December 10, 2016

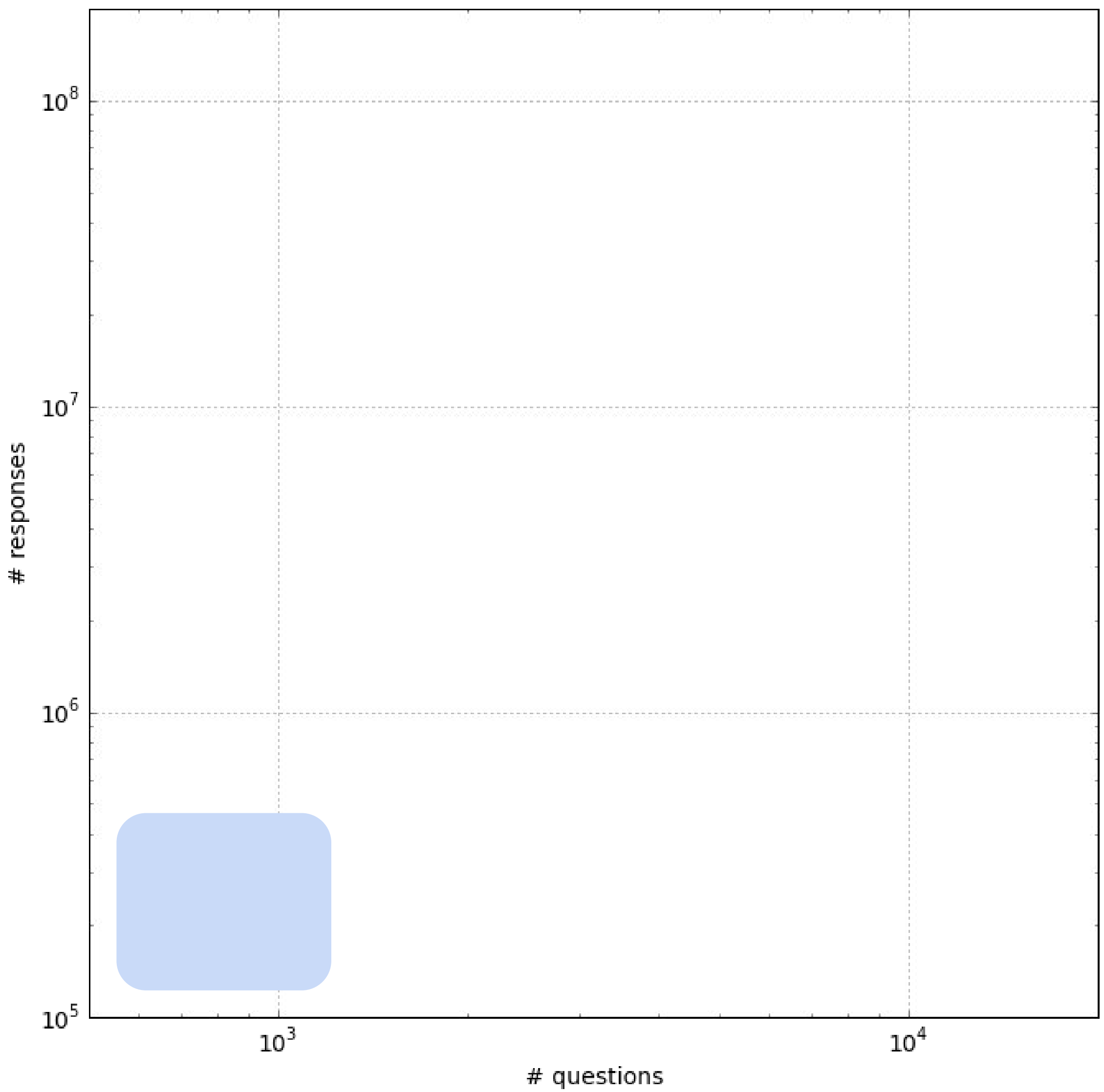
KNEWTON

yan karklin
principal data scientist

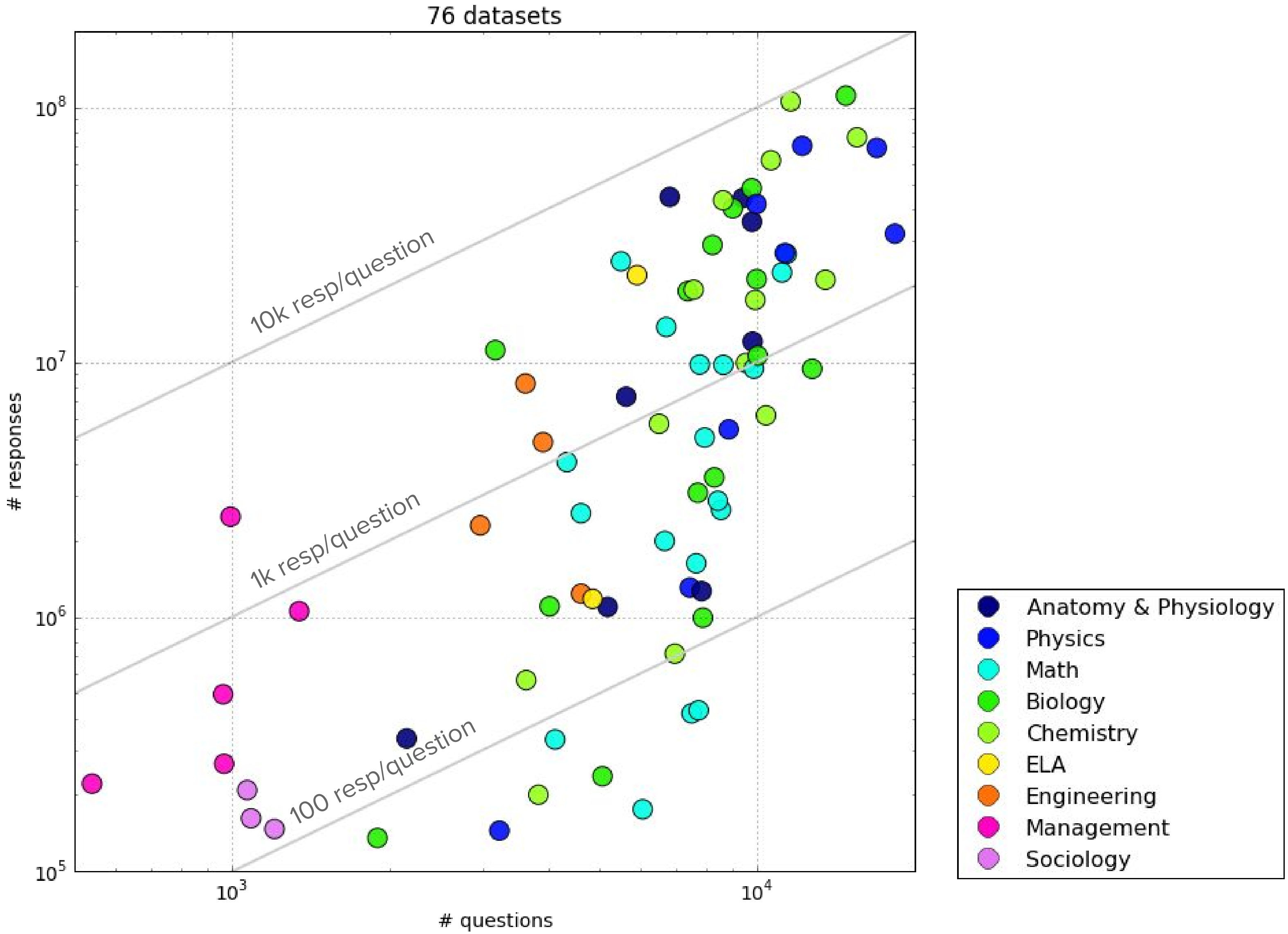
multilayer IRT

big data + little neural nets

big data in education



a subset of Knewton data

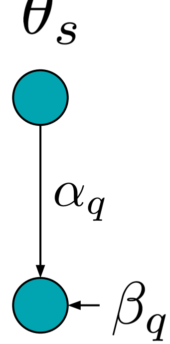


bigger data, more complex models

1. richer ***representation of ability***: learning many dimensions of student ability
2. more powerful ***models of assessment***: interactions between abilities
3. scalable, distributed implementation

