

Chapter 8: Learner Modeling (CS-411)

Tanja Käser

November 11, 2025

Last Week – Foundations of Adaptive Instruction

- *What* it is (definition)
- *How* it works (learner modeling loop)
- *How* it has evolved (history)
- *Where* it applies (dimensions of adaptivity)
- *Why* it works (pedagogical foundations)

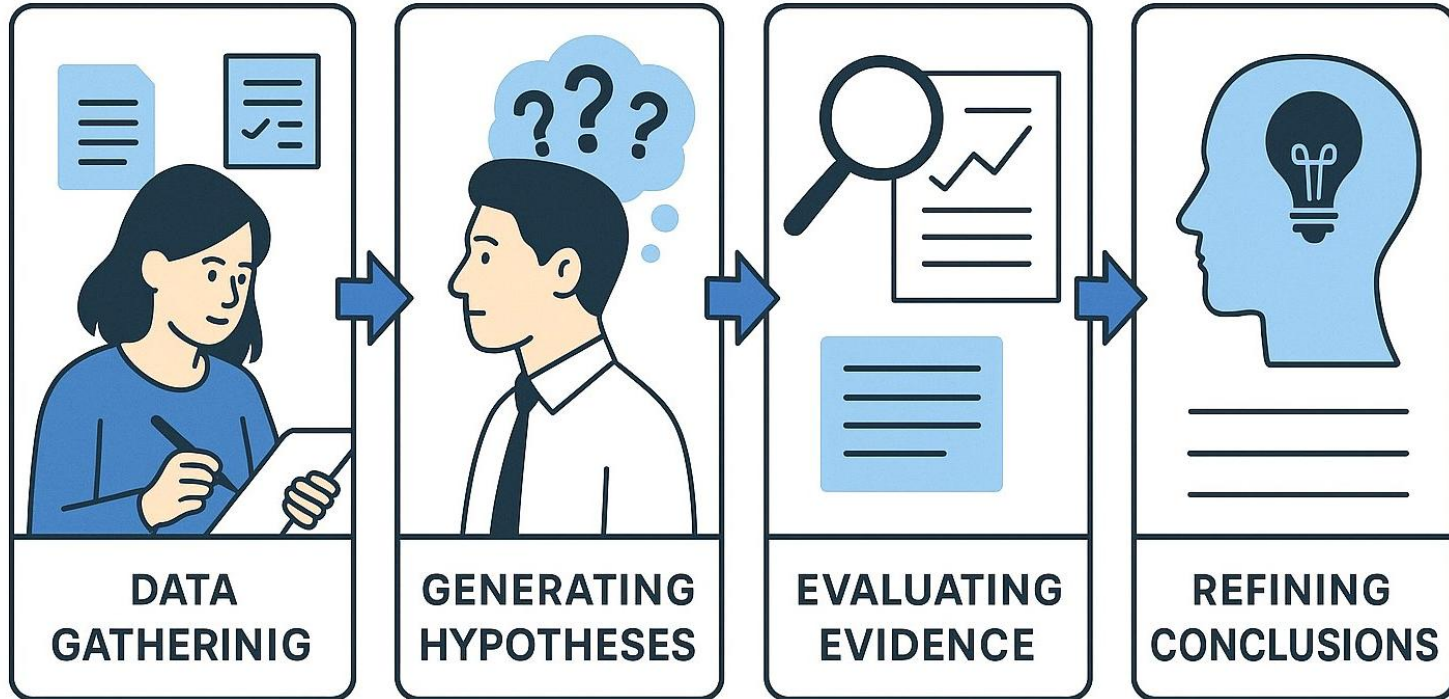
What about adaptive instruction in PS-I?

Explore

Consolidate

Apply

Diagnostic Reasoning in a Pharmacy



PharmaSim – Virtual Pharmacy



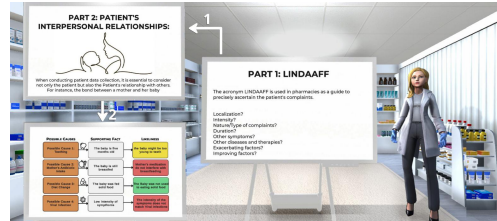
PharmaSim – PS-I Sequence

Explore



Father of a baby with diarrhea

Consolidate

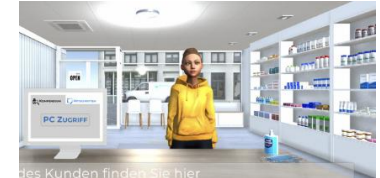


Adaptive Instruction

Apply



Another father of a baby with diarrhea



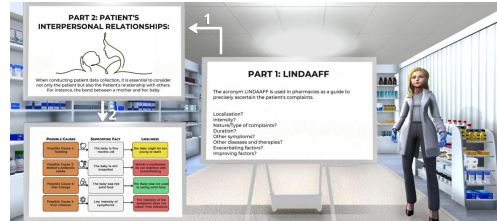
Mother with breastfeeding issues

PharmaSim – PS-I Sequence

Explore



Consolidate



Apply



Father of a baby with diarrhea

Adaptive Instruction

Another father of a baby with diarrhea

Mother with breastfeeding issues

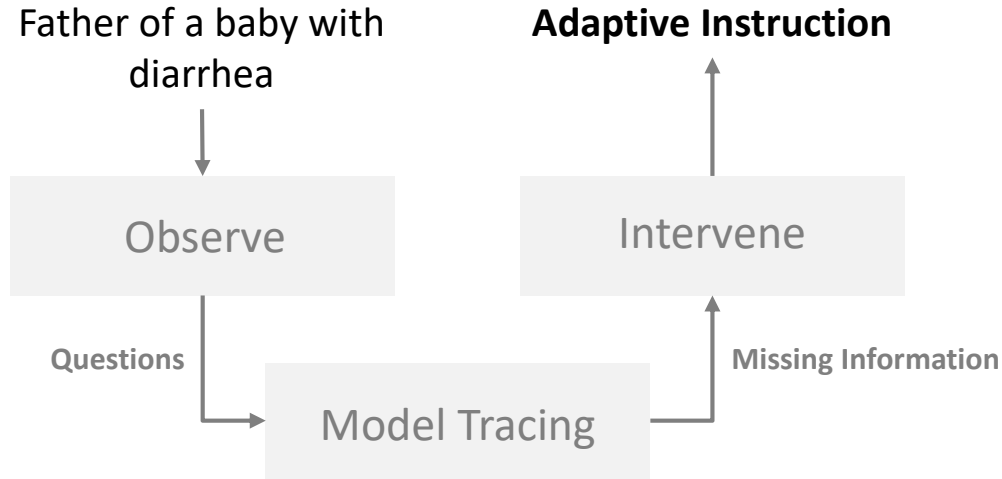
Observe

Intervene

Questions

Missing Information

Model Tracing



Adaptive Instruction

PART 2: PATIENT'S INTERPERSONAL RELATIONSHIPS:



When conducting patient data collection, it is essential to consider not only the patient but also the Patient's relationship with others. For instance, the bond between a mother and her baby

PART 1: LINDAAFF

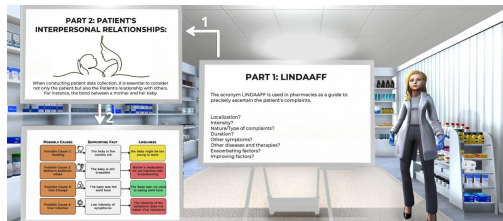
The acronym LINDAAFF is used in pharmacies as a guide to precisely ascertain the patient's complaints.

Localization?
Intensity?
Nature/Type of complaints?
Duration?
Other symptoms?
Other diseases and therapies?
Exacerbating factors?
Improving factors?

POSSIBLE CAUSES	SUPPORTING FACT	LIKELINESS
Possible Cause 1: Teething	The baby is five months old	the baby might be too young to teath
Possible Cause 2: Mother's Antibiotic intake	The baby is still breastfed	Mother's medication do not interfere with breastfeeding
Possible Cause 3: Diet Change	The baby was fed solid food	The Baby was not used to eating solid food
Possible Cause 4: Viral Infection	Low intensity of symphoms	The intensity of the symptoms does not match Viral infections

Control – I-PS sequence

Instruction



Worked Example

Problem-Solving

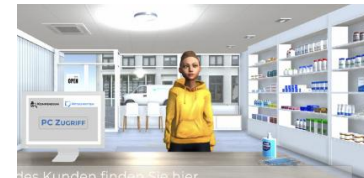


Father of a baby with diarrhea

Transfer



Another father of a baby with diarrhea

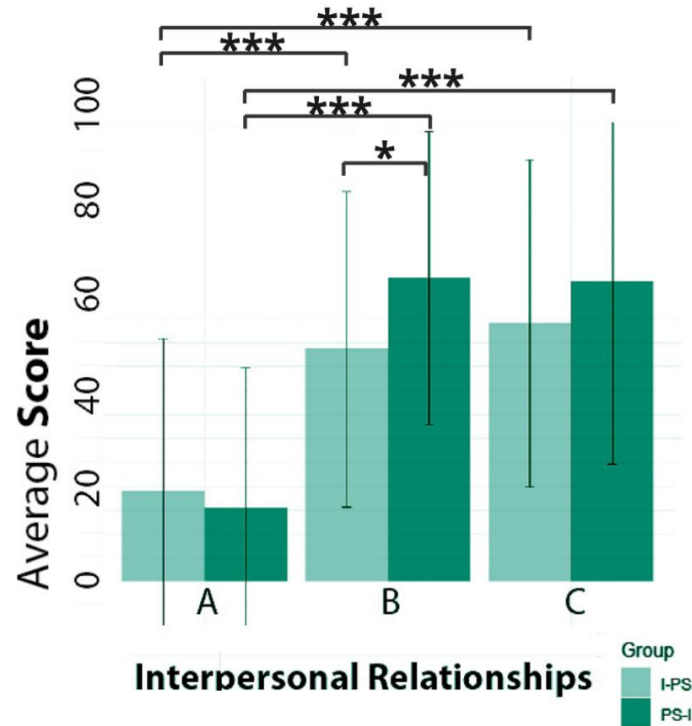


Mother with breastfeeding issues

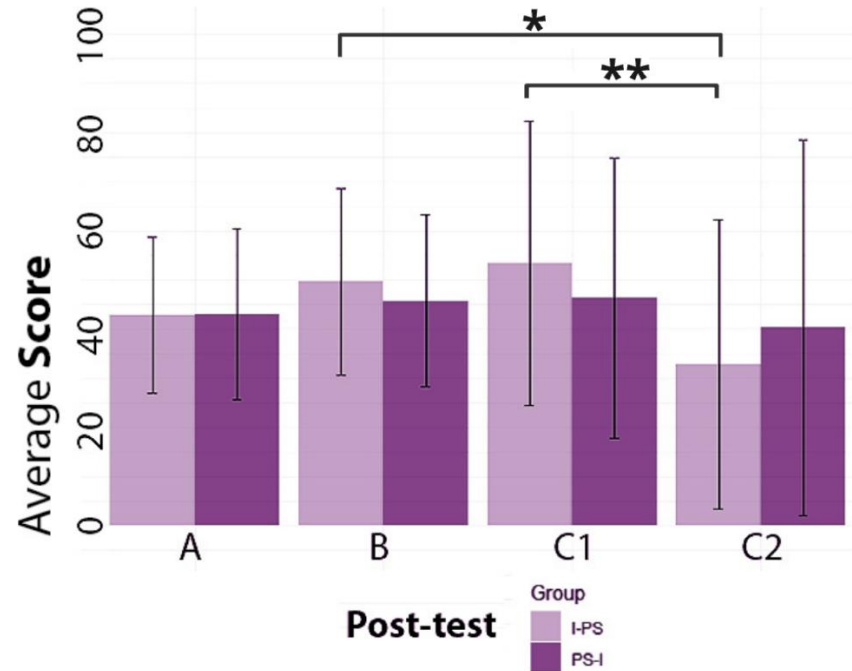
Evaluation

- Measure application of strategies across scenarios:
 - Checklist
 - Interpersonal Relationship
 - Systematically assessing causes

