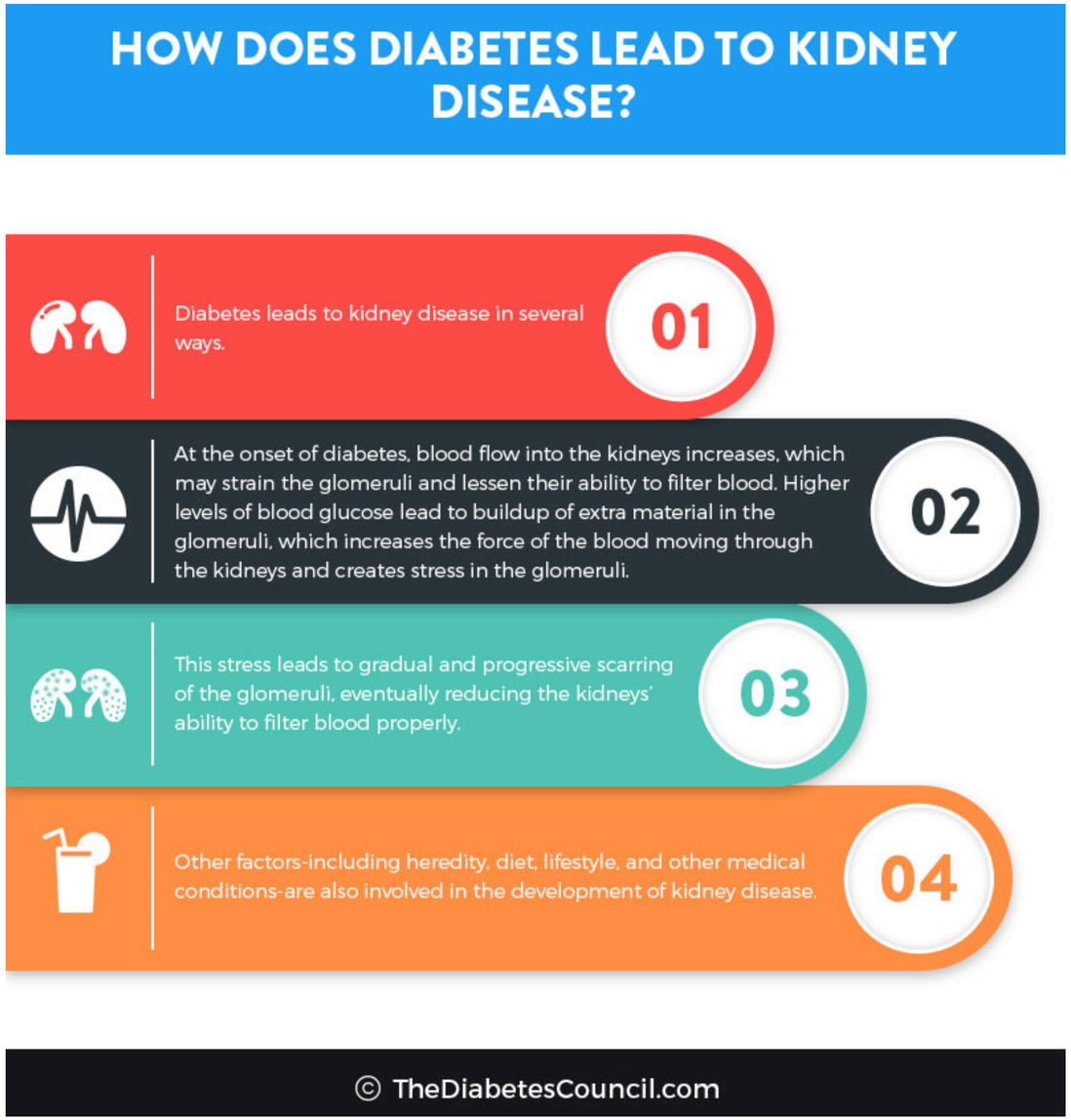
GAMAID: GREEDY CP TENSOR DECOMPOSITION FOR SUPERVISED EHR-BASED DISEASE TRAJECTORY DIFFERENTIATION

JETTE HENDERSON (UT-AUSTIN)
JOYCE C. HO (EMORY UNIVERSITY)
JOYDEEP GHOSH (UT-AUSTIN)

DIABETES & CHRONIC KIDNEY DISORDER

- Diabetes can cause diabetic nephropathy, a type of chronic kidney disease (CKD)
- 23% of diabetic patients suffer from CKD
  - Controlling glycemic levels are more challenging
  - Can result in complication of care