(initial results, work in progress)

representation of student abilities



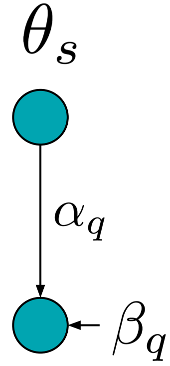
2 parameter Bayesian item response theory (IRT)

- scalar student proficiency
- scalar item discrimination
- scalar item difficulty

$$Pr(r_{s,q} = 1) = \sigma\left(\alpha_q \theta_s - \beta_q\right)$$

2 parameter Bayesian item response theory (IRT)

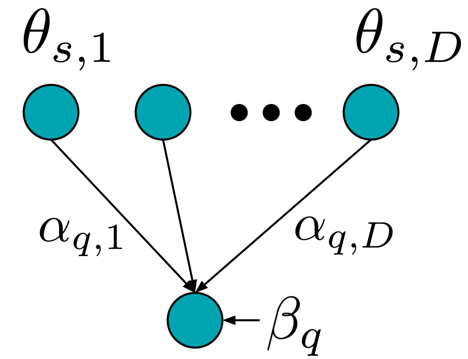
- scalar student proficiency
- scalar item discrimination
- scalar item difficulty



$$Pr(r_{s,q} = 1) = \sigma\left(\alpha_q \theta_s - \beta_q\right)$$

Multivariate IRT [Rasch 1960, Reckase 1972, Lan et al 2014]

- vector of student proficiency
- learned item-skill association vectors
- scalar item difficulty

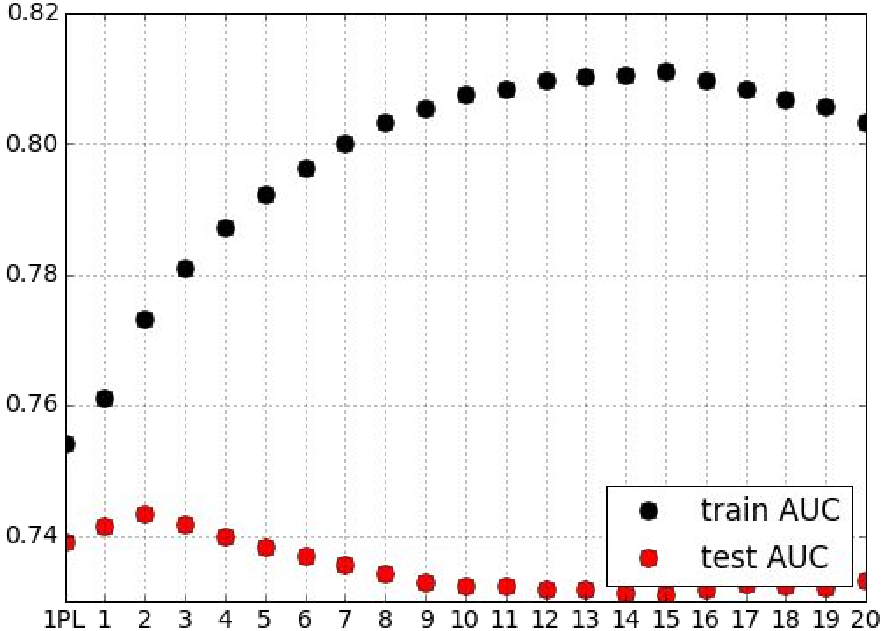


$$Pr(r_{s,q} = 1) = \sigma\left(\sum_{d}^{D} \alpha_{q,d}\theta_{s,d} - \beta_q\right)$$

prediction accuracy

- split responses randomly into train (80%) and test (20%)
- evaluate predictions of P(correct) on test set (AUC)

small data regime (300K events)

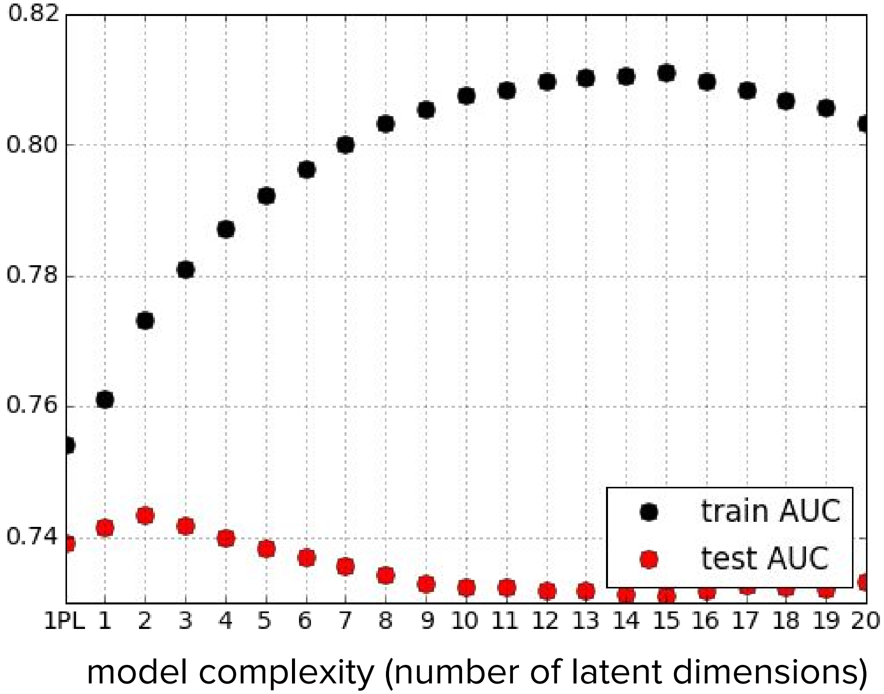


model complexity (number of latent dimensions)

