**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

Table 1. Prediction performance of Group Learning applied in real data sets

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Data set | Prior knowledge | SVM | | Group Learning | |
| FP | FN | FP | FN |
| Hand-written digit matrix | Strong | 100 | 0 | 0 | 0.2 |
| iEEG | Median | Nan\* | Nan\* | 11.2 | 12.7 |
| Gene expression | Weak | 75 | 0 | 56.7 | 5 |

\*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.