PS-I apply interpersonal relationships more



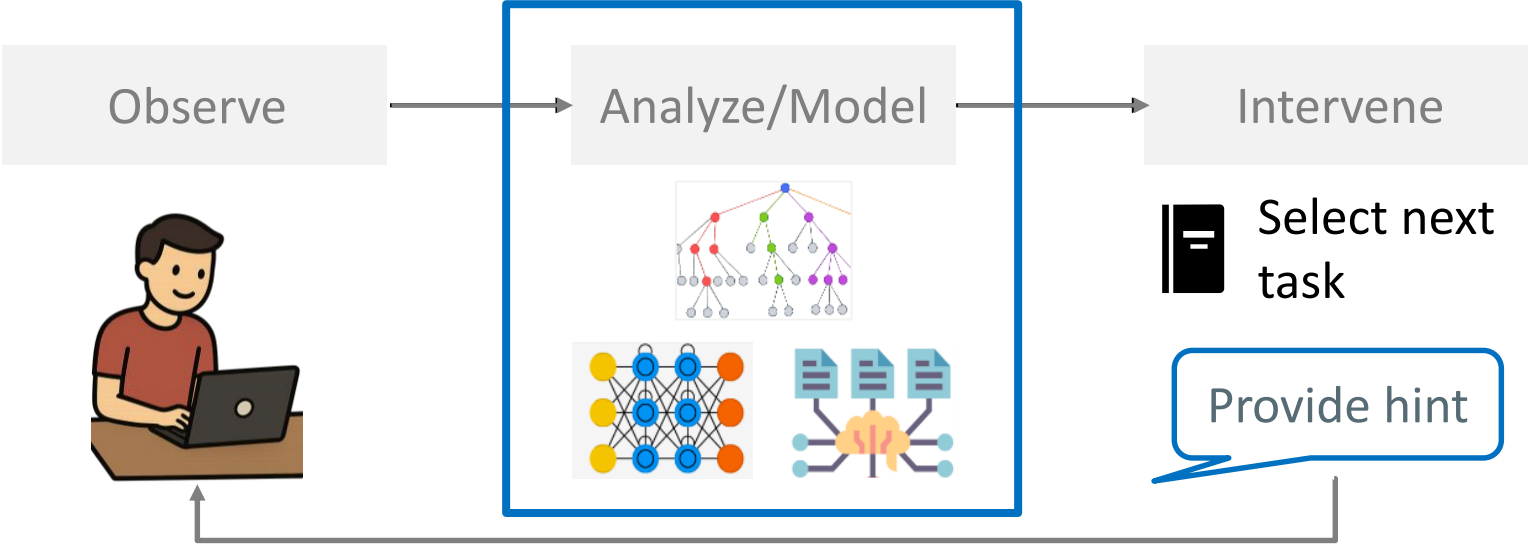
PS-I transfers data interpretation strategy better



Personalized & Generative Approaches in Digital Education

- ① Foundations of Adaptive Instruction
 - Principles and history of adaptive learning systems
- ② **Learner Modeling**
 - **How we represent the learner to drive adaptive instruction**
- ③ Generative AI in Education
 - Understanding generative models and their implications for teaching and learning
- ④ Designing AI-Enhanced Learning Experiences
 - Bridging adaptive instruction and generative AI

How it works – The learner modeling loop



Tracing Student Knowledge

- Is the student learning?
 - Measure what the student *knows* at a specific time t
 - More specifically: knowledge of the student about relevant knowledge components (skills)



Task:	$50 - 23 = ?$	$75 - 12 = ?$	$38 - 14 = ?$
Answer:	27	61	24

Tracing Knowledge – why is it useful?

- Is the student learning?
 - Measure what the student *knows* at a specific time t
 - More specifically: knowledge of the student about relevant knowledge components (skills)
- ➔ Choose the next appropriate activity
- ➔ Know which activities support learning

Tracing Knowledge – why is it useful?

- Is the student learning?
 - Measure what the student *knows* at a specific time t
 - More specifically: knowledge of the student about relevant knowledge components (skills)

➔ Choose the next appropriate activity

➔ Know which activities support learning

What are we measuring?



Task: $50 - 23 = ?$ $75 - 12 = ?$ $38 - 14 = ?$

Answer: 27 61 24

1

0

1

Binary observations of student answers



Subtraction 0-100

1

2

...

n

0

0

1

0

1

1

Predicting future performance



Subtraction 0-100

1

2

...