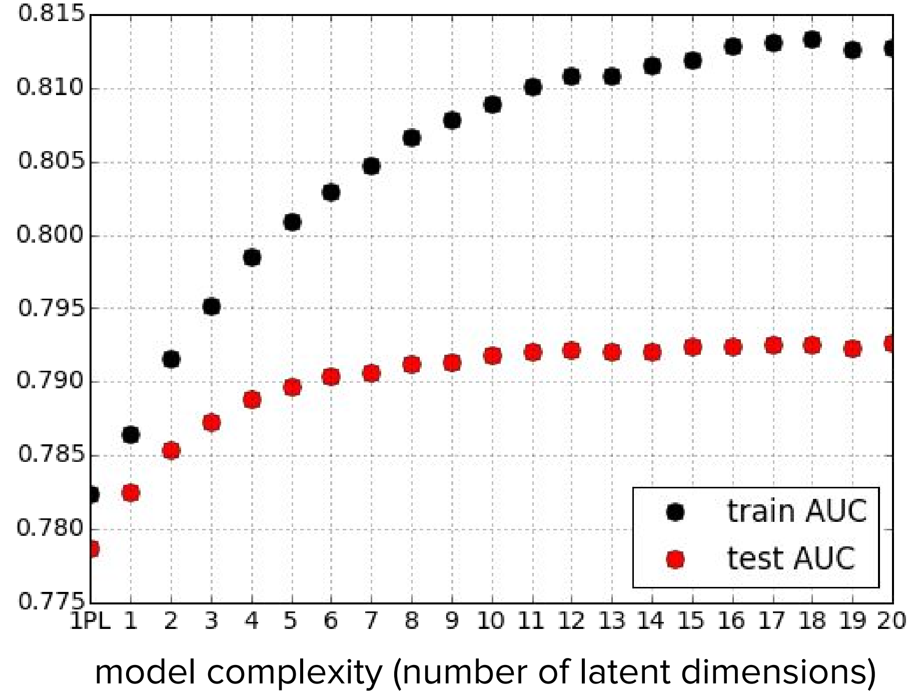
prediction accuracy

- split responses randomly into train (80%) and test (20%)
- evaluate predictions of P(correct) on test set (AUC)

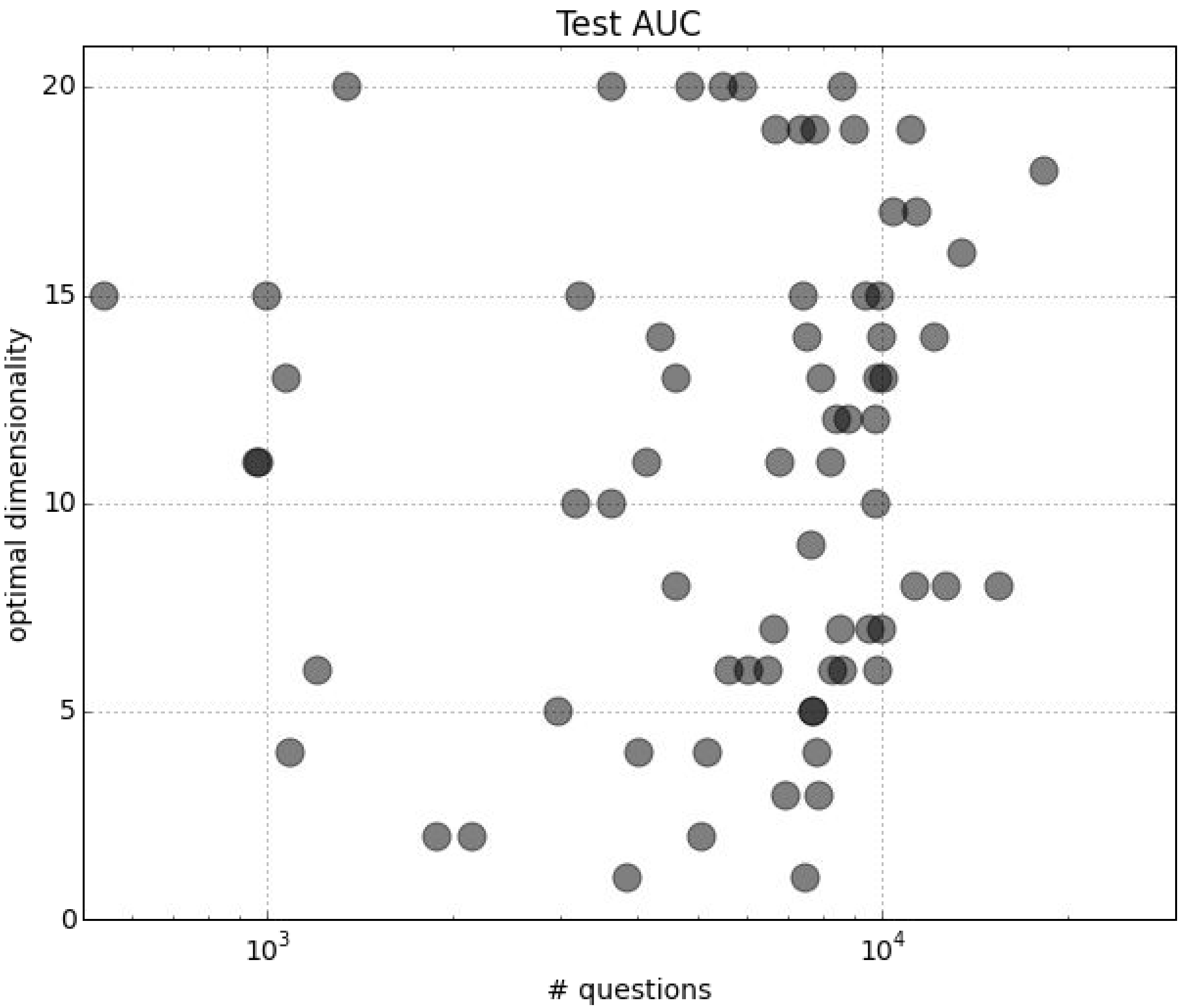
small data regime (300K events)



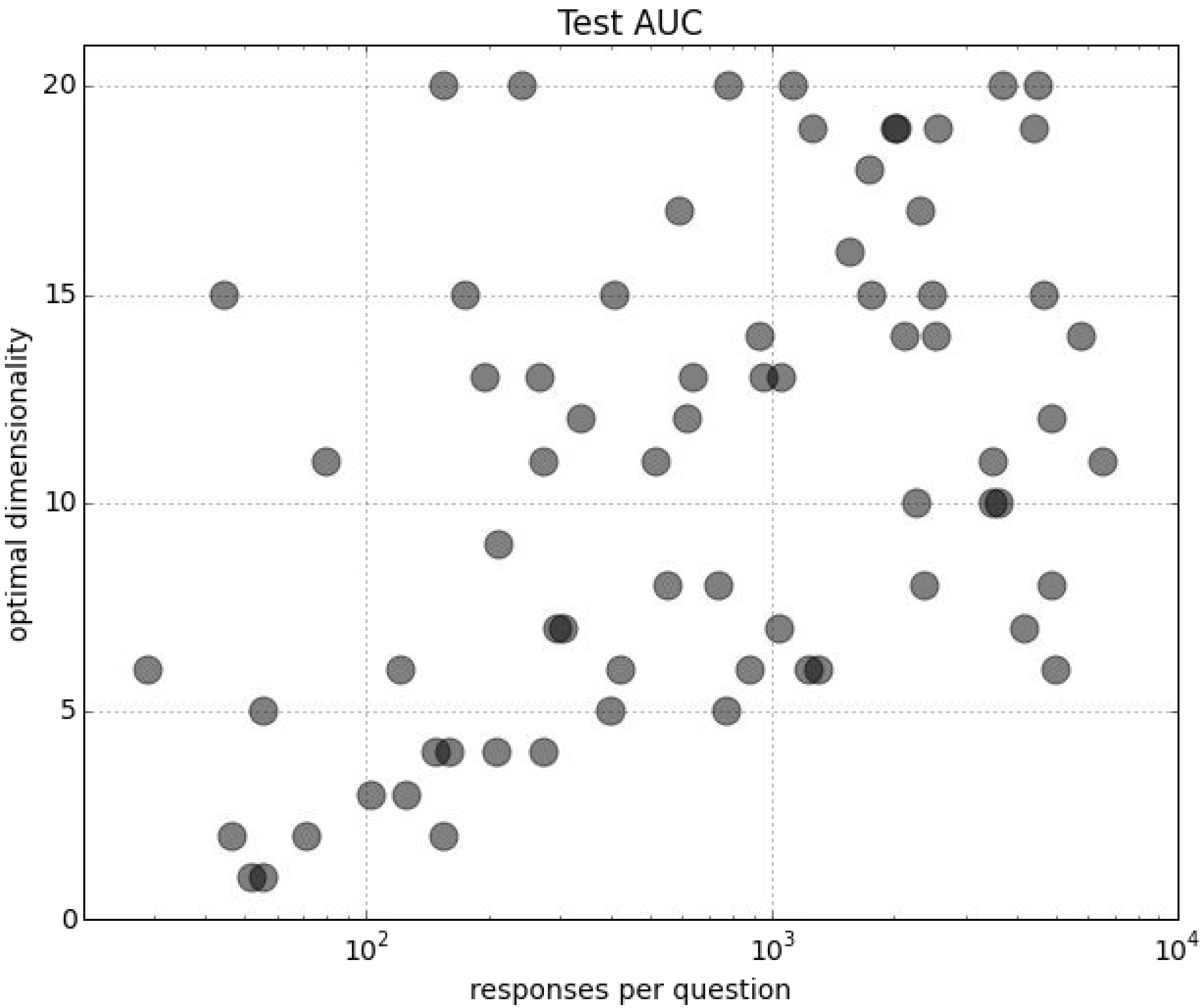
big data regime (25M events)



