

Sparse Common and Distinctive Covariates Logistic Regression

Soogeun Park (s.park_1@tilburguniversity.edu),

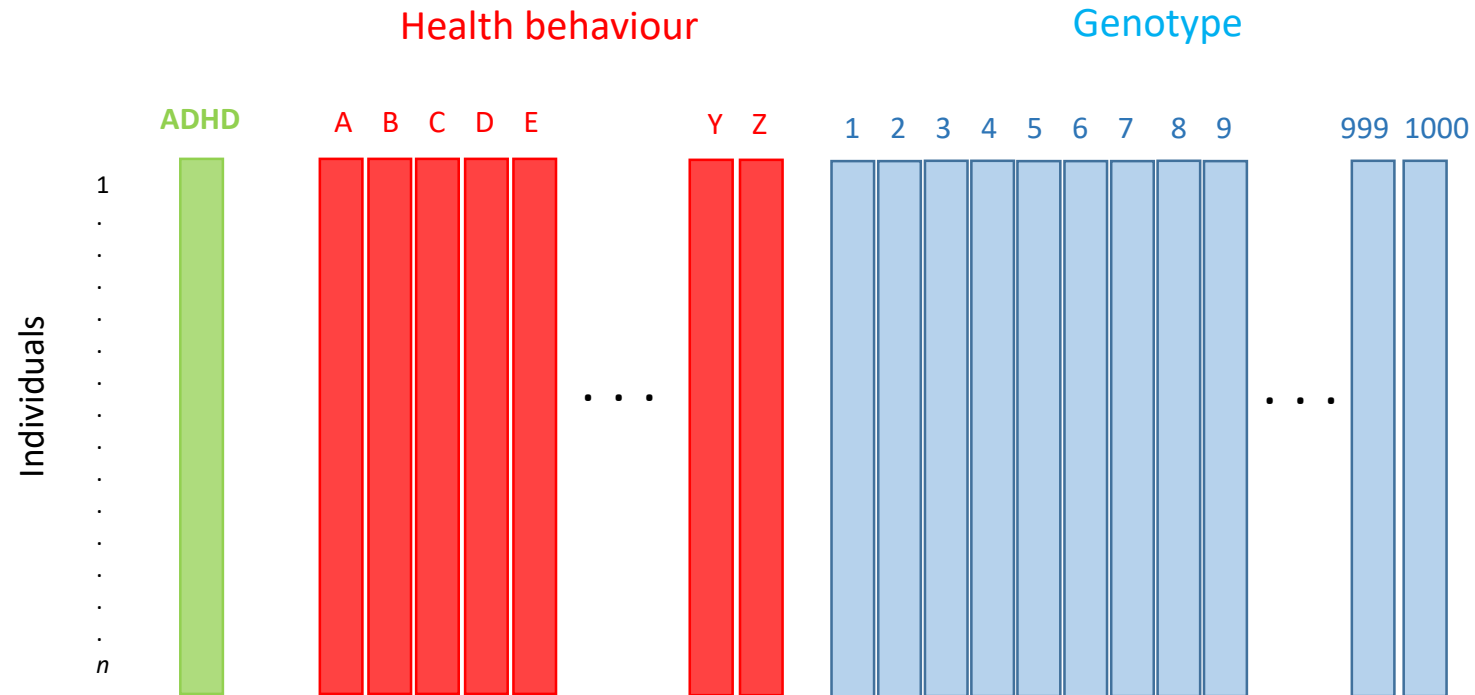
Katrijn Van Deun & Eva Ceulemans

Tilburg University

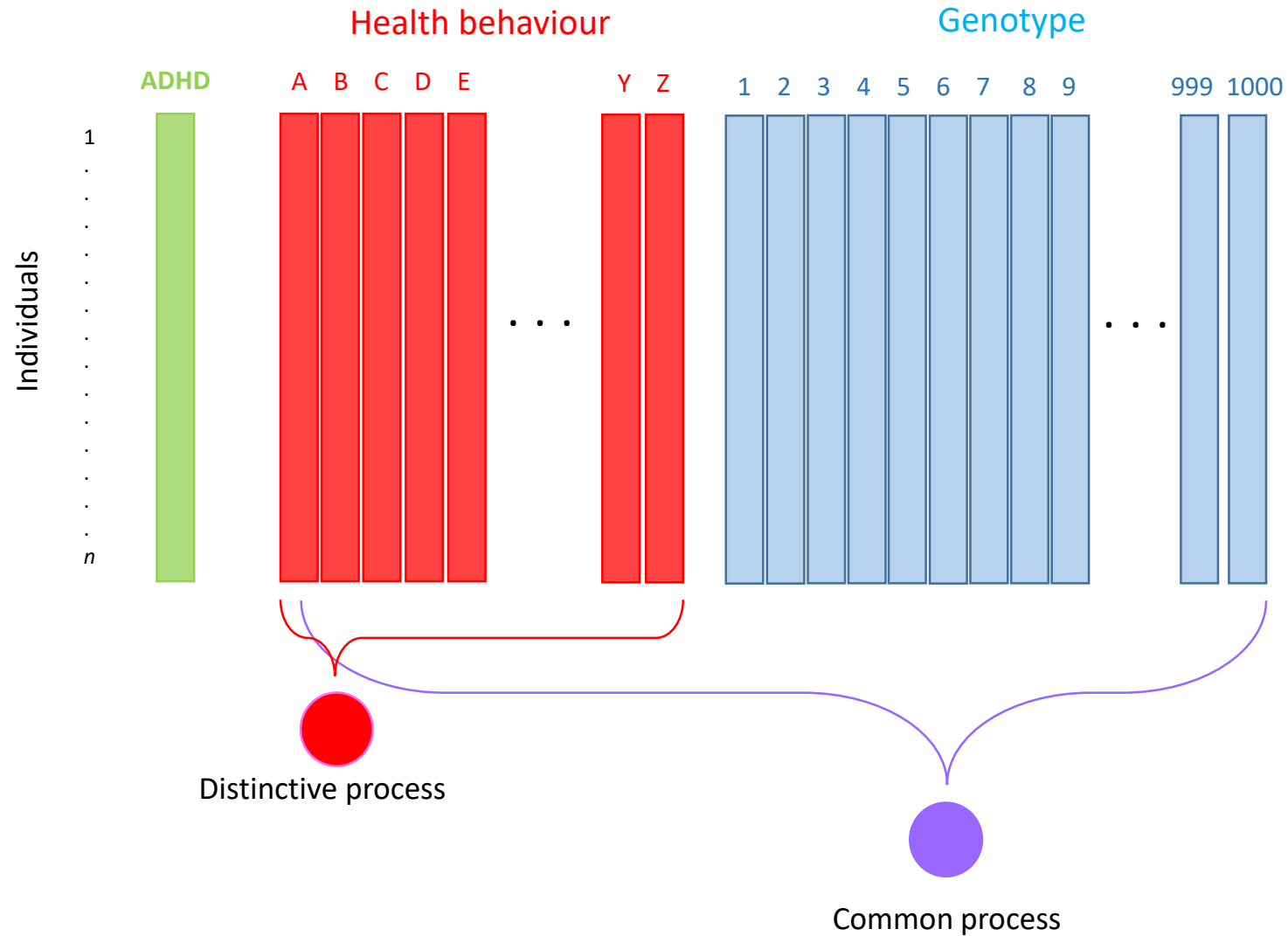
Classification in psychology

- Concerns any construct that is measured by categories
- Eg. In mental health:
 - Depression
 - ADHD
- Classification model investigates the processes in play behind the categories

