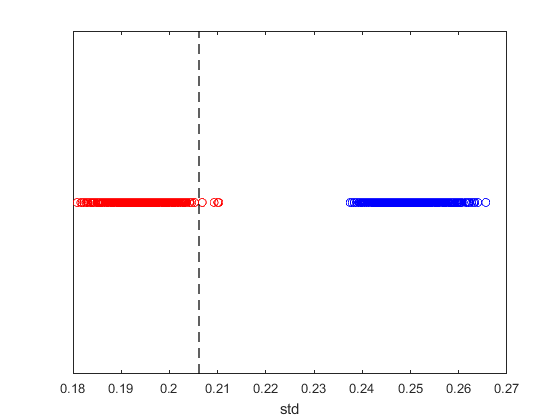
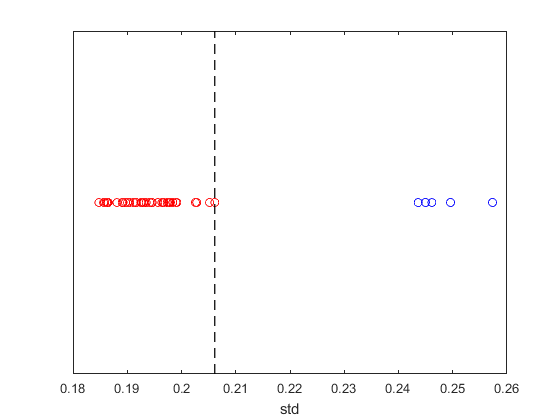
Results:

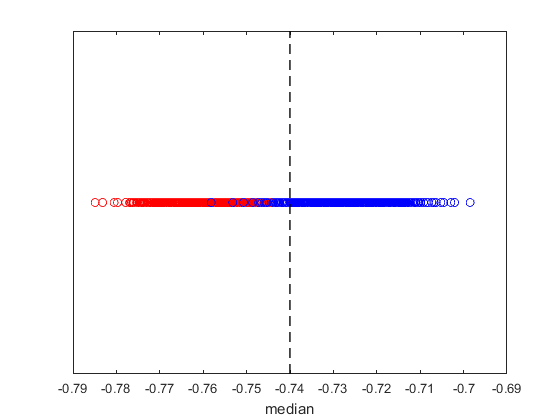
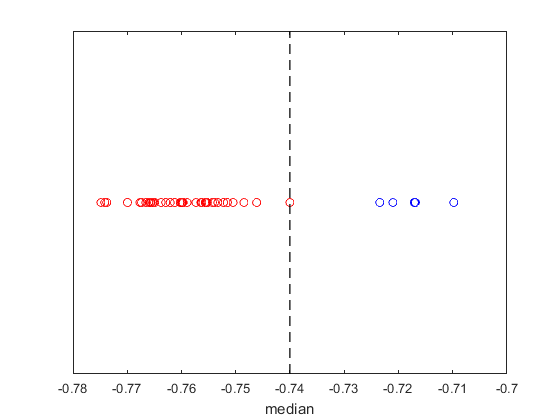
**Table 3.** Prediction performance for Group Learning using different descriptive statistics (Experiment 3)

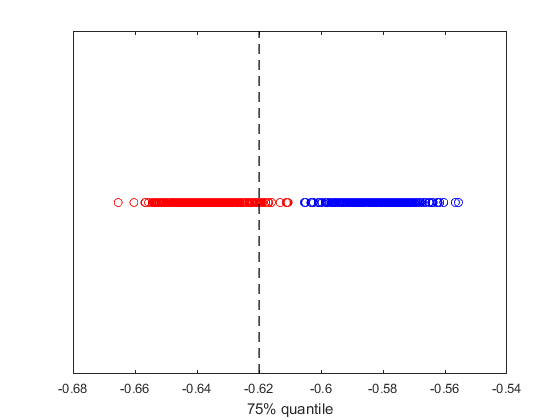
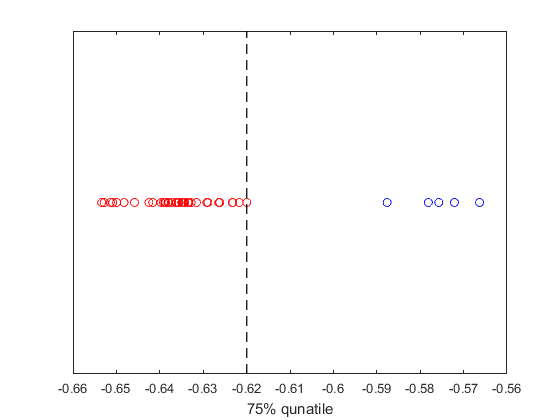
|  |  |  |
| --- | --- | --- |
| Descriptive statistic | SS | SP |
| Mean | 1.00 | 0.98 |
| STD | 1.00 | 0.99 |
| Median | 0.95 | 0.99 |
| 75th percentile | 1.00 | 0.97 |
| 25th percentile | 0.69 | 0.98 |