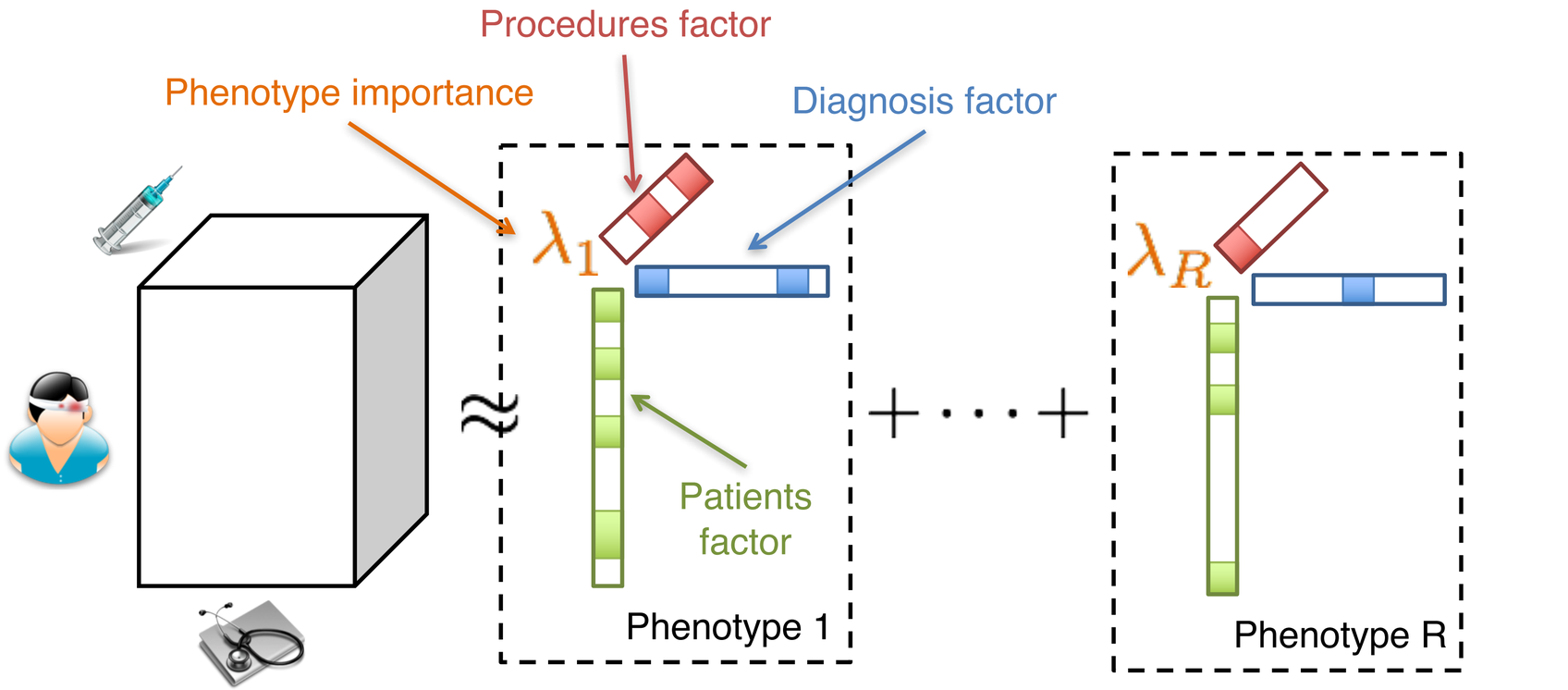


https://www.thediabetescouncil.com/diabetes-and-renal-failure-everything-you-need-to-know/

OVERVIEW: GAMAID

- Exploratory, supervised method to separate diabetic patients into two groups based on risk of developing diabetic nephropathy
- Model high-dimensional electronic health records (EHRs) using tensor to capture multi-way interaction (e.g., procedures used to treat diagnoses for a specific visit)
- Accumulate distinctive computational phenotypes that can differentiate patients with or without a disease