Multiblock data



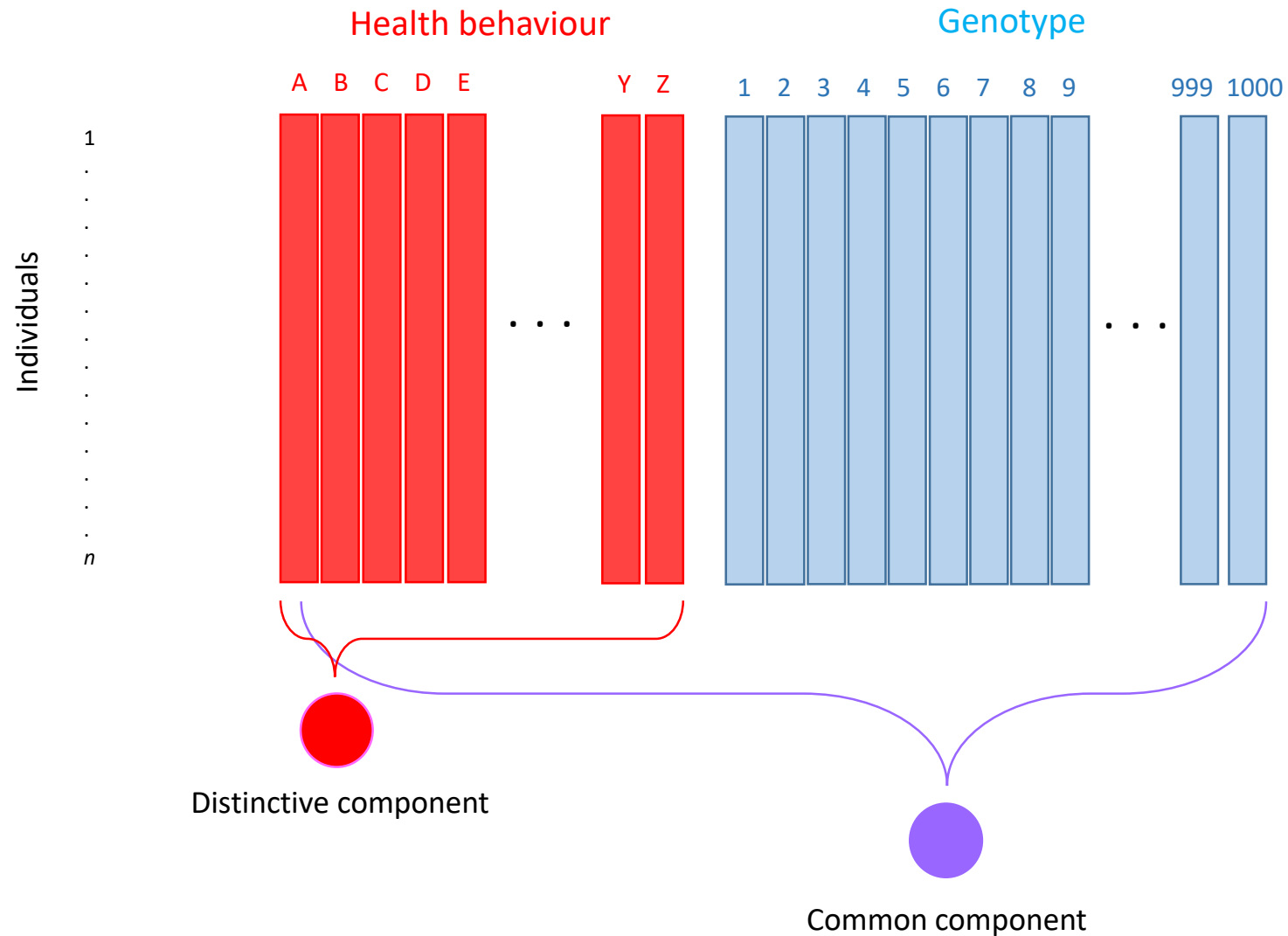
Common and Distinctive processes



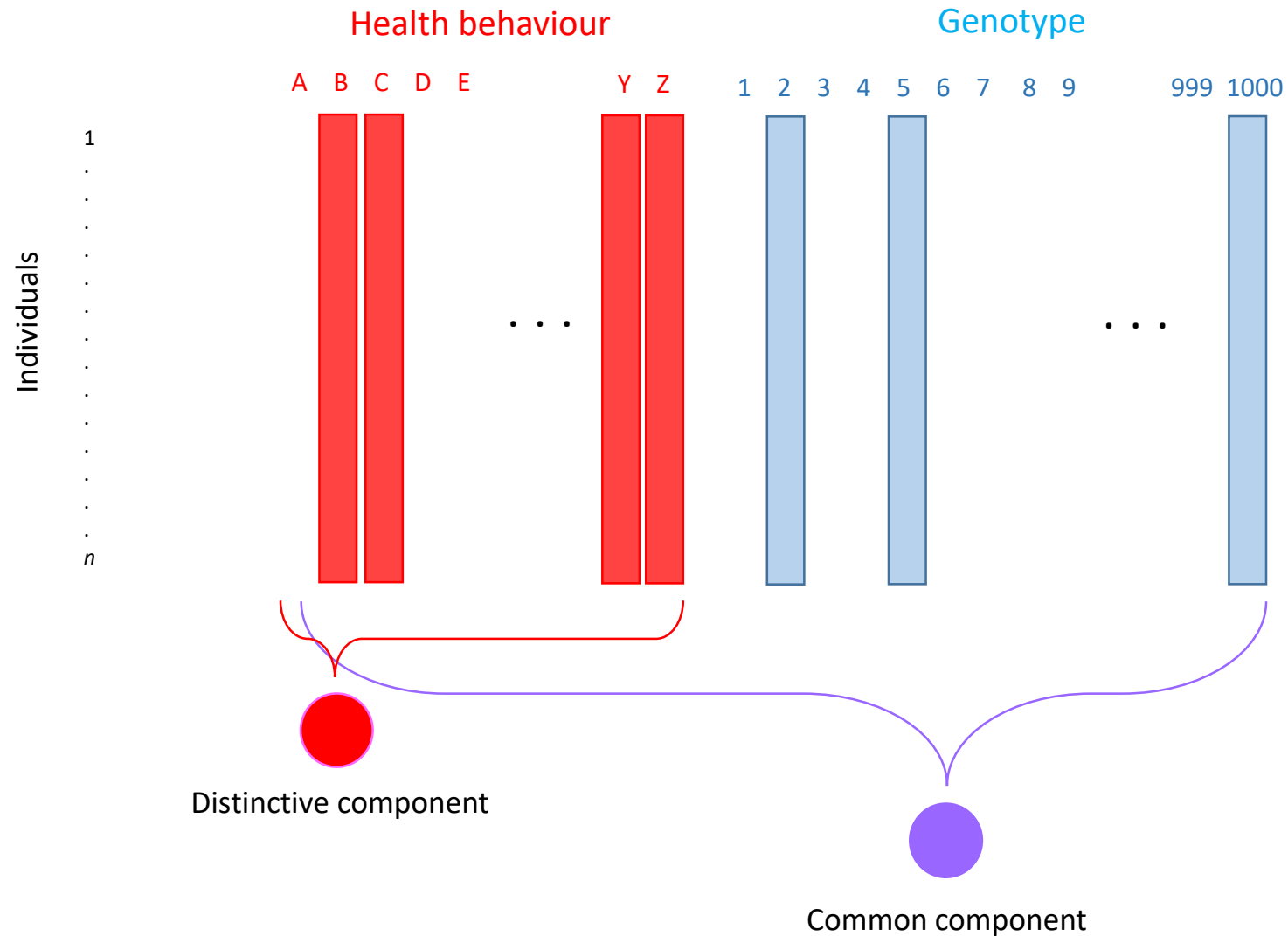