n

n+1

0

0

1

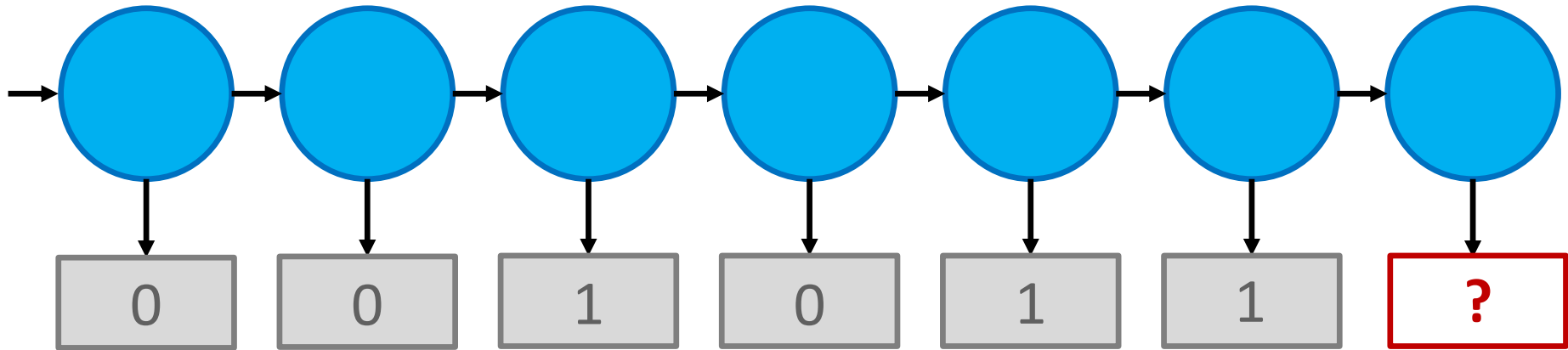
0

1

1

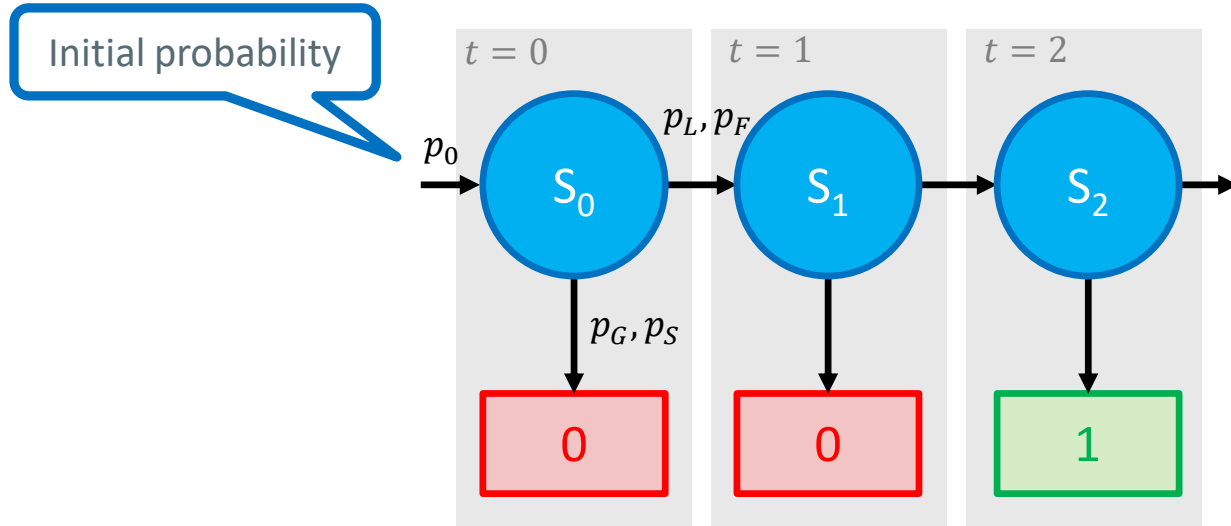
?

Bayesian Knowledge Tracing (BKT)

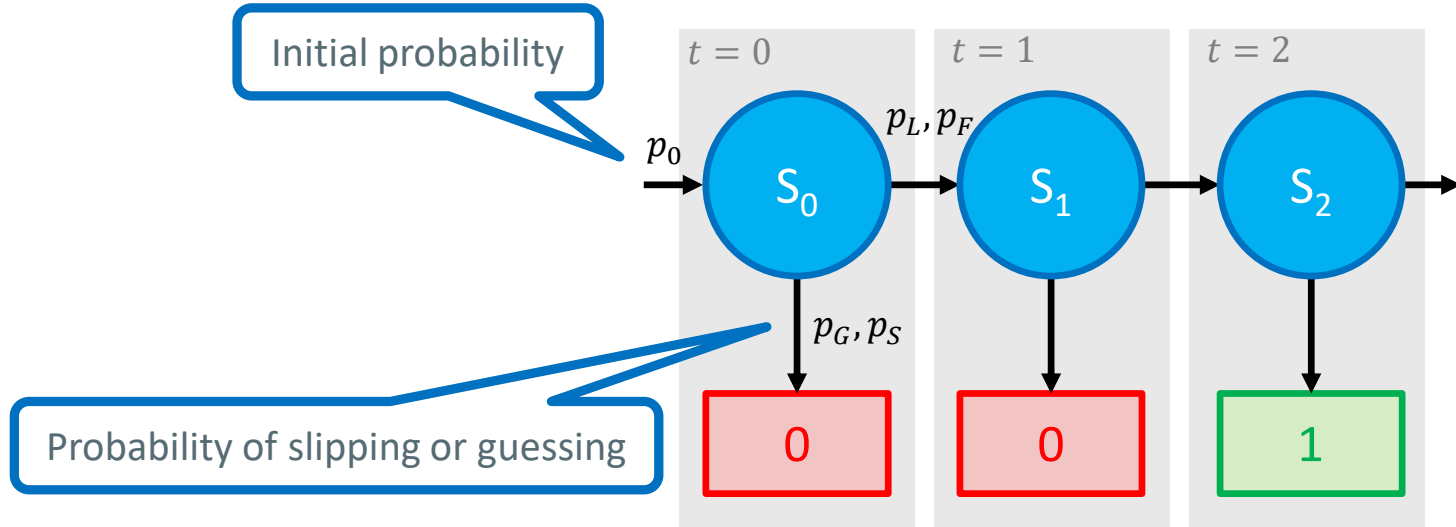


[Corbett & Anderson, UMUAI 1994]