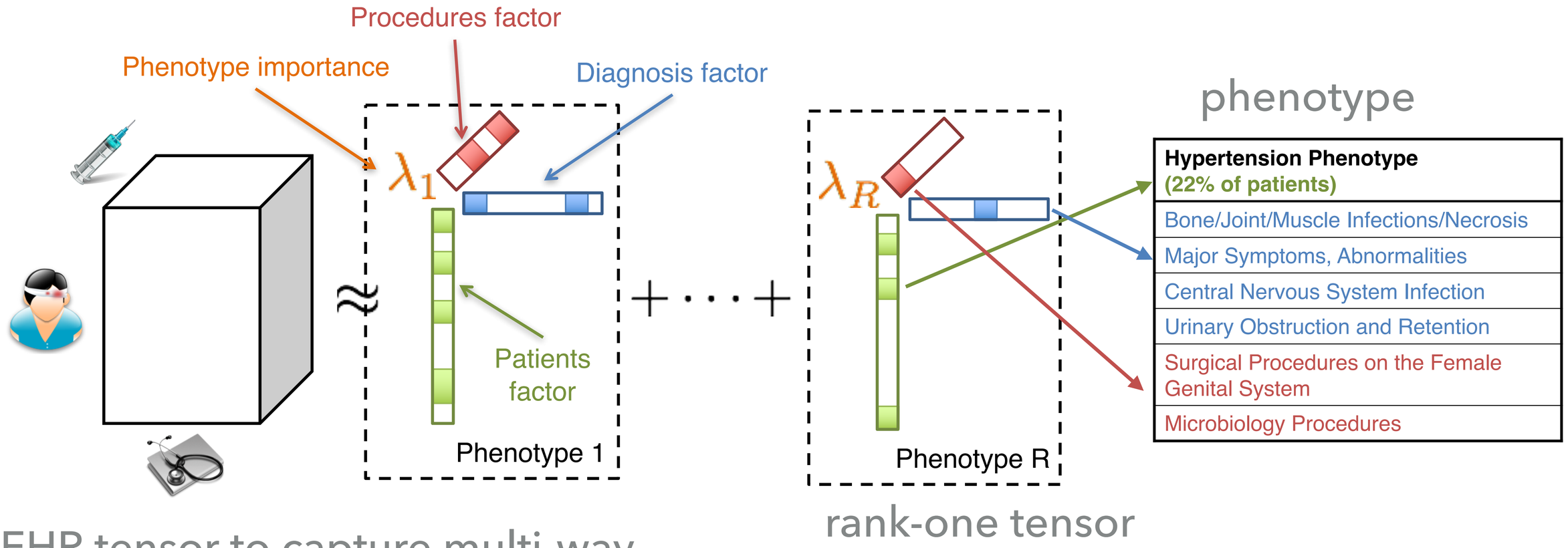
EHR-BASED COMPUTATIONAL PHENOTYPES VIA TENSOR FACTORIZATION



rank-one tensor

EHR tensor to capture multi-way interaction of diagnoses and procedures

EHR-BASED COMPUTATIONAL PHENOTYPES VIA TENSOR FACTORIZATION

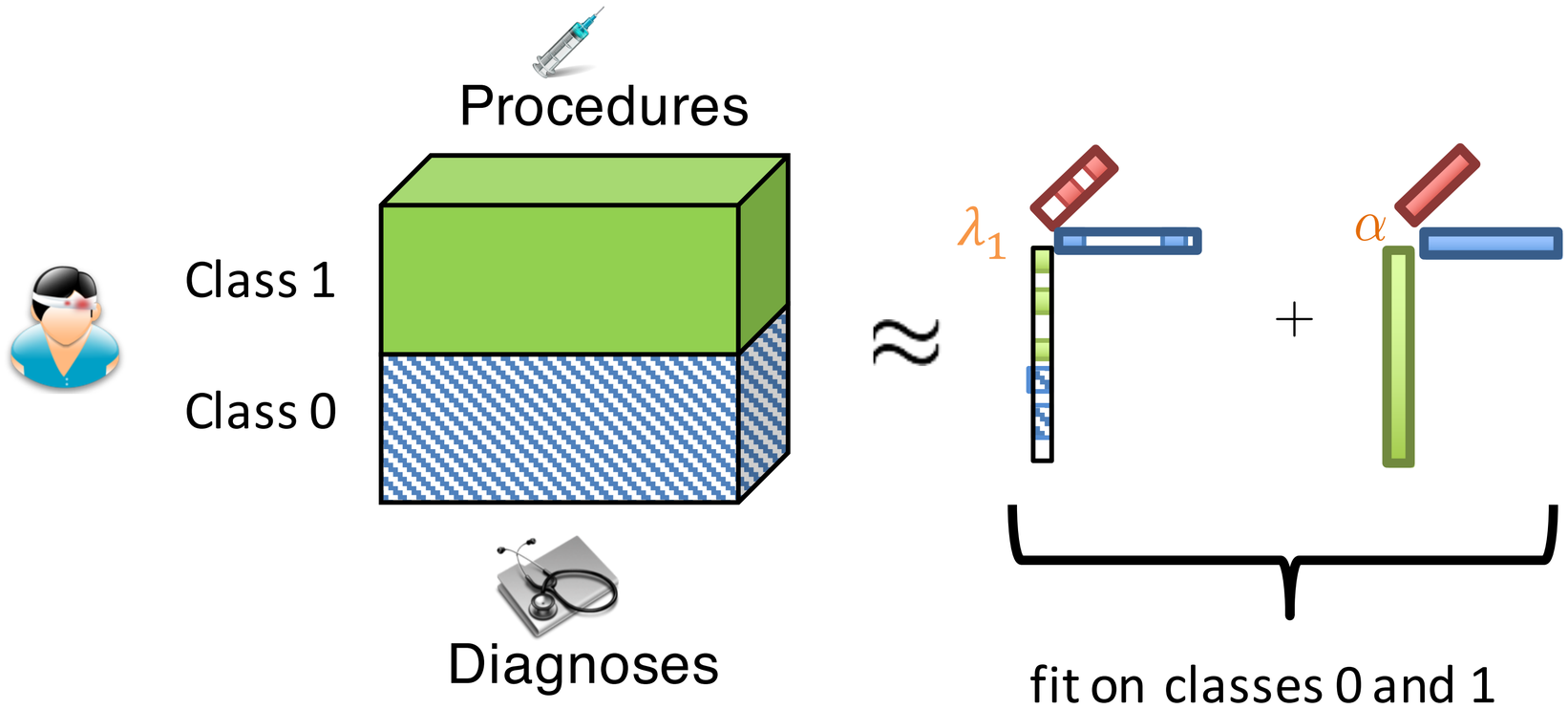


EHR tensor to capture multi-way interaction of diagnoses and procedures

GAMAID: GREEDY CP TENSOR DECOMPOSITION

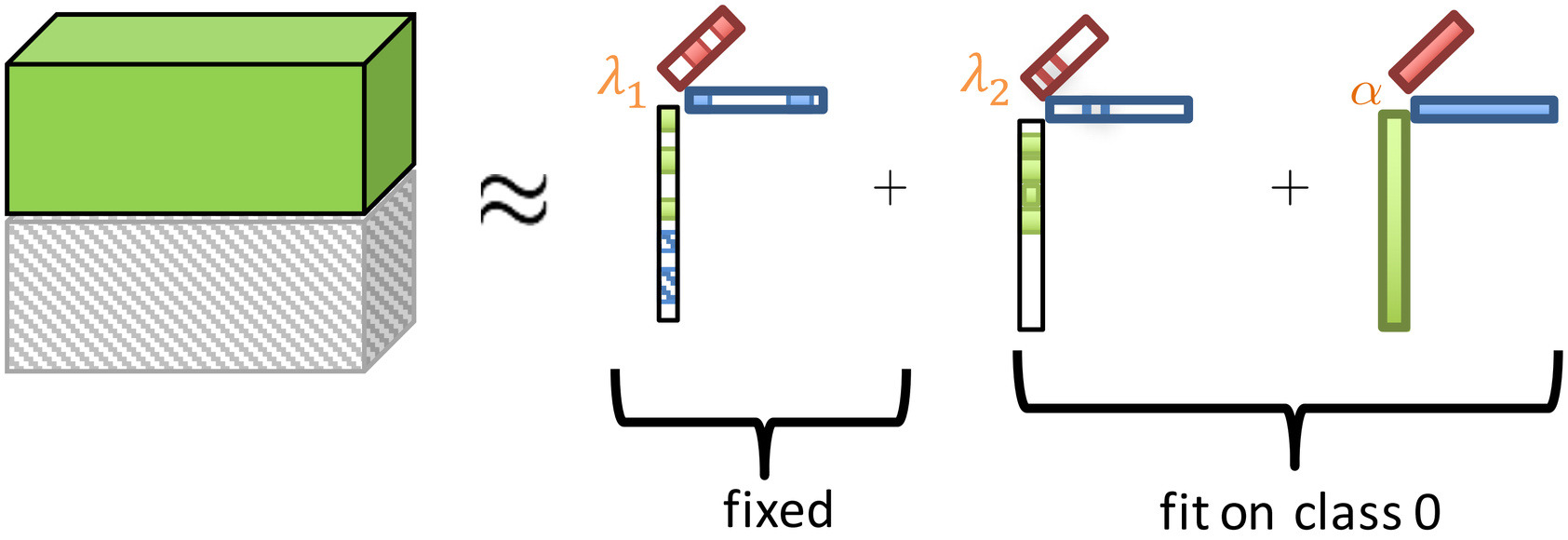