Aims of SCD-Cov-logR

- Classification model for multiblock data
- Identification of common and distinctive predictor processes
- Easier interpretation of the processes via sparse coefficients

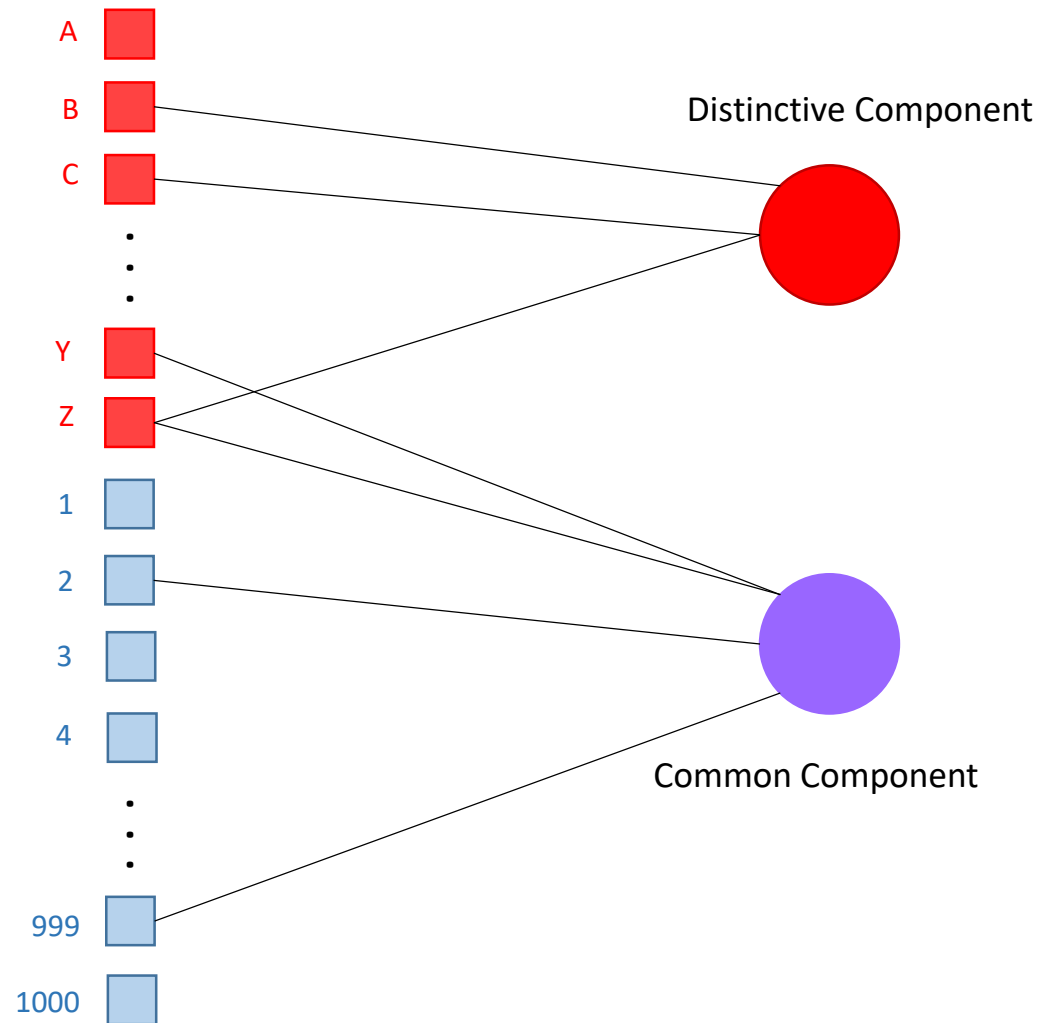
Simultaneous Component Analysis (SCA)



