Semiparametric Sparse Discriminant Analysis

Qing Mai, Florida State University

Session #42, Monday, 1:45PM

Linear discriminant analysis:

- Data: $(\mathbf{X}^i, Y^i)$, $i = 1, \ldots, n$, where $\mathbf{X}^i = (X_1^i, \ldots, X_p^i)$ and $Y^i \in \{+1, -1\}$.
- Model:

$$\mathbf{X} \mid Y = y \sim N(\boldsymbol{\mu}_y, \boldsymbol{\Sigma}).$$

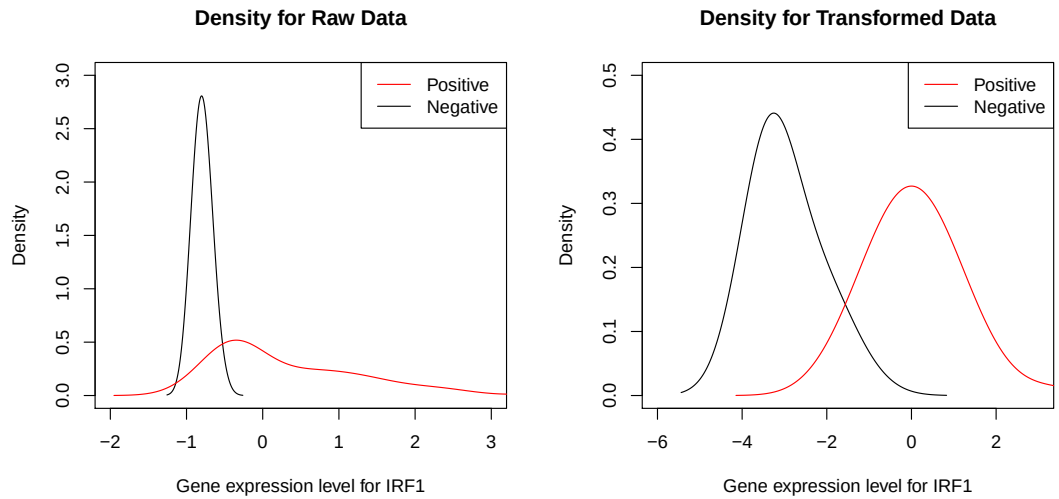
Challenges:

- Cannot be applied to high-dimensional data;
- Is based on rigid assumptions.

A Semiparametric Method

Model:

$$\mathbf{h}(\mathbf{X}) \mid Y = y \sim N(\mu_y, \mathbf{\Sigma})$$



1. Estimation of h;
2. Estimation of a sparse classification rule;
3. Theoretical studies: consistent for ultra-high-dimensional data;
4. Numerical studies.