15-780 - GRADUATE ARTIFICIAL INTELLIGENCE
AI AND EDUCATION II

Shayan Doroudi

April 26, 2017

Series on applications of AI to education.

Lecture	Application	AI Topics
4/24/17	Learning	Machine Learning + Search
4/26/17	**Assessment**	**Machine Learning + Mechanism Design**
5/01/17	Instruction	Multi-Armed Bandits

TODAY

- Item Response Theory (IRT)
- Computerized Adaptive Testing (CAT)
- Calibrated Self-Assessment

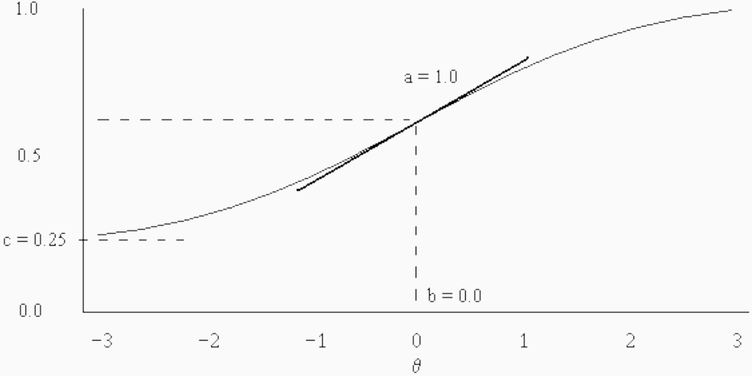
ITEM RESPONSE THEORY (IRT)

- AFM:
  - $\log\left(\frac{p_{ij,T+1}}{1-p_{ij,T+1}}\right) = \theta_i + \sum_k Q_{jk}(\beta_k + \gamma_k T)$
  - $p_{ij,T+1} = \frac{1}{1+\exp(-(\theta_i + \sum_k Q_{jk}(\beta_k + \gamma_k T)))}$
- $p_{ij,T}$: Probability that student i answers question j correctly at opportunity T.
- θ_i: Ability of student i
- β_k: Difficulty of skill k
- γ_k: Learning rate of skill k

- One-Parameter IRT Model (1PL):
  - $\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \theta_i - b_j$
  - $p_{ij} = \frac{1}{1+\exp(-(\theta_i - b_j))}$
- p_{ij}: Probability that student i answers question j correctly.
- θ_i: Ability of student i
- b_j: Difficulty of item j

- Two-Parameter IRT Model (2PL):
    - $\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = a_j(\theta_i - b_j)$
- p_{ij}: Probability that student i answers question j correctly.
- θ_i: Ability of student i
- b_j: Difficulty of item j
- a_j: Discrimination of item j

ITEM RESPONSE THEORY (IRT)



Which of the following is true about the IRT model?

- It is a linear regression model.
- It is a logistic regression model.
- It follows a power law of practice for $P = \left(\frac{p_{ij}}{1-p_{ij}}\right)$.
- It follows an exponential law of practice for $P = \left(\frac{p_{ij}}{1-p_{ij}}\right)$.

ADVANTAGES OF IRT TO CLASSICAL TESTING THEORY

What are some of the advantages of IRT to classical testing theory (add up the score on each item)?

ADVANTAGES OF IRT TO CLASSICAL TESTING THEORY

What are some of the advantages of IRT to classical testing theory (add up the score on each item)?

- Can measure with much more precision.

ADVANTAGES OF IRT TO CLASSICAL TESTING THEORY

What are some of the advantages of IRT to classical testing theory (add up the score on each item)?

- Can measure with much more precision.
- Can obtain a standard error of measurement.

ADVANTAGES OF IRT TO CLASSICAL TESTING THEORY

What are some of the advantages of IRT to classical testing theory (add up the score on each item)?

- Can measure with much more precision.
- Can obtain a standard error of measurement.
- Can give different tests to different students without compromising rankings.

ADVANTAGES OF IRT TO CLASSICAL TESTING THEORY

What are some of the advantages of IRT to classical testing theory (add up the score on each item)?