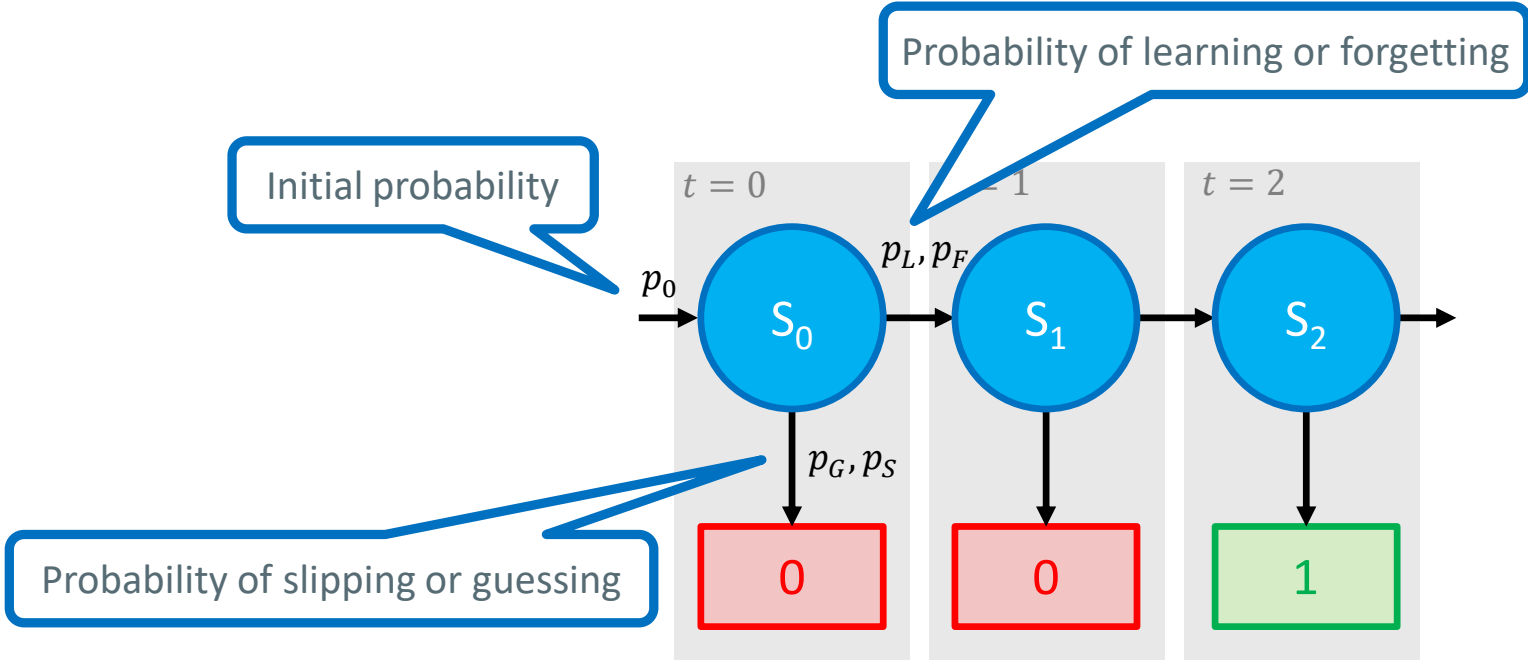
BKT parameters are interpretable



BKT parameters are interpretable

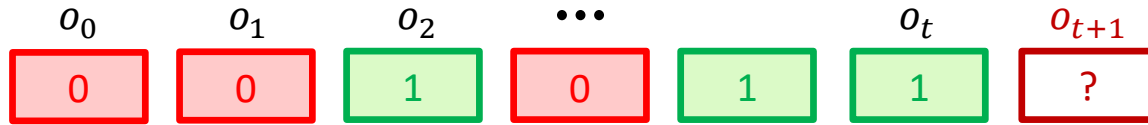


BKT parameters are interpretable



Two tasks need to be solved in practice

- Given a model with parameters $\theta = \{p_0, p_L, p_F, p_S, p_G\}$ and a sequence of observations $\mathbf{o} = [o_0, \dots, o_t]$ from a student s , predict o_{t+1}



Inference

Inference in BKT models

$$\mathbf{o}_{t-1} = [o_0, \dots, o_{t-1}]$$

Equations for time $t = 0$:

Belief about latent state before observation

$$p(s_0 = 1) = p_0$$

Predicted observation at time t

$$p(o_0 = 1) = (1 - p_S) \cdot p_0 + p_G \cdot (1 - p_0)$$

$$p(o_0 = 0) = p_S \cdot p_0 + (1 - p_G) \cdot (1 - p_0)$$

Posterior: belief about latent state after observation

$$p_{s_0|1} = \frac{(1 - p_S) \cdot p_0}{(1 - p_S) \cdot p_0 + p_G \cdot (1 - p_0)} \quad \left. \vphantom{p_{s_0|1}} \right\} p_{s_0|o_0}$$

$$p_{s_0|0} = \frac{p_S \cdot p_0}{p_S \cdot p_0 + (1 - p_G) \cdot (1 - p_0)}$$

Equations for time steps $t = 1, \dots, T$:

$$p_{s_t|o_{t-1}} = (1 - p_F) \cdot p_{s_{t-1}|o_{t-1}} + p_L \cdot (1 - p_{s_{t-1}|o_{t-1}})$$

$$p(o_t = 1|\mathbf{o}_{t-1}) = (1 - p_S) \cdot p_{s_t|o_{t-1}} + p_G \cdot (1 - p_{s_t|o_{t-1}})$$

$$p(o_t = 0|\mathbf{o}_{t-1}) = p_S \cdot p_{s_t|o_{t-1}} + (1 - p_G) \cdot (1 - p_{s_t|o_{t-1}})$$

$$p_{s_t|1,o_{t-1}} = \frac{(1 - p_S) \cdot p_{s_t|o_{t-1}}}{(1 - p_S) \cdot p_{s_t|o_{t-1}} + p_G \cdot (1 - p_{s_t|o_{t-1}})} \quad \left. \vphantom{p_{s_t|1,o_{t-1}}} \right\} p_{s_t|o_t}$$

$$p_{s_t|0,o_{t-1}} = \frac{p_S \cdot p_{s_t|o_{t-1}}}{p_S \cdot p_{s_t|o_{t-1}} + (1 - p_G) \cdot (1 - p_{s_t|o_{t-1}})}$$