- Construct three tensors
  - One with data from both classes: $\mathcal{X}_{(01)}$
  - Two with data from each class: $\mathcal{X}_{(0)}, \mathcal{X}_{(1)}$
- Angular constraint to encourage diversity between discovered phenotypes
- Greedy algorithm to iteratively fit the best rank-one tensor

GAMAID: THE PROCESS



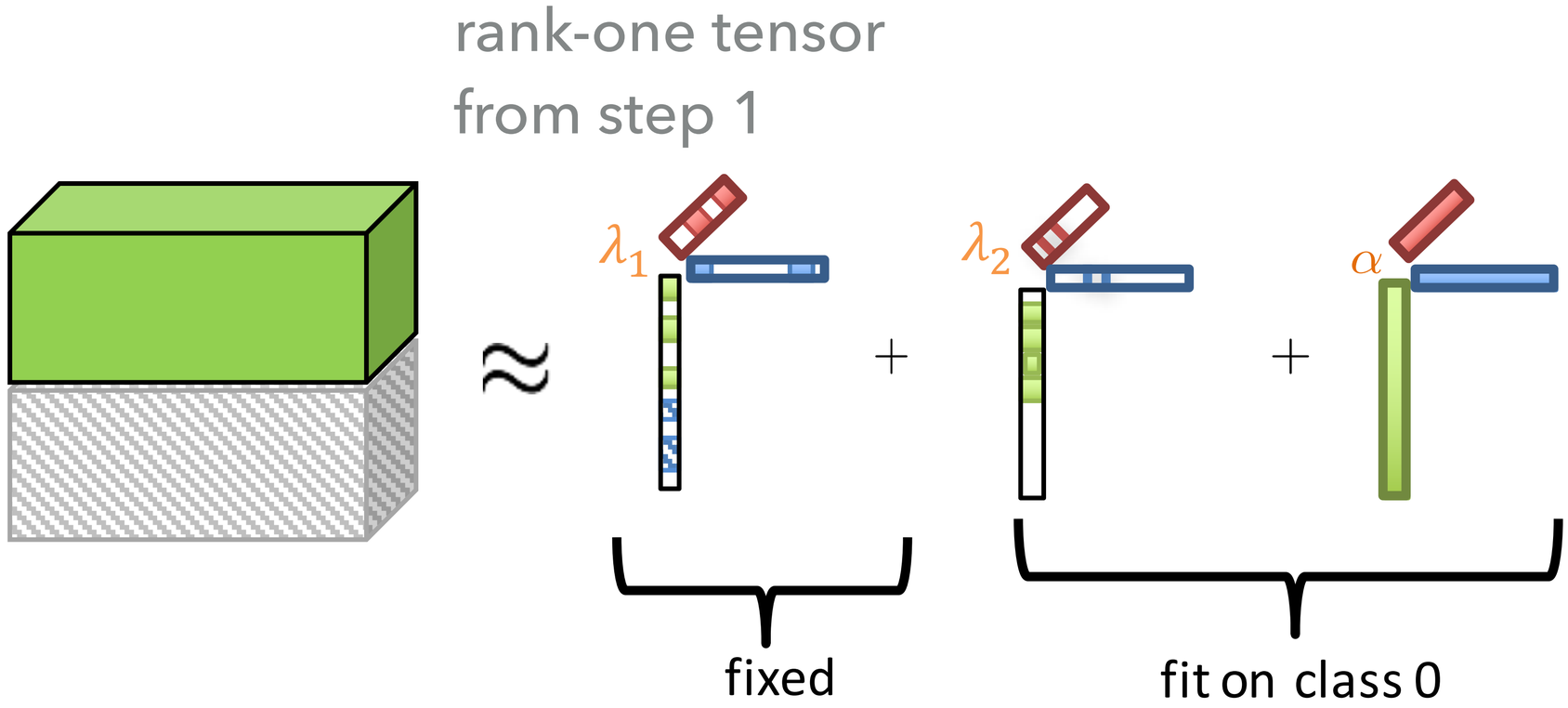
Step 1: Fit best rank-one approximation to tensor with both patients

