Preliminaries

Used Datasets - Synthetic

Synthetic Dataset

A

Synthetic Dataset B

- Nr. dimensions: 11

- Nr. dimensions: 100

- Binary Classification Problem
- $x_i^j \sim N(0,1)$
- $p(y_i = 1 \mid x_i) = \dfrac{1}{1 + \text{logit}(x_i)}$

✤ Fixed ture explanations:

- Syn 1: $\exp\ (x_i^1 x_i^2)$
- Syn 2: $\exp\ (\sum_{j=3}^{6} (x_i^j)^2 - 4)$
- Syn3: $\exp\{-10 * sin(2x_i^7) + 2\,|x_i^8| + x_i^9 + \exp(-x_i^{10})\}$

✤ Variable ture explanations:

- Syn 4: If $x_i^{11} < 0$, logit follows Syn 1;
  Otherwise, logit follows Syn 2.
- Syn 5: If $x_i^{11} < 0$, logit follows Syn 1;
  Otherwise, logit follows Syn 3.
- Syn 6: If $x_i^{11} < 0$, logit follows Syn 2;
  Otherwise, logit follows Syn 3.

Preliminaries

Used Datasets - Synthetic

- Binary Classification Problem
- $x_i^j \sim N(0,1)$
- $p(y_i = 1 \mid x_i) = \dfrac{1}{1 + \text{logit}(x_i)}$

Synthetic Dataset A

- Nr. dimensions: 11

Synthetic Dataset B

- Nr. dimensions: 100

✤ Fixed ture explanations:

- Syn 1: exp $(x_i^1 x_i^2)$
- Syn 2: exp $(\sum_{j=3}^{6} (x_i^j)^2 - 4)$
- Syn3: $\exp\{-10 * sin(2x_i^7) + 2\,|x_i^8| + x_i^9 + \exp(-x_i^{10})\}$

✤ Variable ture explanations:

- Syn 4: If $x_i^{11} < 0$, logit follows Syn 1; Otherwise, logit follows Syn 2.
- Syn 5: If $x_i^{11} < 0$, logit follows Syn 1; Otherwise, logit follows Syn 3.
- Syn 6: If $x_i^{11} < 0$, logit follows Syn 2; Otherwise, logit follows Syn 3.

Preliminaries

Used Datasets - Text

- Internet Movie Database (IMDB) dataset
  - A large collection of 50,000 movie reviews
  - Binary labeled: positive (rating ≥ 7); negative (rating ≤ 4)
  - Commonly employed as a benchmark for evaluating IFS methods