- Can measure with much more precision.
- Can obtain a standard error of measurement.
- Can give different tests to different students without compromising rankings.
- Computerized Adaptive Testing!

COMPUTERIZED ADAPTIVE TESTING (CAT)

Why might we want a test to be adaptive?

Start with a set of calibrated test items.

Start with a set of calibrated test items.

1. Based on our estimate of a students ability $\hat{\theta}$ choose the item that will give us the most information to get a more precise measure of the student's ability.

Start with a set of calibrated test items.

1. Based on our estimate of a students ability $\hat{\theta}$ choose the item that will give us the most information to get a more precise measure of the student's ability.
2. Student answers question.

Start with a set of calibrated test items.

1. Based on our estimate of a students ability $\hat{\theta}$ choose the item that will give us the most information to get a more precise measure of the student's ability.
2. Student answers question.
3. Update $\hat{\theta}$ using maximum likelihood estimation.

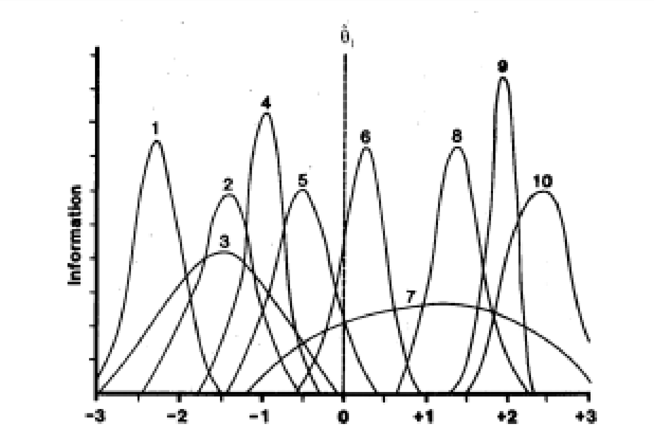
Start with a set of calibrated test items.

1. Based on our estimate of a students ability $\hat{\theta}$ choose the item that will give us the most information to get a more precise measure of the student's ability.
2. Student answers question.
3. Update $\hat{\theta}$ using maximum likelihood estimation.
4. Repeat steps 1-3 until termination.

- Fisher Information: $\mathcal{I}(\theta) = \mathbb{E}\left[(\frac{\partial}{\partial\theta}\log p(X;\theta))^2|\theta\right]$

- Fisher Information: $\mathcal{I}(\theta) = \mathbb{E}\left[(\frac{\partial}{\partial\theta}\log p(X;\theta))^2|\theta\right]$
- 1PL Information: $\mathcal{I}_j(\theta_i) = p_{ij}(1 - p_{ij})$

- Fisher Information: $\mathcal{I}(\theta) = \mathbb{E}\left[(\frac{\partial}{\partial\theta} \log p(X;\theta))^2 | \theta\right]$
- 1PL Information: $\mathcal{I}_j(\theta_i) = p_{ij}(1 - p_{ij})$
- 2PL Information: $\mathcal{I}_j(\theta_i) = a_i^2 p_{ij}(1 - p_{ij})$



Weiss, D. J. (2004). Computerized adaptive testing for effective and efficient measurement in counseling and education. Measurement and Evaluation in Counseling and Development, 37(2), 70.

calibrated self-assessment

labutov, i., & studer, c. calibrated self-assessment.

educational data mining, 2016.

How do we grade free-form questions in large courses (e.g., MOOCs)?

SELF-ASSESSMENT

Ask student how likely they are to have answered a question correctly!

Want a strategy proof mechanism to elicit student correctness.

Want a strategy proof mechanism to elicit student correctness.

Can use quadratic scoring rule:

$$S_{ij} = \begin{cases} c_{ij} & \text{if item } j \text{ correct} \\ -\frac{1}{2}c_{ij}^2 & \text{if item } j \text{ incorrect} \end{cases}$$

where c_{ij} is a score proposed by student i on item j.

