Week 2 Video 6

Types of Validity

Many types of validity

Generalizability

- Does your model remain predictive when used in a new data set?
- Underlies the cross-validation paradigm that is common in data mining
- Knowing the context the model will be used in drives what kinds of generalization you should study

Generalizability Fail

- Model of boredom is built on data from 3 students
- Model fails when applied to new students

Ecological Validity

- Do your findings apply to real-life situations outside of research settings?

- For example, if you build a detector of student behavior in lab settings, will it work in real classrooms?

Ecological Validity Fail

- Detector of Off-Task Behavior is built based on data from lab study where students use the software one at a time

- Detector is then applied to classroom data

Ecological Validity Subtle Fail

- Model predicting high school dropout is built on data from 300 students, all from middle-class suburban schools
- Model is cross-validated at student level
- Model fails when applied to urban students

Construct Validity

- Does your model actually measure what it was intended to measure?

Construct Validity

- Does your model actually measure what it was intended to measure?
- One interpretation: does your model fit the training data?

Construct Validity

- Does your model actually measure what it was intended to measure?
- One interpretation: does your model fit the training data?
- But is your training data correct?

Construct Validity Fail

- You're trying to detect from disciplinary records which students will end up in alternative school

- But your label of "alternative school" also includes students with cognitive or developmental disabilities sent to a special school

Predictive Validity

- Does your model predict not just the present, but the future as well?

- "It is difficult to make predictions, especially about the future." – Niels Bohr

Substantive Validity

- Do your results matter?
- Are you modeling a construct that matters?

- If you model X, what kind of scientific findings or impacts on practice will this model drive?

- Can be demonstrated by predicting future things that *matter*

Substantive Validity

- For example, we know that boredom correlates strongly with
  - Disengagement
  - Learning Outcomes
  - Standardized Exam Scores
  - Attending College Years Later

Substantive Validity

- By comparison, whether someone prefers visual or verbal learning materials doesn't even seem to predict very reliably whether they learn better from visual or verbal learning materials
  (See lit review in Pashler et al., 2008)

Content Validity

- From testing; does the test cover the full domain it is meant to cover?

- For behavior modeling, an analogy would be, does the model cover the full range of behavior it's intended to?
  - A model of gaming the system that only captured systematic guessing but not hint abuse (cf. Baker et al, 2004; my first model of this)
  - Would have lower content validity than a model which captured both (cf. Baker et al., 2008)

Conclusion Validity

- Are your conclusions justified based on the evidence?

Many Dimensions of Validity

- Important to address them all

End of Week 2

- See you next week