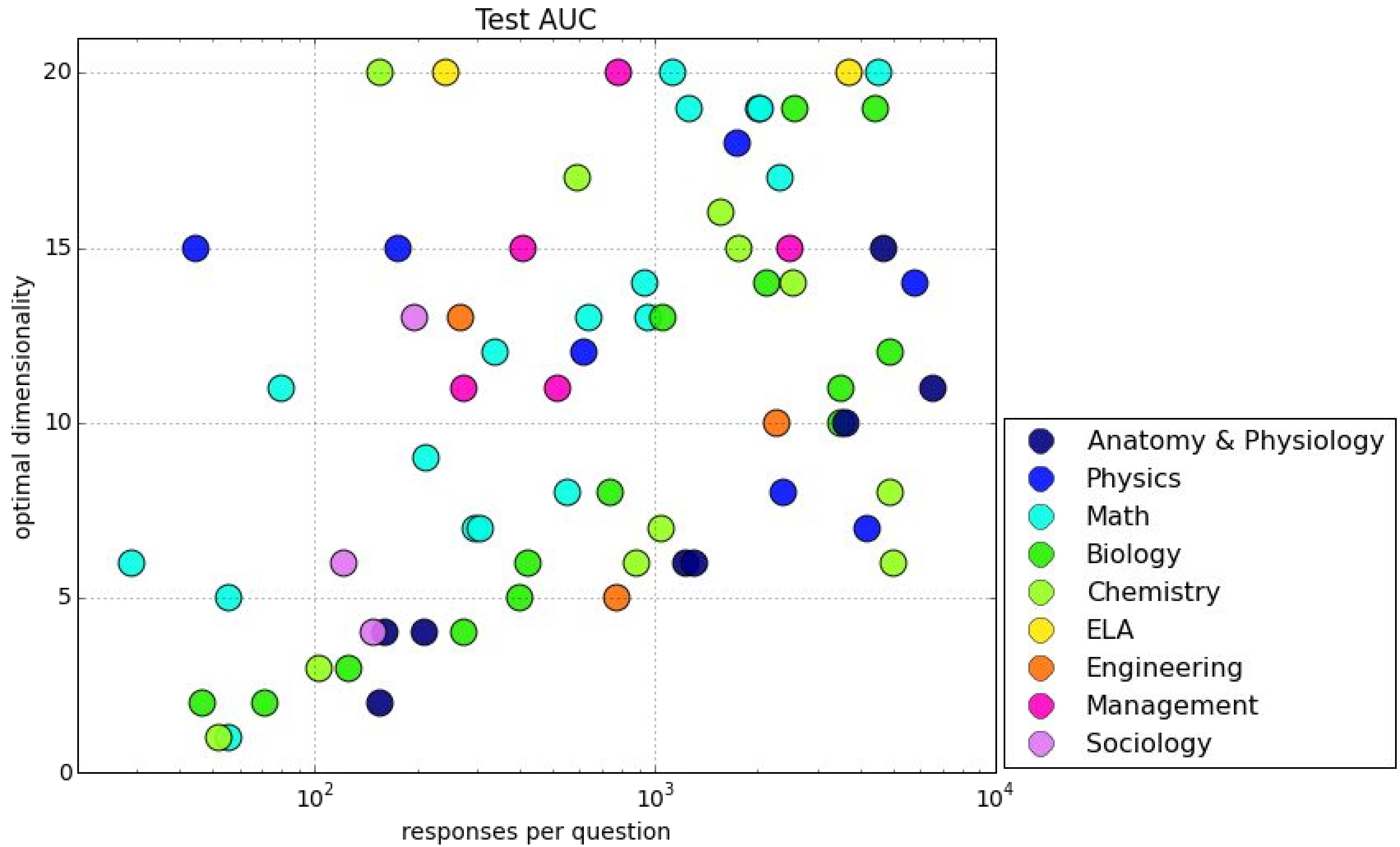
big data: more gain from complex models



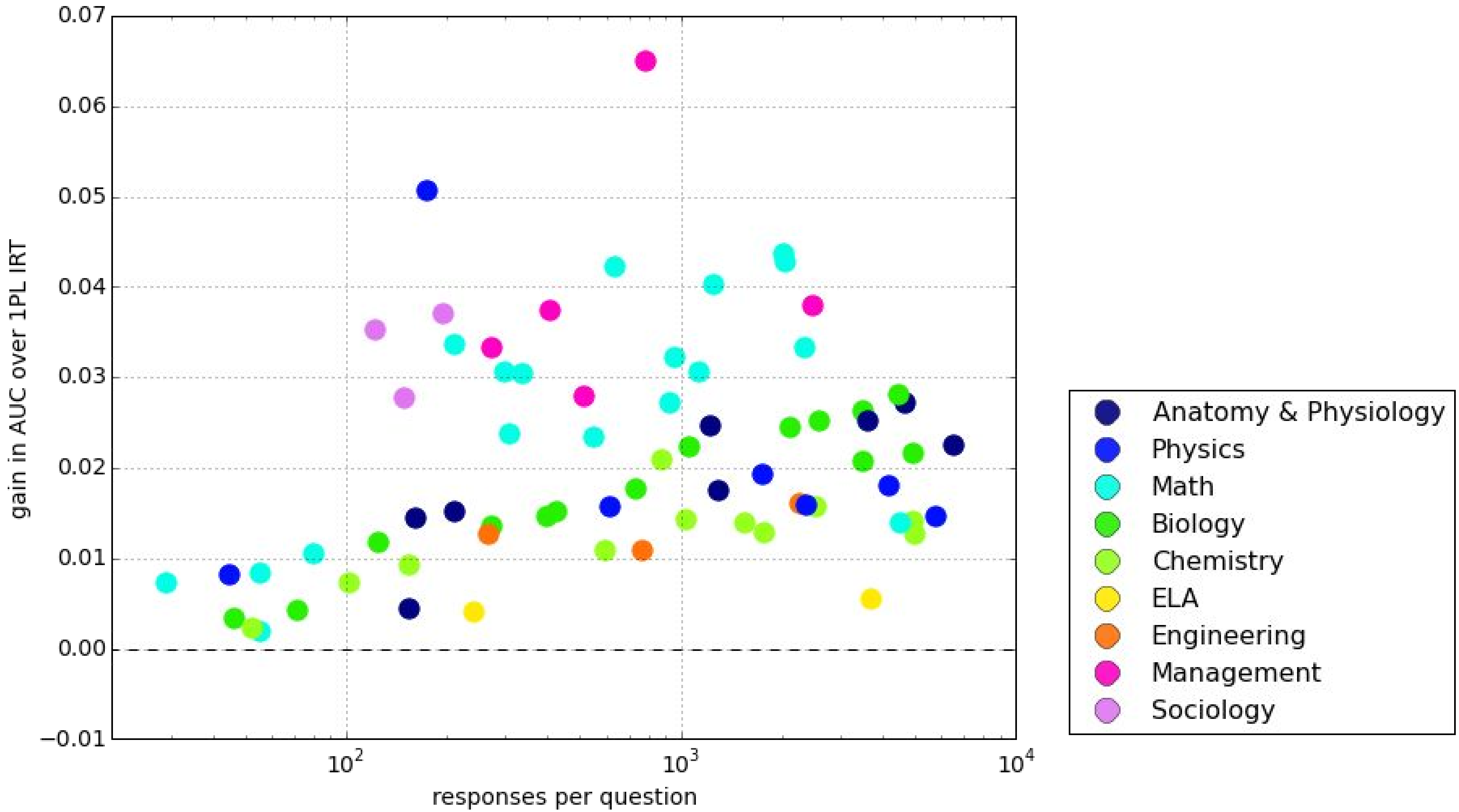
big data: more gain from complex models



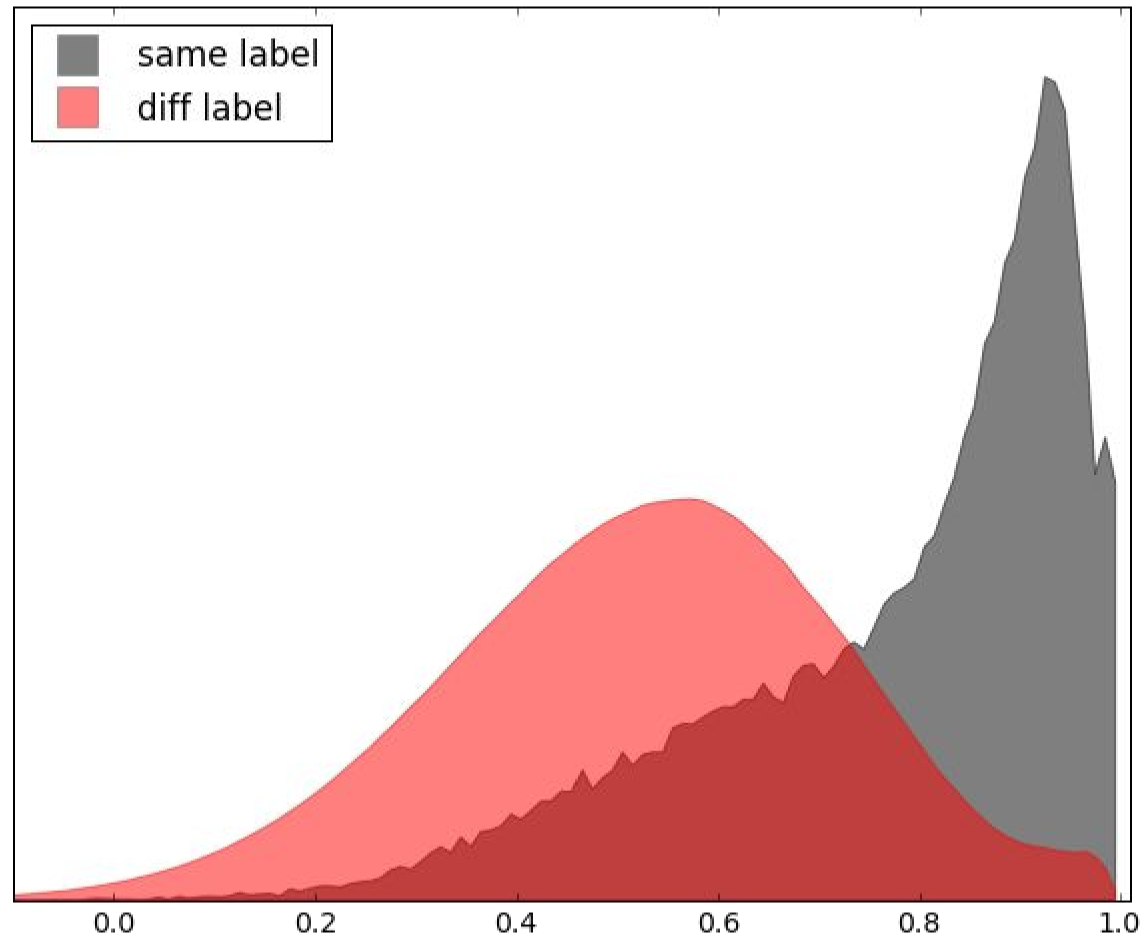
big data: more gain from complex models



prediction accuracy gains



learned item-concept mappings vs labels

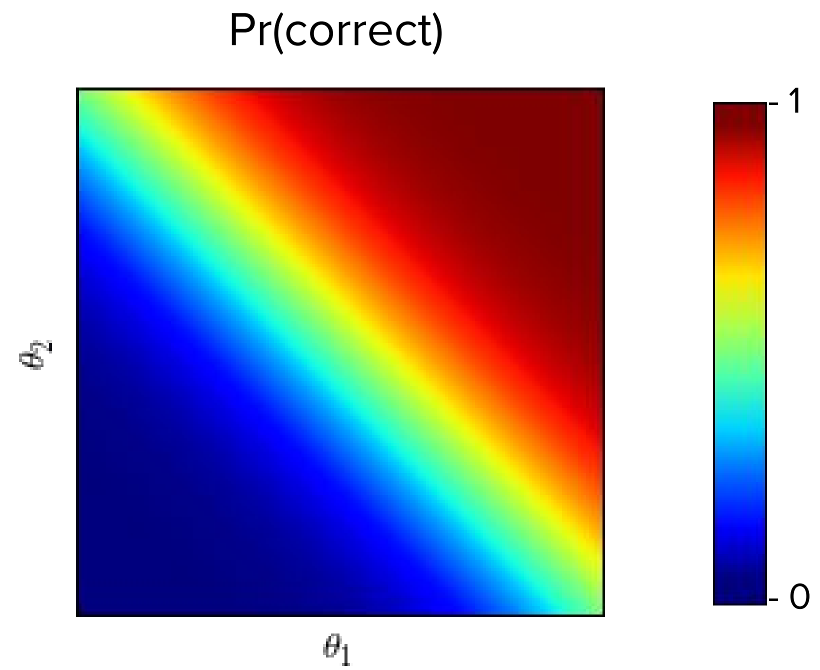


cosine similarity between pairs of 19-dim item-concept mappings ($D(\vec{\alpha}_i, \vec{\alpha}_j)$)

assessment models

assessment (measurement) models

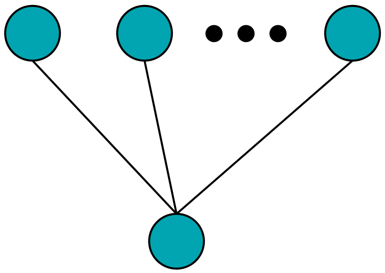
- compensatory $\sigma\left(\sum_{d}^{D} \alpha_{q,d}\theta_{s,d} - \beta_q\right)$
- non-compensatory
- our approach: learn inter-skill interactions using general functions