GAMAID: THE PROCESS



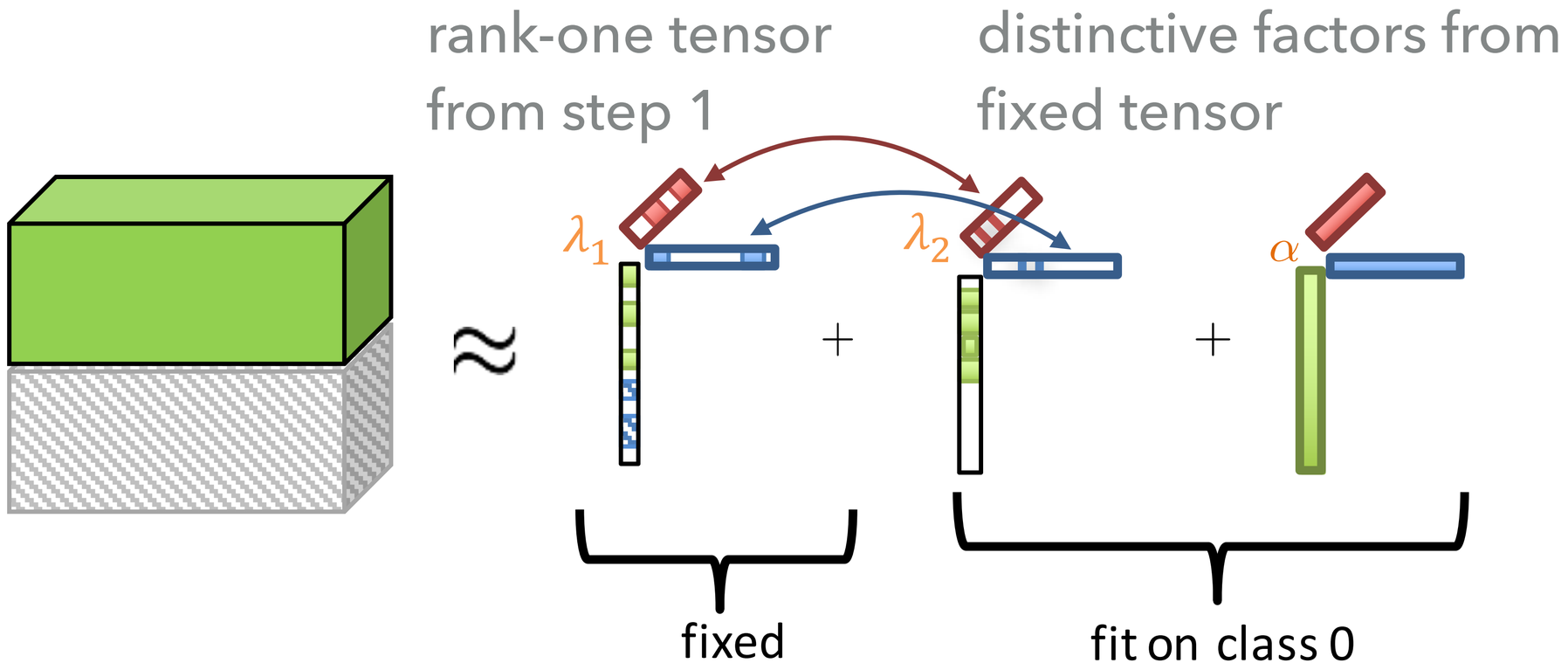
Step 2: Fit best rank-one approximation to class 1 tensor that is distinct from the first rank-one approximation

GAMAID: THE PROCESS



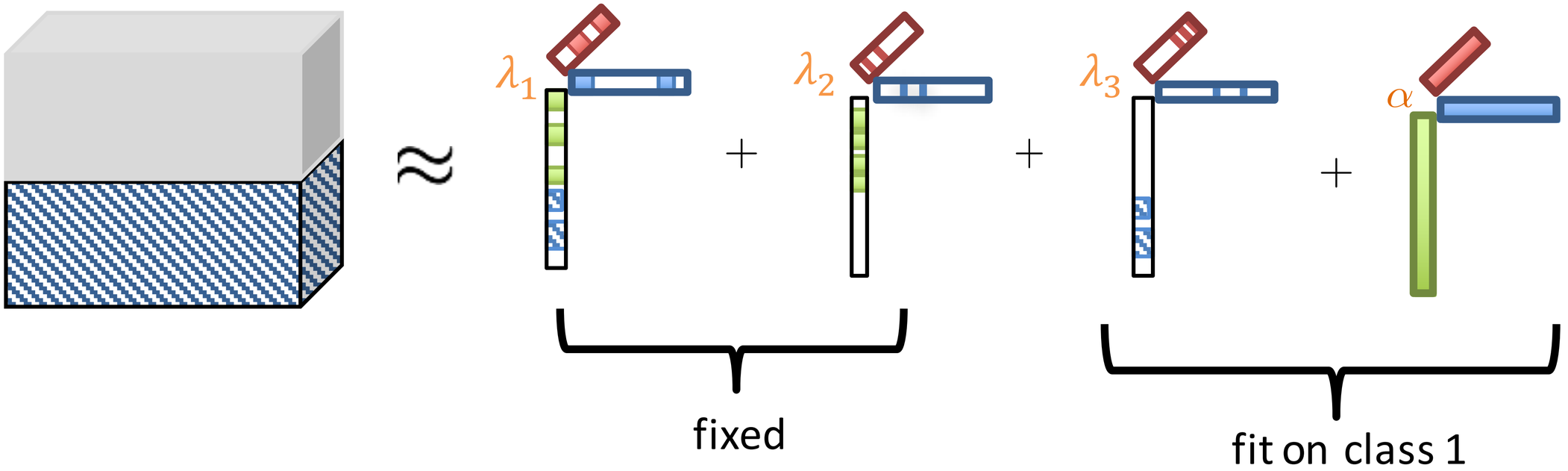
Step 2: Fit best rank-one approximation to class 1 tensor that is distinct from the first rank-one approximation