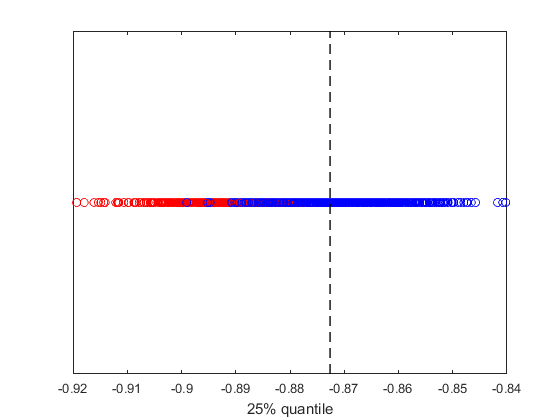
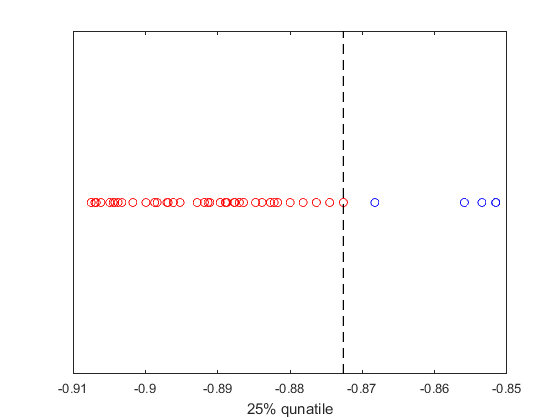


Fig. 5. Histograms of projections of training (top-left corner) and test results (top-right corner). The corresponding decision spaces and thresholds estimated from descriptive statistics are shown in the lower subfigures (from top to bottom: mean, std, median, 75th percentile, and 25th percentile).

1. *Group learning for real-life data*

* Hand written digit matrix: (detailed information in Cherkassky et al., 2019)

*- Positive class:* matrix includes 720 images of even digits (‘0, 2, 4, 6, 8’) and 80 images of digit ‘1’

*- Negative class:* all images in the matrix are even digits

- Ratio between positive and negative samples: 1:8

* iEEG data: (detailed information in Cherkassky et al., 2019)

*- Positive class:* 4-hr preictal iEEG segments

*- Negative class:* 4-hr interictal iEEG segments

- Ratio between positive and negative samples: 1:8

* Gene expression data: (detailed information in Ramaswamy et al., 2003, Díaz-Uriarte et al., 2006)

*- Positive class:* metastatic adenocarcinoma

*- Negative class:* primary adenocarcinomas

- Ratio between positive and negative samples: 12:64

Results:

Table 4. Prediction performance of Group Learning applied in real data sets

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data set | Prior knowledge | SVM | | Group Learning | |
| SS | SP | SS | SP |
| Hand-written digit matrix | Strong | 0 | 1.00 | 1.00 | ~1.00 |
| iEEG | Median | Nan\* | Nan\* | 0.89 | 0.97 |
| Gene expression | Weak | 0.25 | 1.00 | 0.43 | 0.95 |

\*The feature dimension of iEEG samples (i.e., 4hr segment) is too high for SVM

Reference:

1. Cherkassky, Vladimir, Hsiang-Han Chen, and Han-Tai Shiao. "Group Learning for High-Dimensional Sparse Data." *2019 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2019.
2. Ramaswamy, Sridhar, et al. "A molecular signature of metastasis in primary solid tumors." *Nature genetics* 33.1 (2003): 49-54.
3. Díaz-Uriarte, Ramón, and Sara Alvarez De Andres. "Gene selection and classification of microarray data using random forest." *BMC bioinformatics* 7.1 (2006): 3.