deep item response functions

- flexible model for non-compensatory relationships
- utilize neural net optimization machinery

$$Pr(r_{s,q} = 1) = \sigma\left(\sum_{d}^{D} \alpha_{q,d}\theta_{s,d} - \beta_q\right)$$



extend the linear-nonlinear architecture in multivariate IRT by making it deep

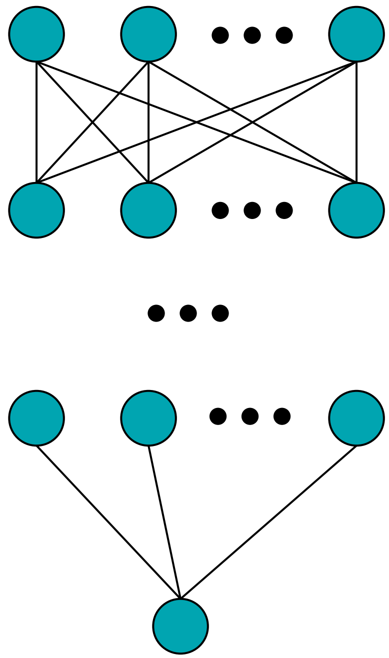
deep item response functions

- flexible model for non-compensatory relationships
- utilize neural net optimization machinery
- "deep" **and** probabilistic

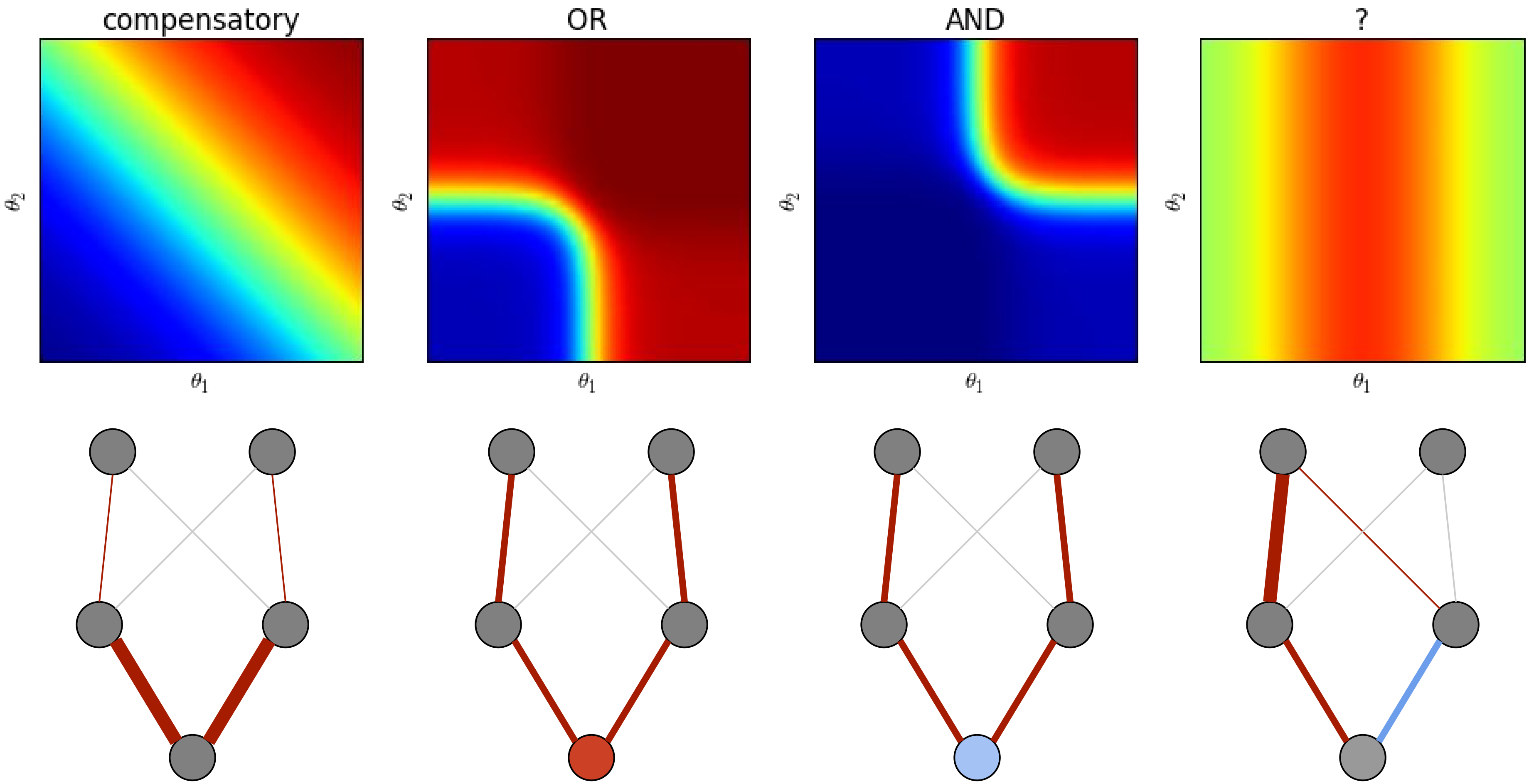
$$h_i^{(0)} = \theta_{is}$$

$$h_i^{(l)} = f\left(\sum_j w_{qij}^{(l)} h_j^{(l-1)} + \beta_{qi}^{(l)}\right)$$

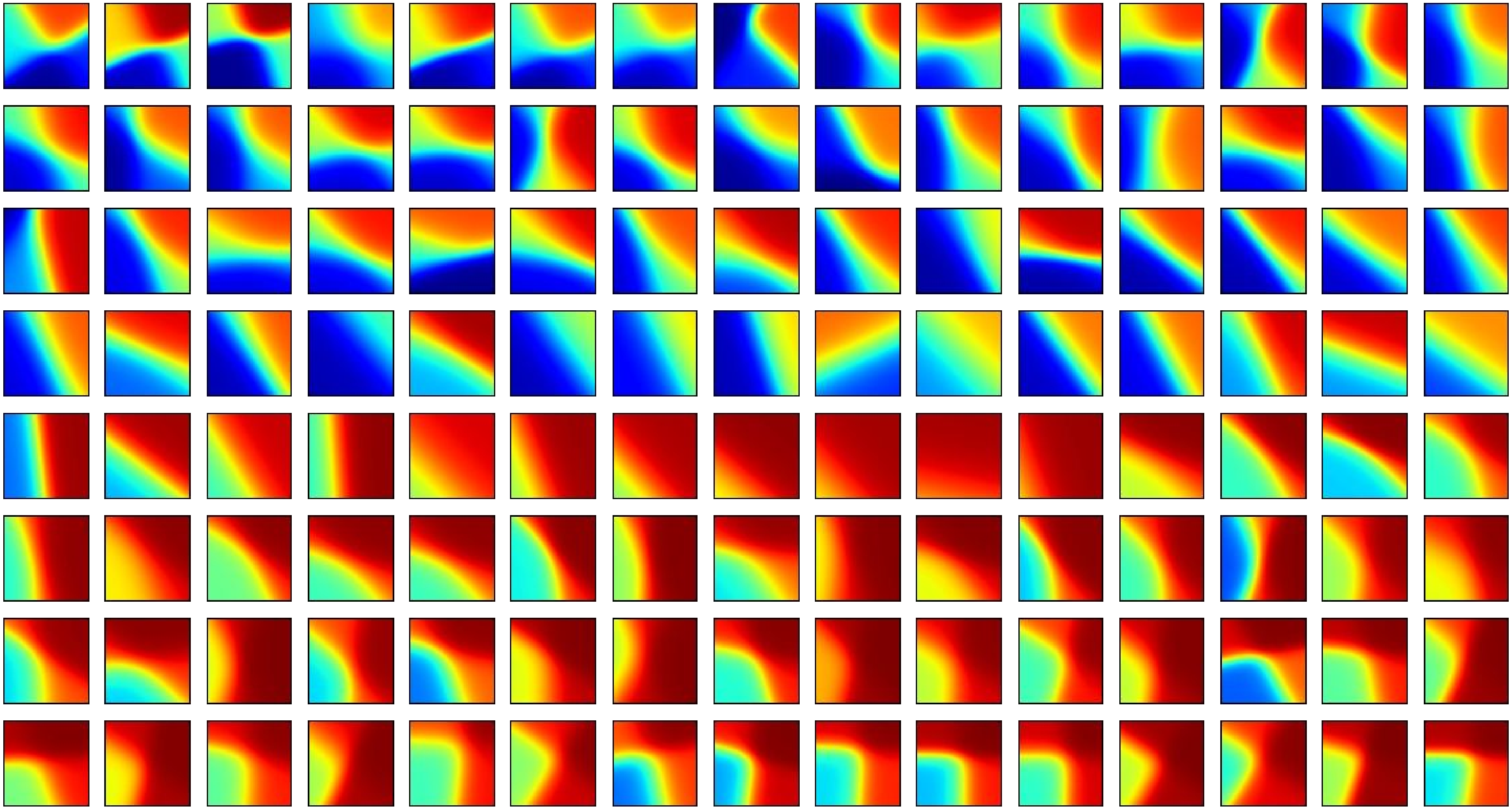
$$Pr(r_{s,q} = 1) = h^{(L)} = \sigma\left(\sum_j w_{qj}^{L} h_j^{(L-1)} + \beta_q^{(L)}\right)$$