Simultaneous Component Analysis (SCA)

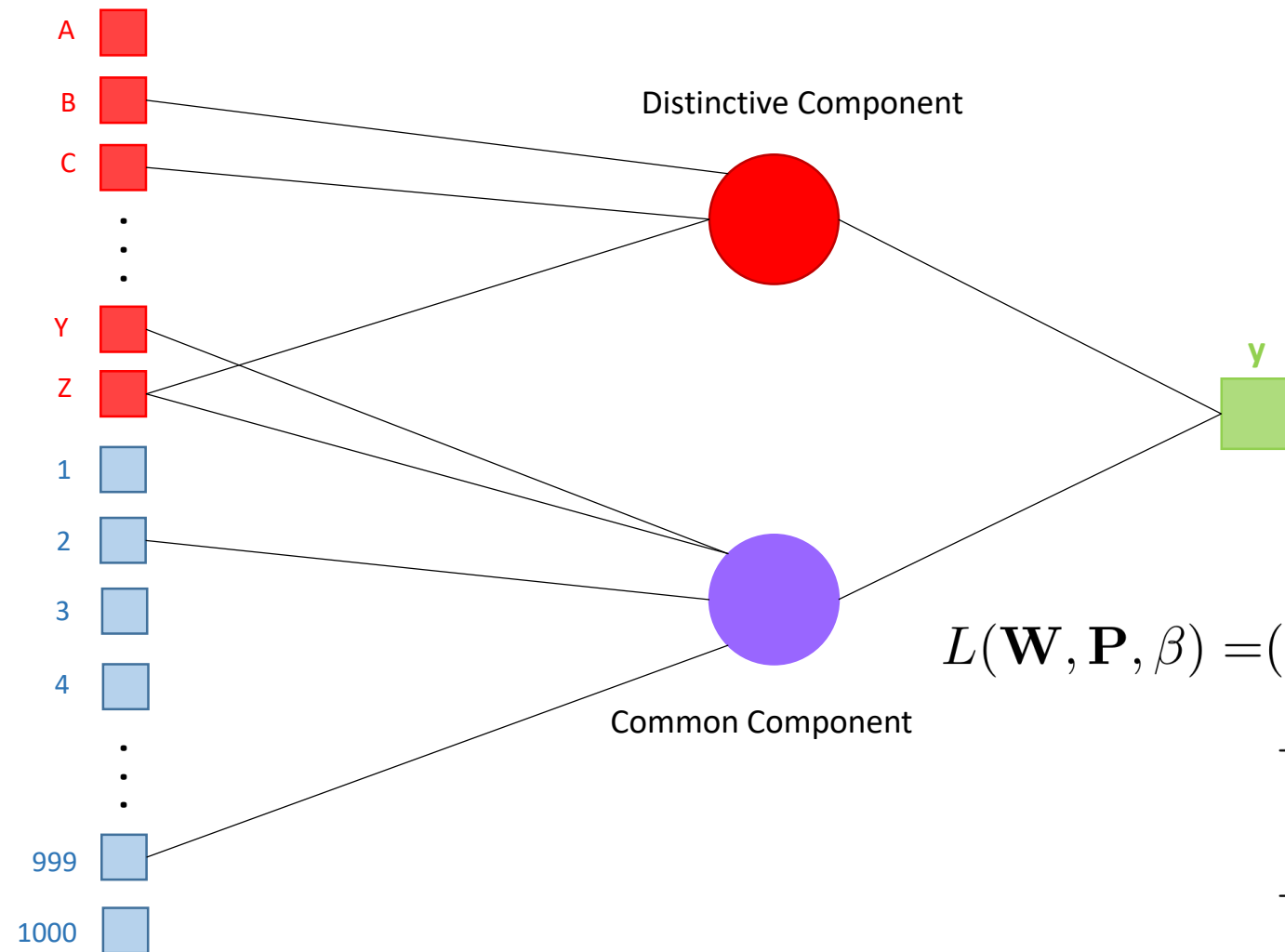


Sparse Common and Distinctive SCA



$$L(\mathbf{W}, \mathbf{P}) = \|\mathbf{X} - \mathbf{X}\mathbf{W}\mathbf{P}^T\|_2^2 + \lambda_L |\mathbf{W}|_1 + \lambda_R \|\mathbf{W}\|_2^2$$