Making predictions using a BKT model

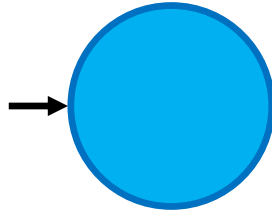
$$p_0 = 0.5$$

$$p_S = 0.2$$

$$p_G = 0.3$$

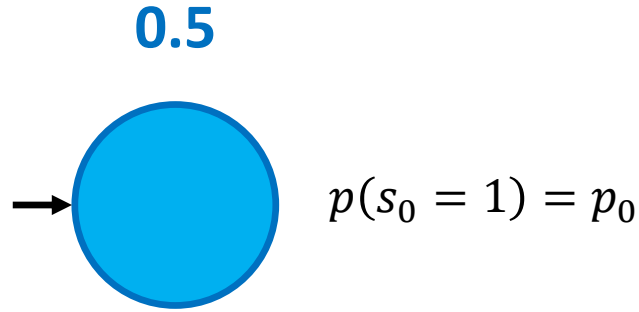
$$p_L = 0.4$$

$$p_F = 0.0$$



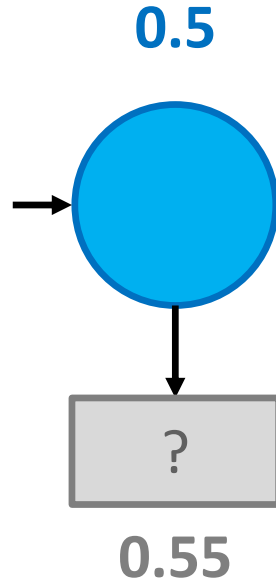
Making predictions using a BKT model

$p_0 = 0.5$
$p_S = 0.2$
$p_G = 0.3$
$p_L = 0.4$
$p_F = 0.0$



Making predictions using a BKT model

$p_0 = 0.5$
$p_S = 0.2$
$p_G = 0.3$
$p_L = 0.4$
$p_F = 0.0$

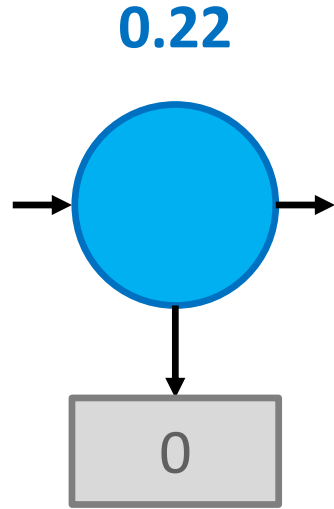


$$p(s_0 = 1) = p_0$$

$$p(o_0 = 1) = (1 - p_S) \cdot p_0 + p_G \cdot (1 - p_0)$$

Making predictions using a BKT model

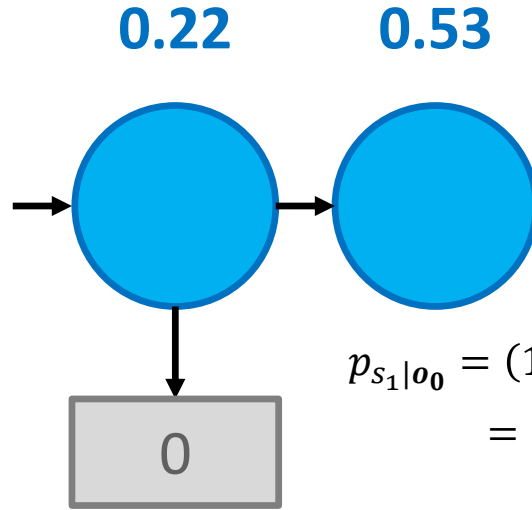
$p_0 = 0.5$
$p_S = 0.2$
$p_G = 0.3$
$p_L = 0.4$
$p_F = 0.0$



$$p_{s_0|0} = \frac{p_S \cdot p_0}{1 - p(o_0 = 1)} = \frac{0.2 \cdot 0.5}{0.45}$$

Making predictions using a BKT model

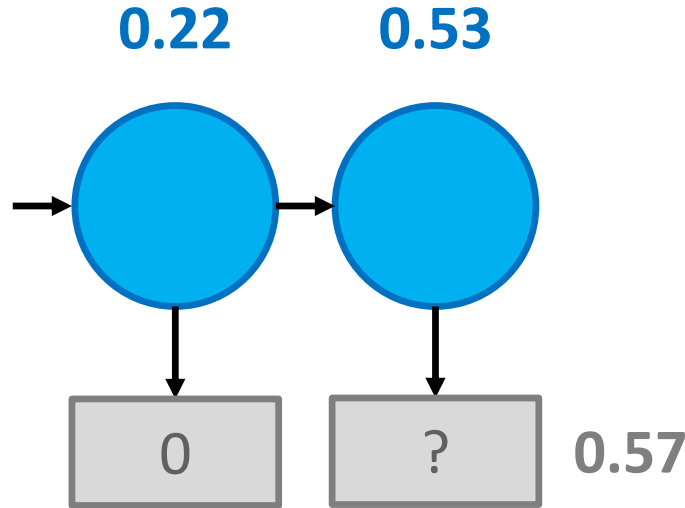
$$\begin{aligned} p_0 &= 0.5 \\ p_S &= 0.2 \\ p_G &= 0.3 \\ p_L &= 0.4 \\ p_F &= 0.0 \end{aligned}$$



$$\begin{aligned} p_{S_1|o_0} &= (1 - p_F) \cdot p_{S_1|o_0} + p_L \cdot (1 - p_{S_1|o_0}) \\ &= 0.22 + 0.4 \cdot 0.78 \end{aligned}$$

Making predictions using a BKT model

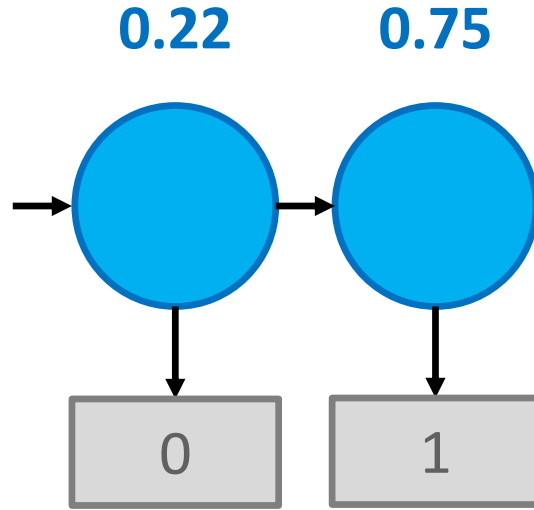
$$\begin{aligned} p_0 &= 0.5 \\ p_S &= 0.2 \\ p_G &= 0.3 \\ p_L &= 0.4 \\ p_F &= 0.0 \end{aligned}$$



$$\begin{aligned} p(o_1 = 1 | \mathbf{o}_0) &= (1 - p_S) \cdot p_{s_1 | \mathbf{o}_0} + p_G \cdot (1 - p_{s_1 | \mathbf{o}_0}) \\ &= 0.8 \cdot 0.53 + 0.3 \cdot 0.47 \end{aligned}$$

Making predictions using a BKT model

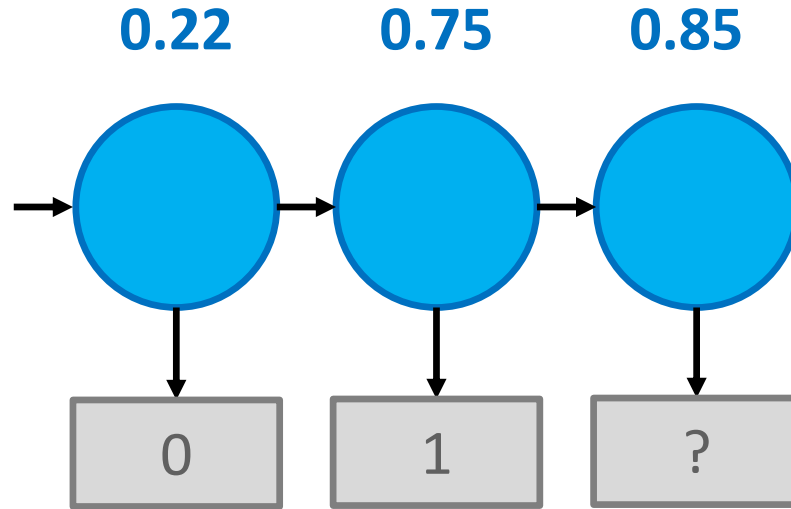
$$\begin{aligned} p_0 &= 0.5 \\ p_S &= 0.2 \\ p_G &= 0.3 \\ p_L &= 0.4 \\ p_F &= 0.0 \end{aligned}$$



$$\begin{aligned} p_{S_1|1,o_0} &= \frac{(1 - p_S) \cdot p_{S_1|o_0}}{(1 - p_S) \cdot p_{S_1|o_0} + p_G \cdot (1 - p_{S_1|o_0})} \\ &= \frac{0.8 \cdot 53}{0.57} \end{aligned}$$