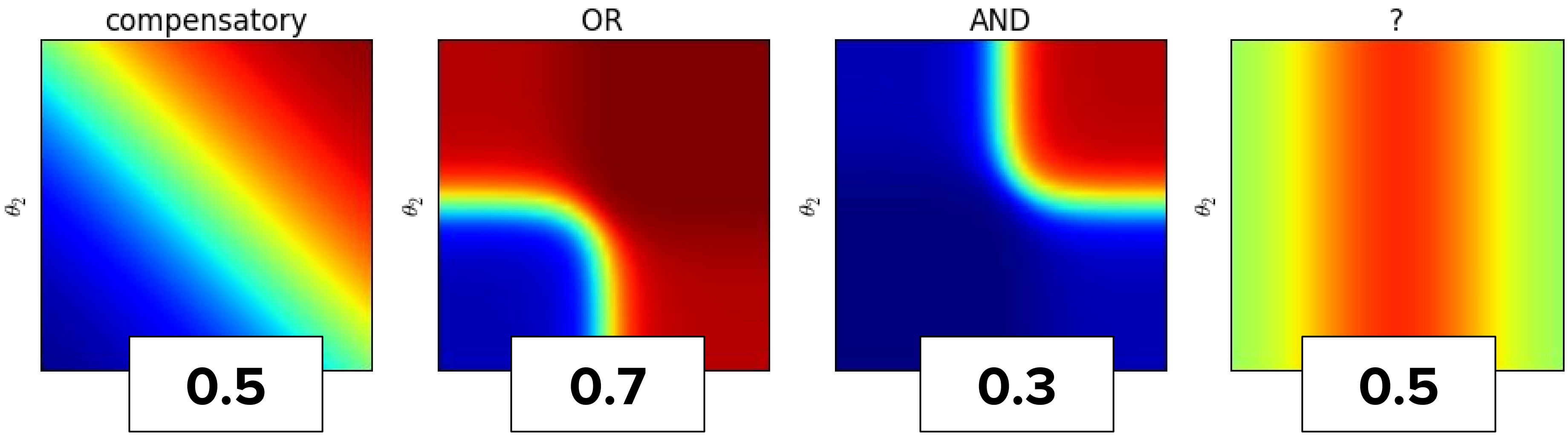
item response surfaces

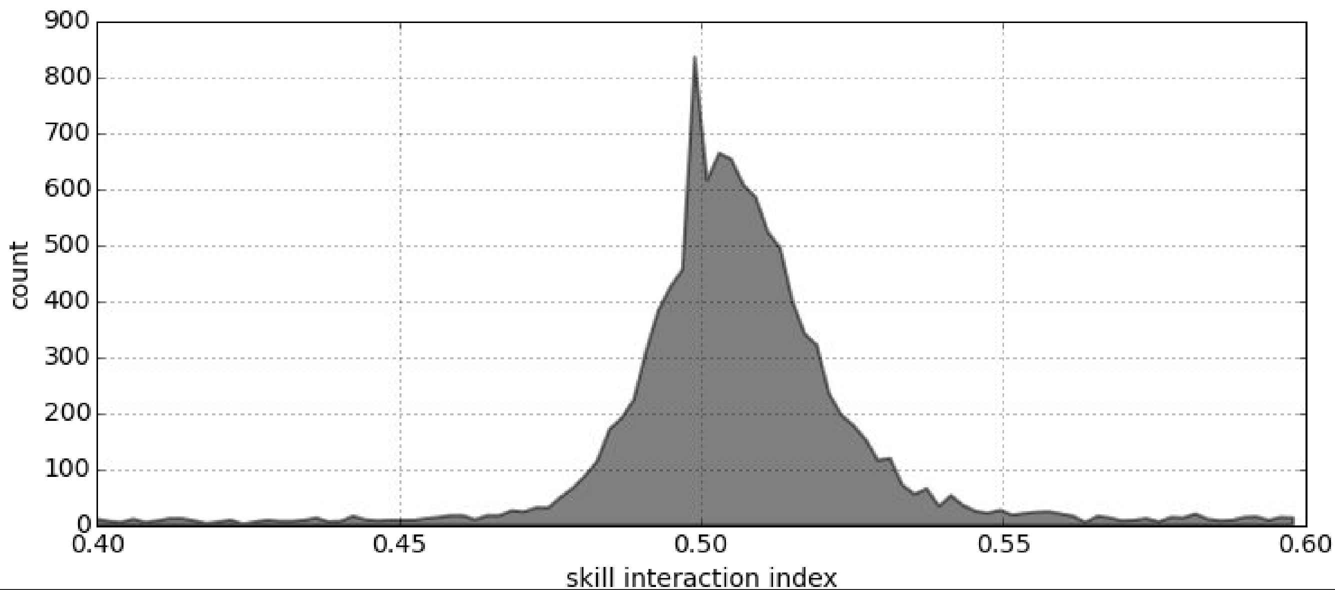


learned item response surfaces

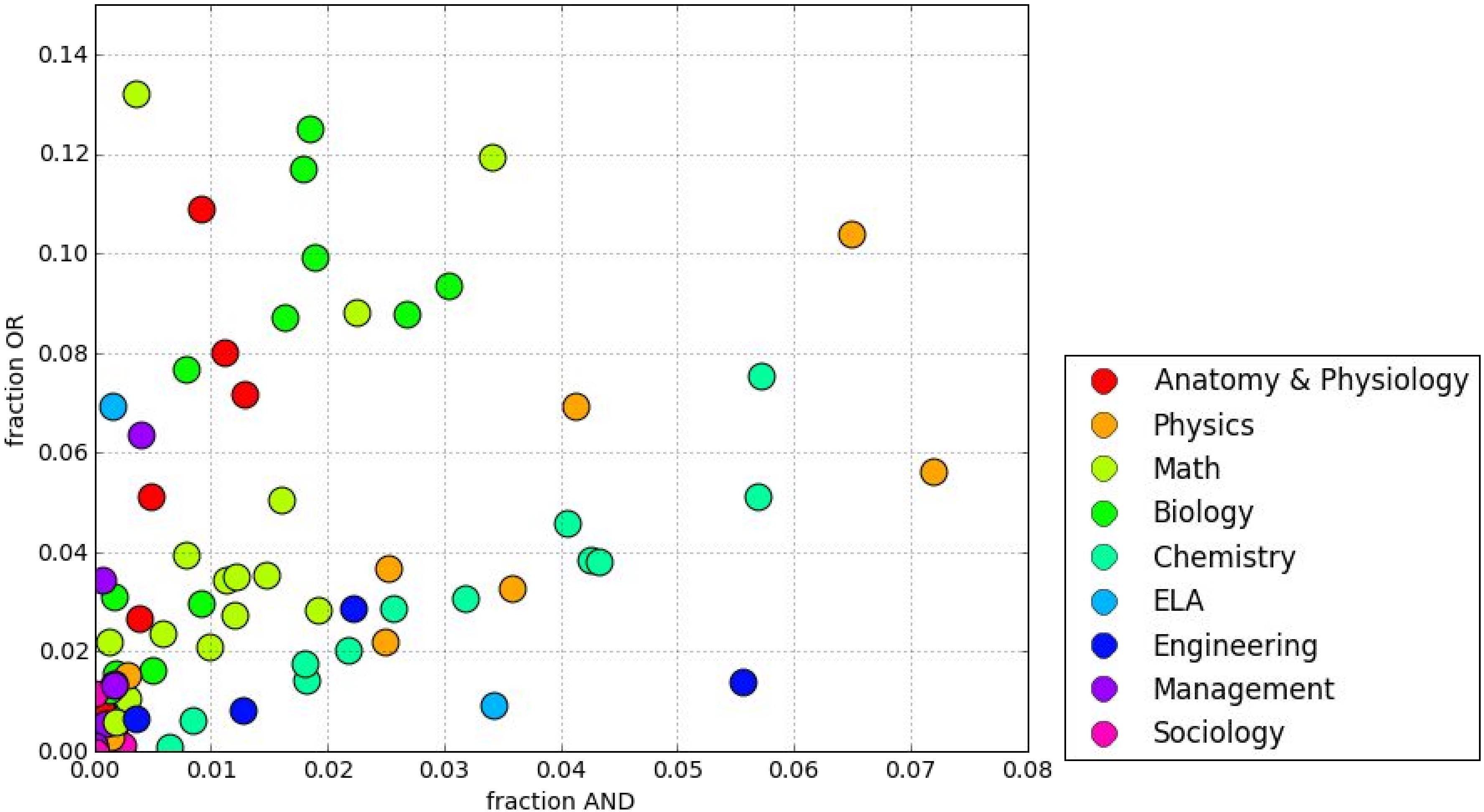


categorizing response surfaces

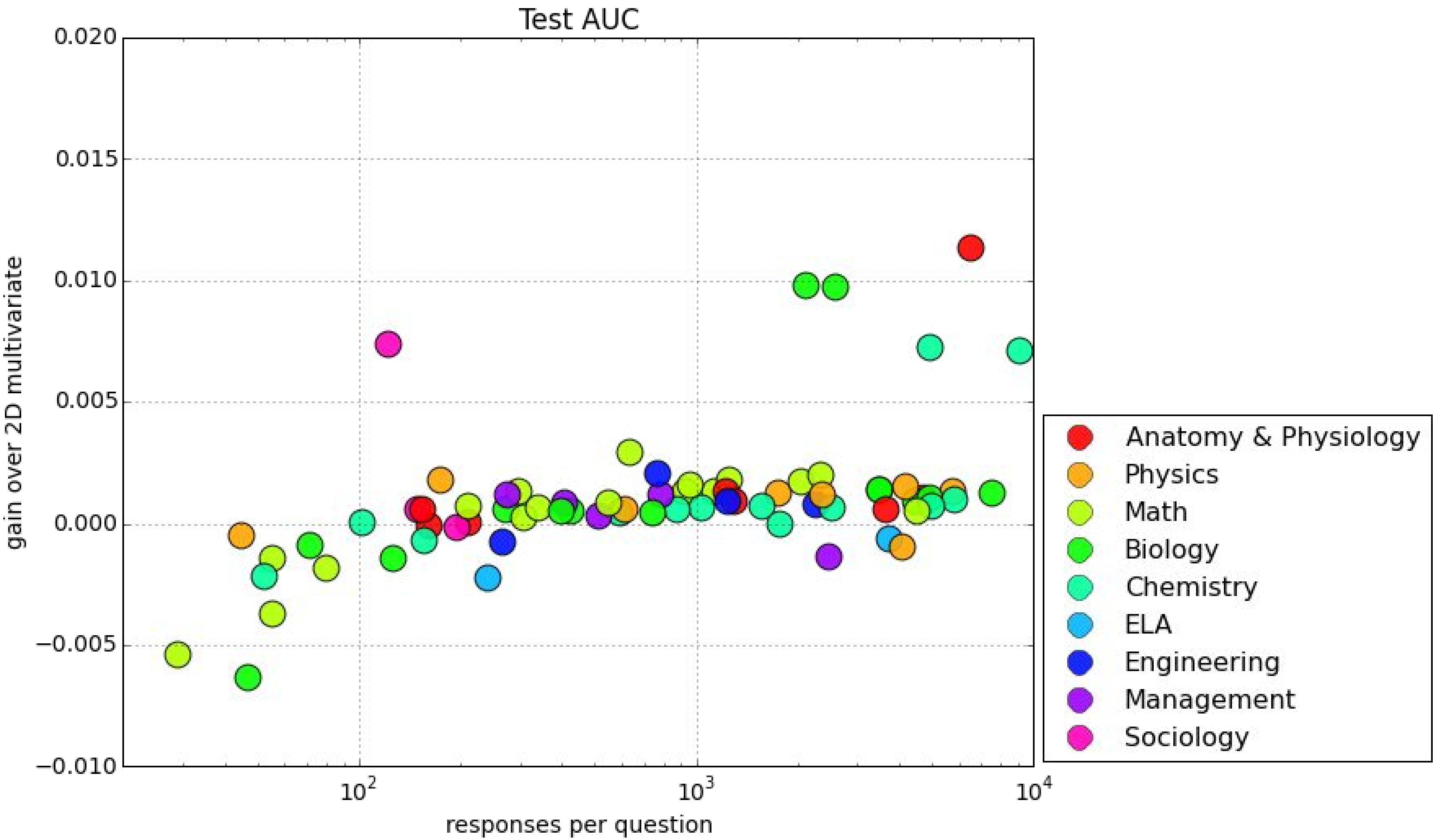




skill-interaction statistics



prediction accuracy



next steps

multivariate IRT

- compare performance to other models (hierarchical items, temporal profs.)
- sparse priors on item-concept mappings yield more interpretable params?
- combine expert and machine-learned concept mappings

multilayer IRT

- deal with convergence issues
- larger neural net item response models
- correlate with item types, domains

Thank You

yan karklin

principal data scientist

Knewton, Inc.

yan@knewton.com

Twitter: @knewton