GAMAID: THE PROCESS



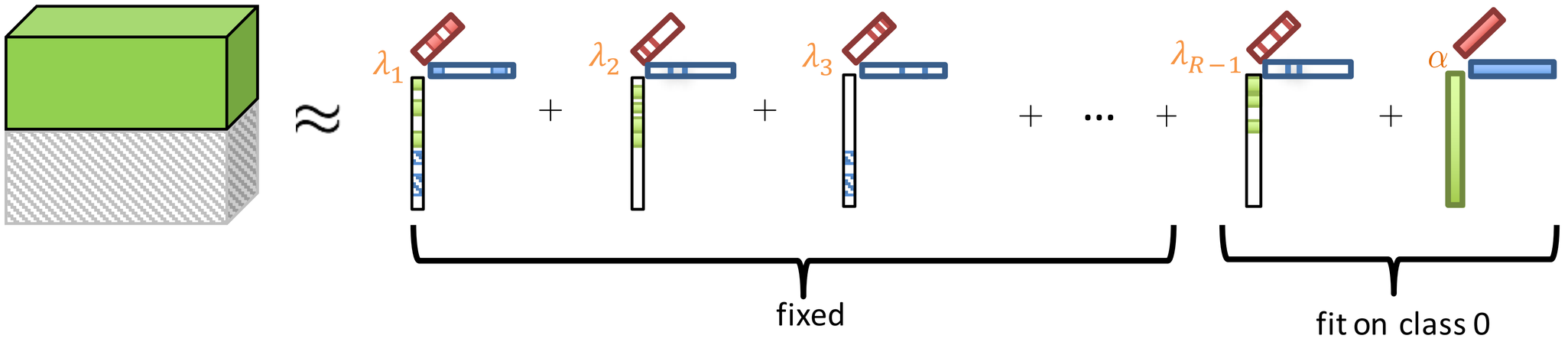
Step 2: Fit best rank-one approximation to class 1 tensor that is distinct from the first rank-one approximation

GAMAID: THE PROCESS



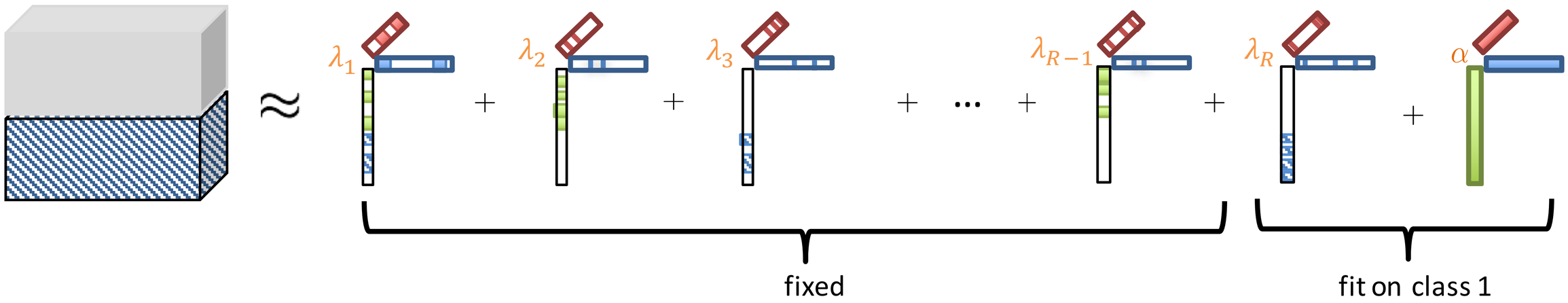
Step 3: Fit best rank-one approximation to class 0 tensor that is distinct from the other two rank-one approximation

GAMAID: THE PROCESS



Step R-1: Fit best rank-one approximation to class 1 tensor that is distinct from the other R-2 rank-one approximations