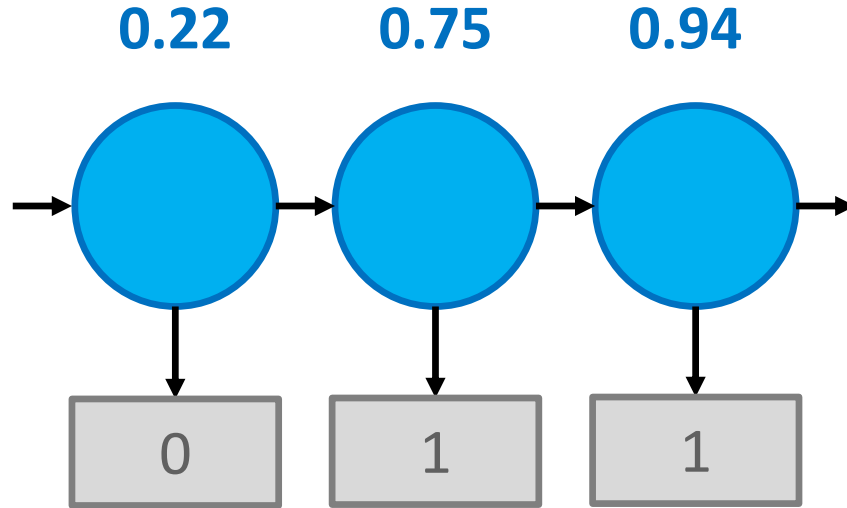
Making predictions using a BKT model

$$\begin{aligned} p_0 &= 0.5 \\ p_S &= 0.2 \\ p_G &= 0.3 \\ p_L &= 0.4 \\ p_F &= 0.0 \end{aligned}$$



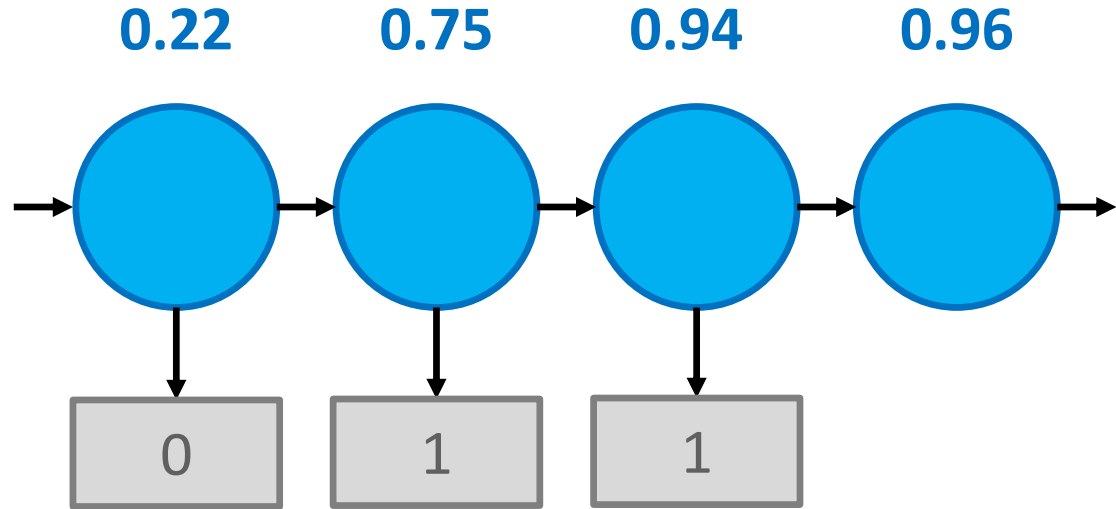
Making predictions using a BKT model

$p_0 = 0.5$
$p_S = 0.2$
$p_G = 0.3$
$p_L = 0.4$
$p_F = 0.0$



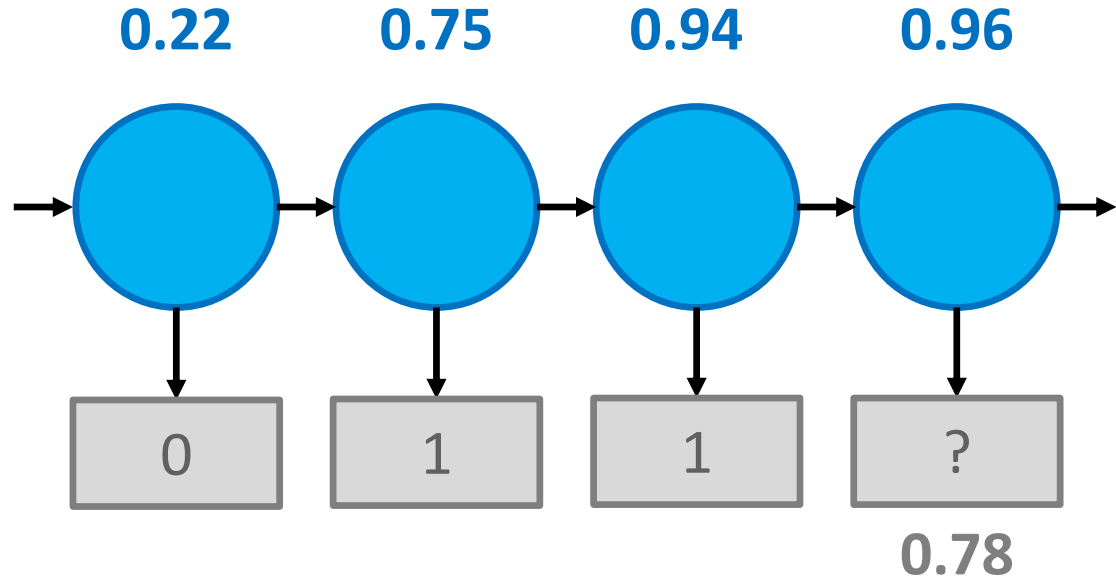
Making predictions using a BKT model

$$\begin{aligned} p_0 &= 0.5 \\ p_S &= 0.2 \\ p_G &= 0.3 \\ p_L &= 0.4 \\ p_F &= 0.0 \end{aligned}$$