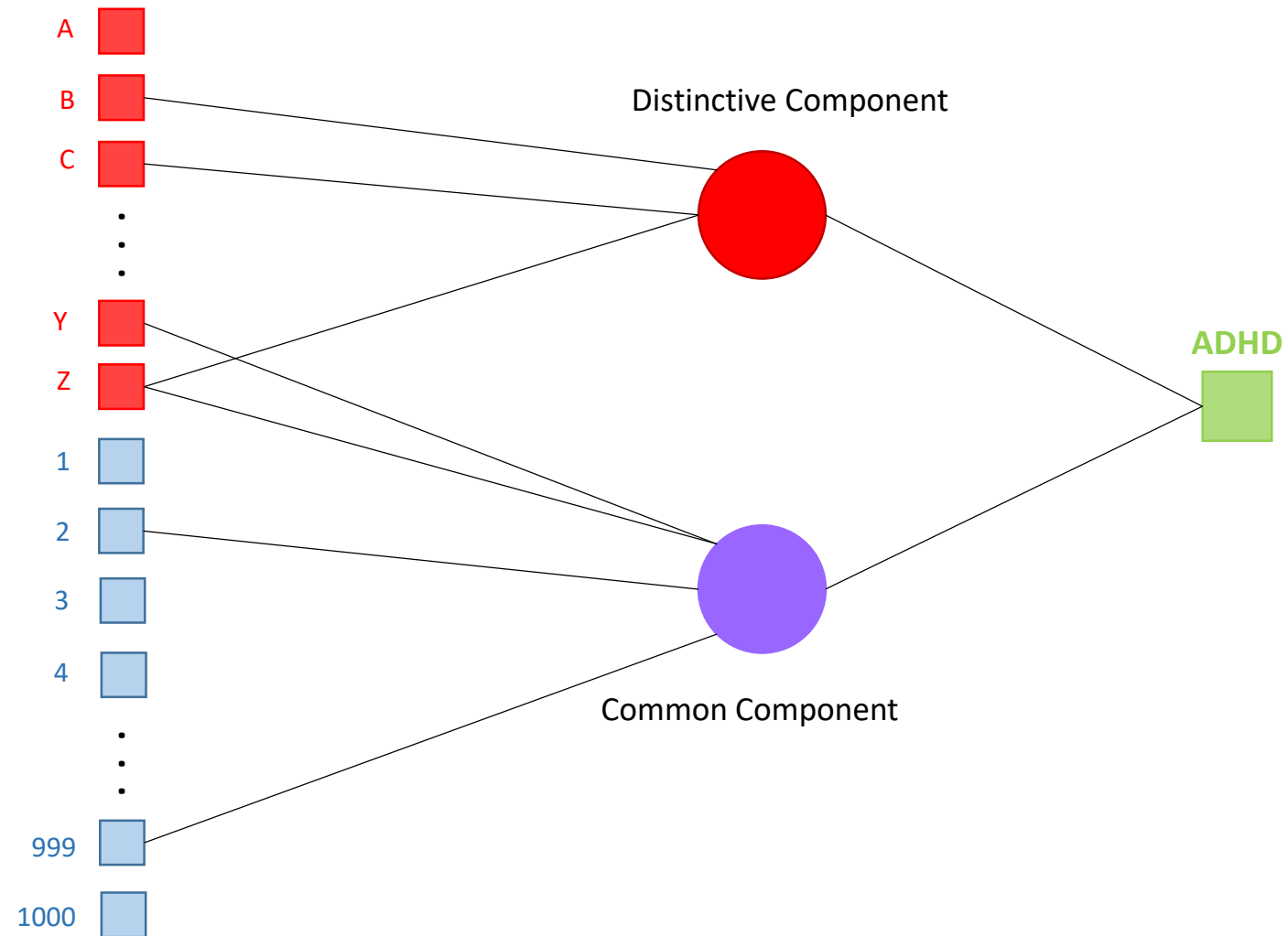
Sparse Common and Distinctive PCovR



$$L(\mathbf{W}, \mathbf{P}, \beta) = (1 - \alpha) \|\mathbf{X} - \mathbf{X}\mathbf{W}\mathbf{P}^T\|_2^2 + (\alpha) \|\mathbf{y} - \mathbf{X}\mathbf{W}\beta\|_2^2$$

$$+ \lambda_L |\mathbf{W}|_1 + \sum_r^R \sum_k^K \lambda_{Gr} \sqrt{J_k} \left\| \mathbf{w}_r^{(k)} \right\|_2$$

SCD-Cov-logR



SCD-Cov-logR objective function

$$\begin{aligned} L(\mathbf{W}, \mathbf{P}^{(X)}, \mathbf{p}^{(y)}, p_0^{(y)}) = & \alpha \left(- \sum_i^N y_i (p_0^{(y)} + \mathbf{x}_i^T \mathbf{W} \mathbf{p}^{(y)}) - \log(1 + e^{(p_0^{(y)} + \mathbf{x}_i^T \mathbf{W} \mathbf{p}^{(y)})}) \right) \\ & + (1 - \alpha) \sum_i^N \left\| \mathbf{x}_i - \mathbf{x}_i^T \mathbf{W} (\mathbf{P}^{(X)})^T \right\|_2^2 \\ & + \sum_r^R \lambda_{Lr} \|\mathbf{w}_r\|_1 + \sum_r^R \sum_k^K \lambda_{Gr} \sqrt{J_k} \left\| \mathbf{w}_r^{(k)} \right\|_2 + \lambda_R \left\| \mathbf{p}^{(y)} \right\|_2^2 \end{aligned}$$