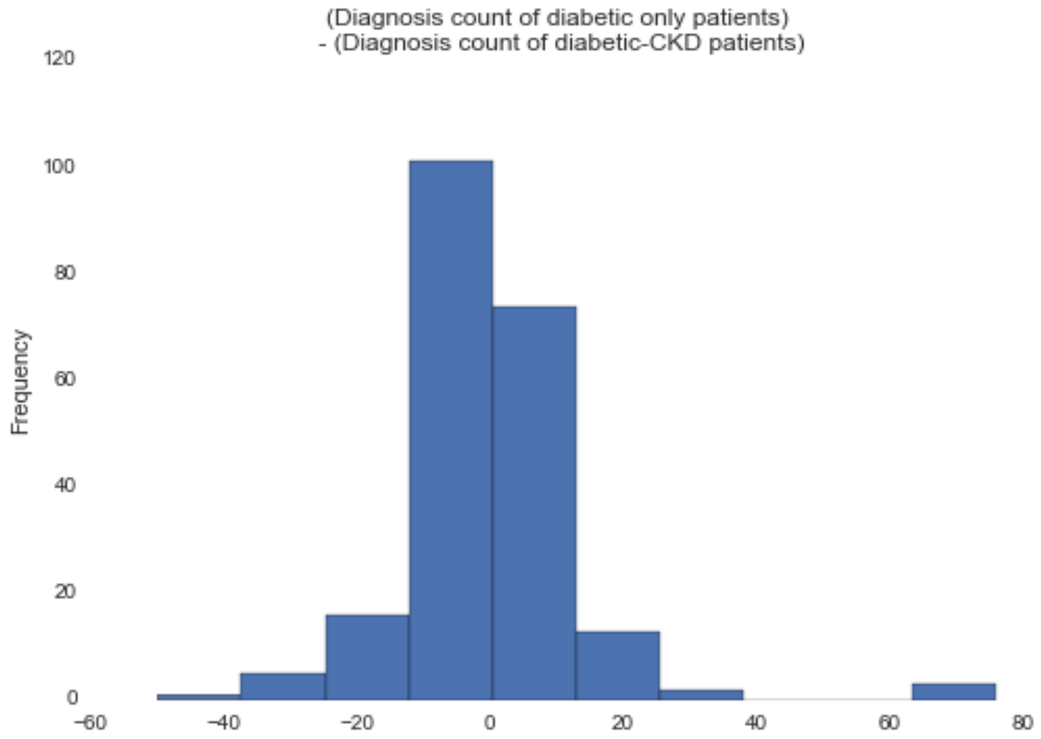
GAMAID: THE PROCESS



Step R: Fit best rank-one approximation to class 0 tensor that is distinct from the other R-1 rank-one approximations

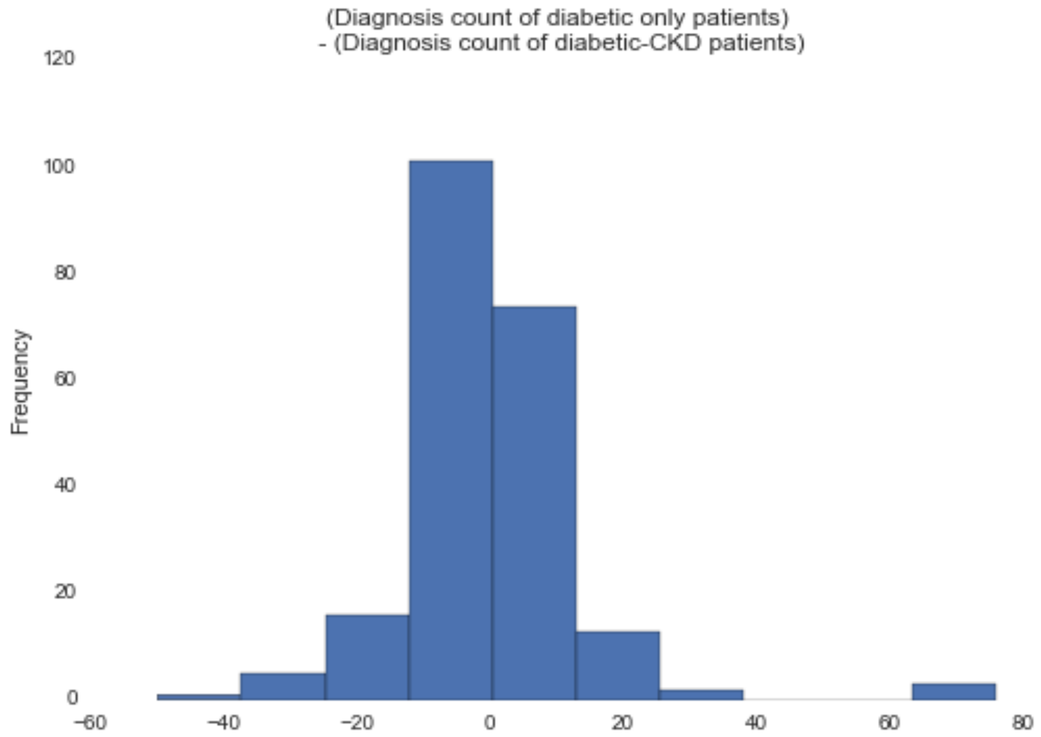
DATA: CMS DE-SYNPUF

- CMS Linkable 2008-2010 Medicare Data Entrepreneurs' Synthetic Public Use File (DE-SYNPUF)
- Focus on diabetic patients in 2009 and 2010
  - Class 1: 1,492 patients who did not have a CKD flag in 2009 but had one in 2010
  - Class 0: 1,625 patients who did not have a CKD flag in 2009 or 2010
- 3177 x 66 (diagnoses) x 198 (procedures)