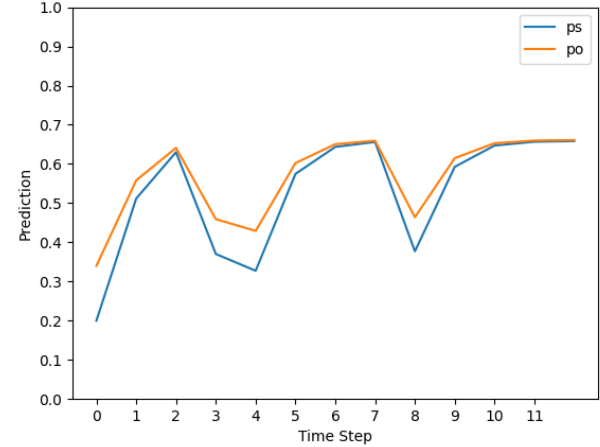
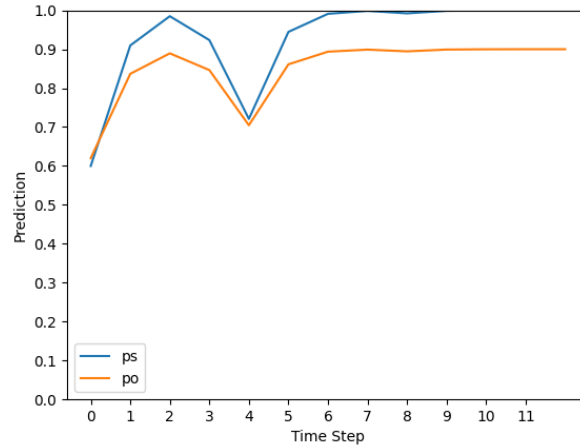
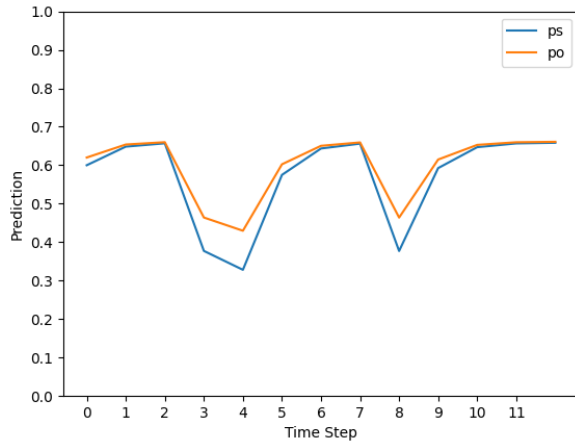


Making predictions using a BKT model

$$\begin{aligned} p_0 &= 0.5 \\ p_S &= 0.2 \\ p_G &= 0.3 \\ p_L &= 0.4 \\ p_F &= 0.0 \end{aligned}$$



Which parameter was changed?



$$p_O = 0.6 \quad p_S = 0.1$$

$$p_G = 0.2 \quad p_L = 0.3$$

$$p_F = 0.3$$

$$o = [1,1,0,0,1,1,1,0,0,1,1]$$

Check your hypotheses

Go to:

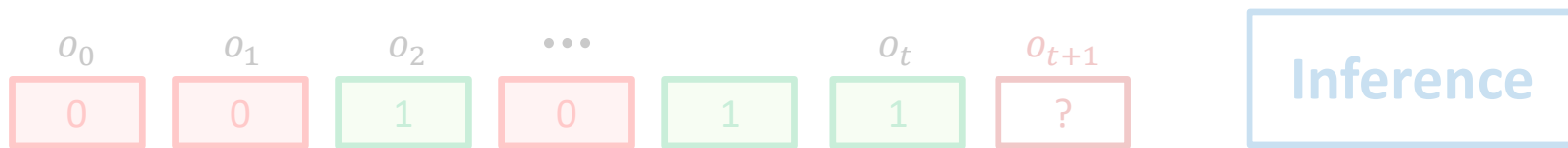
noto.epfl.ch

Clone the GitHub repository:

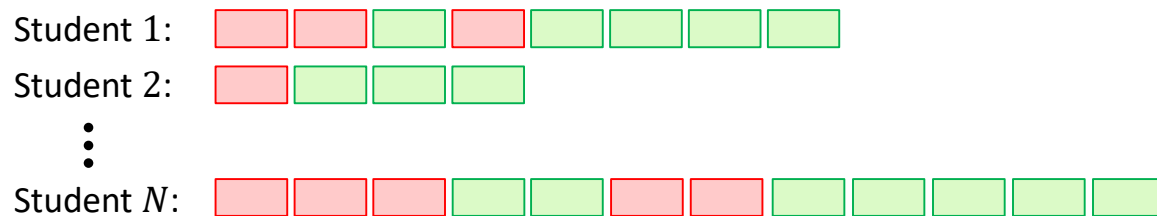
https://github.com/epfl-ml4ed/de_2025

Two tasks need to be solved in practice

- Given a model with parameters $\theta = \{p_0, p_L, p_F, p_S, p_G\}$ and a sequence of observations $\mathbf{o} = [o_0, \dots, o_t]$ from a student s , predict o_{t+1}



- Given sequences of observations $\mathbf{o} = [o_0, \dots, o_T]$ of N students, learn the parameters $\theta = \{p_0, p_L, p_F, p_S, p_G\}$ that maximize the likelihood of the observed data



BKT Assumptions

- Knowledge can be divided into different skills
- Definition of skills is accurate/detailed enough
- Each task corresponds to a single skill (original)
- There is **no** connection between the skills
- Mastery can be achieved through practice
- There is no forgetting: $p_F = 0$ (original)

ASSITments – ITS for math learning

- ASSISTments is a free tool for assigning and assessing math problems and homework
- All math problems (tasks/items) are associated to a specific skill/knowledge component
- 4,217 middle-school students
- 525,534 observations

Your Turn – BKT Training & Inference

- In the student notebook, you have:
 - A trained BKT model for six selected skills
 - A data frame containing the predictions of the BKT model for each observation in the test set
- Your tasks:
 - 1) Plot the BKT predictions for a specific student and skill. What do you observe? Does the model seem to describe the data well?
 - 2) Compute and plot the RMSE per skills. Interpret the results.

Summary – Knowledge Tracing

- Bayesian Knowledge Tracing and Learning Curves
- There are lots of other modeling approaches
 - Factor-based models (AFM & PFA)
 - Deep learning models (DKT)