Application: 500 Family data

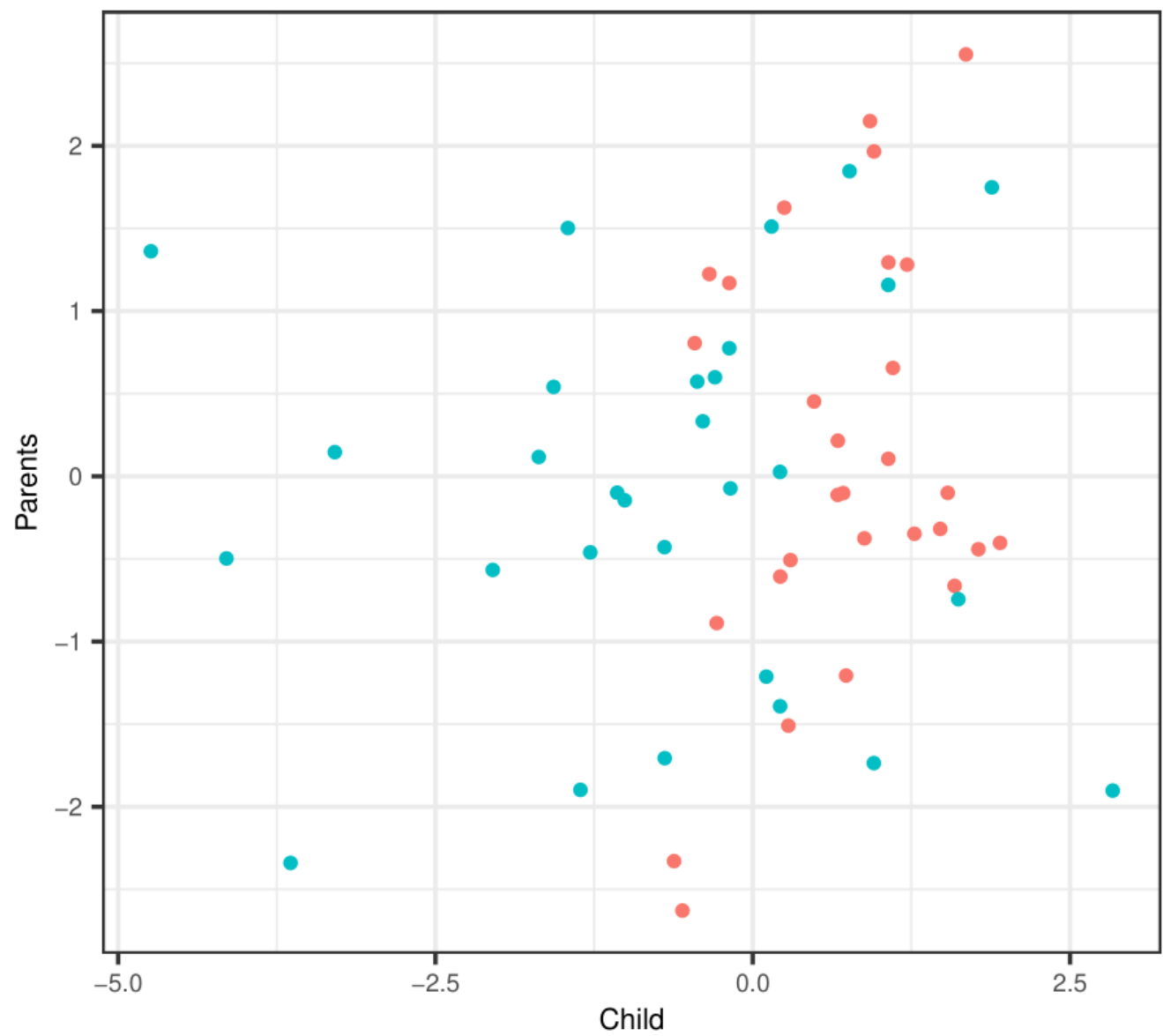
- Questionnaire data from different members of the same family concerning their well-being and family relationship
- 8 predictor variables from mother
- 8 predictor variables from father
- 6 predictor variables from child
- Each observation concerns a family unit
- Outcome variable: academic achievement (under vs. over)

$$\hat{W}_C$$

	Child	Parents
Mother		
Relationship with partners	0	0.231
Argue with partners	0	0.234
Childs bright future	0	0
Activities with children	0	0
Feeling about parenting	0	0.141
Communication with children	0	0.342
Argue with children	0	0.129
Confidence about oneself	0	0.395
Father		
Relationship with partners	0	0.077
Argue with partners	0	0.215
Childs bright future	0	0
Activities with children	0	0
Feeling about parenting	0	0
Communication with children	0	0
Argue with children	0	0.279
Confidence about oneself	0	0.007
Child		
Self confidence/esteem	0.258	0
Social life and extracurricular	0.303	0
Importance of friendship	0.468	0
Self Image	0.410	0
Happiness	0.357	0
Confidence about the future	0.265	0

Logistic regression coefficients

Estimated	
Child	0.631
Parents	0.043
Intercept	-0.050



Overachievement Underachievement

Thank you for tuning in

Soogeun Park (s.park_1@tilburguniversity.edu)

References

- **Sparse Common and Distinctive SCA:**
 - de Schipper, N. C., & Van Deun, K. (2019). Revealing the joint mechanisms in traditional data linked with big data. *Zeitschrift für Psychologie*.
- **Sparse PCovR:**
 - Van Deun, K., Cromptvoets, E. A., & Ceulemans, E. (2018). Obtaining insights from high-dimensional data: sparse principal covariates regression. *BMC bioinformatics*, 19(1), 1-13.
- **Sparse Common and Distinctive PCovR:**
 - Park, S., Ceulemans, E., & Van Deun, K. (2021). Sparse common and distinctive covariates regression. *Journal of Chemometrics*, 35(2), e3270.

Github (under construction)

- Sparse Common and Distinctive Covariates Logistic Regression:
 - <https://github.com/soogs/SCD-Cov-logR>