Student wants to maximize:

$$\mathbb{E}\left[S_{ij}\right] = c_{ij}p_{ij} - \frac{1}{2}c_{ij}^2(1 - p_{ij})$$

Student wants to maximize:

$$\mathbb{E}\left[S_{ij}\right] = c_{ij}p_{ij} - \frac{1}{2}c_{ij}^2(1 - p_{ij})$$

Maximized when $c_{ij} = \frac{p_{ij}}{1-p_{ij}}$.

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \hat{\theta}_i - \hat{b}_j$$

- p_{ij}: Student i's estimated probability that they answer question j correctly.
- $\hat{\theta}_i$: Student i's estimate of their own ability
- $\hat{b}_j$: Student i's estimate of the difficulty of item j

$$\log\left(\frac{p_{ij}}{1-p_{ij}}\right) = \hat{\theta}_i - \hat{b}_j$$

- p_{ij}: Student *i*'s estimated probability that they answer question *j* correctly.
- $\hat{\theta}_i$: Student *i*'s estimate of their own ability
- $\hat{b}_j$: Student *i*'s estimate of the difficulty of item *j*
- Assume $\hat{\theta}_i - \hat{b}_j \sim \mathcal{N}(\theta_i - b_j, \sigma^2)$

$$\log(c_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \theta_i - b_j + \epsilon$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

$$\log(c_{ij}) = \log\left(\frac{p_{ij}}{1 - p_{ij}}\right) = \theta_i - b_j + \epsilon$$

$$\epsilon \sim \mathcal{N}(0, \sigma^2)$$

Can be estimated using linear regression!

Is the mechanism fair?

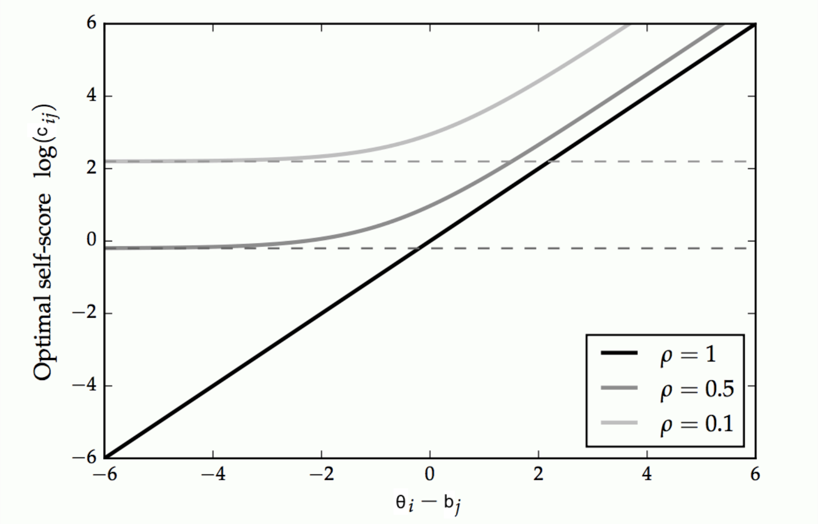
$$\mathbb{E}\left[S_{ij}\right] = c_{ij}p_{ij} - \frac{1}{2}c_{ij}^2(1 - p_{ij})$$

- Yes, it is fair.
- No, it will give inflate scores of higher ability students and deflate scores of lower ability students.
- No, it will deflate scores of higher ability students and inflate scores of lower ability students.

What happens when we don't actually grade student answers?

If each item is graded with probability ρ:

$$\mathbb{E}\left[S_{ij}\right] = \rho(c_{ij}p_{ij} - \frac{1}{2}c_{ij}^2(1 - p_{ij})) + (1 - \rho)c_{ij}$$



Labutov, I., & Studer, C. Calibrated Self-Assessment. Educational Data Mining, 2016.

- IRT allows for more precise measuring of student abilities.

SUMMARY

- IRT allows for more precise measuring of student abilities.
- Is used for computerized adaptive testing.

- IRT allows for more precise measuring of student abilities.
- Is used for computerized adaptive testing.
- Can combine mechanism design with IRT to elicit scores from students.