DATA: CMS DE-SYNPUF

- CMS Linkable 2008-2010 Medicare Data Entrepreneurs' Synthetic Public Use File (DE-SYNPUF)
- Focus on diabetic patients in 2009 and 2010
  - Class 1: 1,492 patients who did not have a CKD flag in 2009 but had one in 2010
  - Class 0: 1,625 patients who did not have a CKD flag in 2009 or 2010
- 3177 x 66 (diagnoses) x 198 (procedures)



diagnoses that appear in class 1 than class 0

EXPERIMENTAL RESULTS: GAMAID

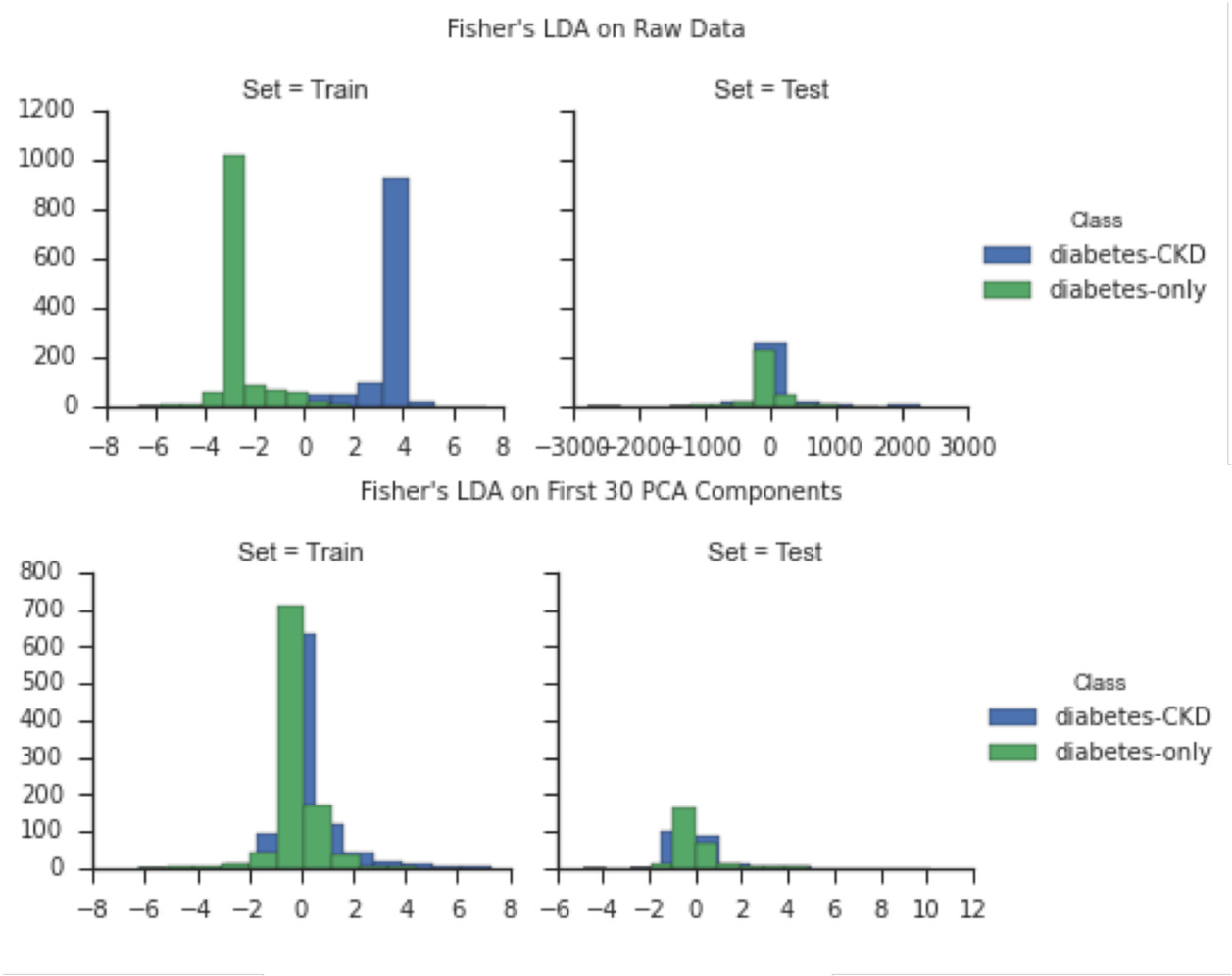
Phenotype 1
Other gastrointestinal disorders
Chronic obstructive pulmonary disease and bronchiectasis
Fluid and electrolyte disorders
Abdominal pain
Chronic ulcer of skin
Other circulatory disease
Cardiac dysrhythmias
Suture of skin and subcutaneous tissue
Routine chest X-ray
Other Laboratory
Electrocardiogram
Nonoperative urinary system measurements
Microscopic examination (bacterial smear, culture, toxicology)
Other diagnostic procedures (interview, evaluation, consultation)

Phenotype 6
Spondylosis; intervertebral disc disorders; other back problems
Physical therapy exercises, manipulation, and other procedures

Phenotype 9
Other gastrointestinal disorders
Other diagnostic radiology and related techniques

Phenotype	% Class 1	% Class 0	% Population
1	0.52	0.48	0.08
2	0.49	0.51	0.80
3	0.48	0.52	0.10
4	0.48	0.52	0.21
5	0.48	0.52	0.17
6	0.54	0.46	0.09
7	0.00	0.00	0.00
8	0.48	0.52	0.08
9	0.62	0.38	0.01

COMPARISON TO FISHER'S LDA



Model	SVM F1 Score
LDA	0.4783
LDA + PCA	0.3914
GAMAID	0.5106

DISCUSSION + CONCLUSION

- Greedy CP decomposition with angular constraints
  - Potential to tease out phenotypes of diverging disease population
  - Exploratory method that produces easy to interpret results from high-dimensional data
- Future work
  - Understand tuning parameters (sparsity, fitting class 0 vs class 1, etc.)
  - Test framework on other sets of diseases

Q&A

- Contact information:
  - Jette Henderson (jette@ices.utexas.edu)
  - Joyce C. Ho (joyce.c.ho@emory.edu)
  - Joydeep Ghosh (jghosh@utexas.edu)

TEXAS
The University of Texas at Austin

EMORY
UNIVERSITY