Knowledge Tracing - Continued

Machine Learning for Behavioral Data

April 4, 2022

Today's Topic

Week	Lecture/Lab
1	Introduction
2	Data Exploration
3	Regression
4	Classification
5	Model Evaluation
6	Knowledge Tracing
7	Knowledge Tracing
8	Time Series Prediction

Supervised learning on time series:
- **Probabilistic graphical models**
- **Neural networks: LSTM, GRU, etc.**

Getting ready for today's lecture…

- **If not done yet**: clone the repository containing the Jupyter notebook and data for today's lecture into your Noto workspace
- Set up the Tensorflow kernel in Noto (follow the instructions in the student notebook)
- SpeakUp room for today's lecture:

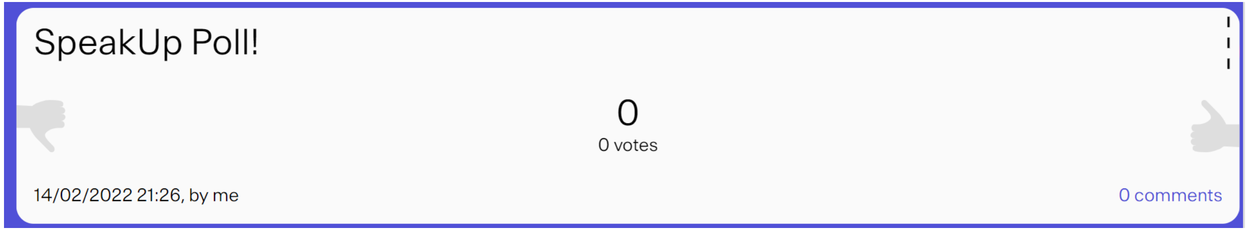
https://go.epfl.ch/mlbd-lecture

Short quiz about the past…

[KT] BKT does account for students guessing the correct answer.

a) True

b) False

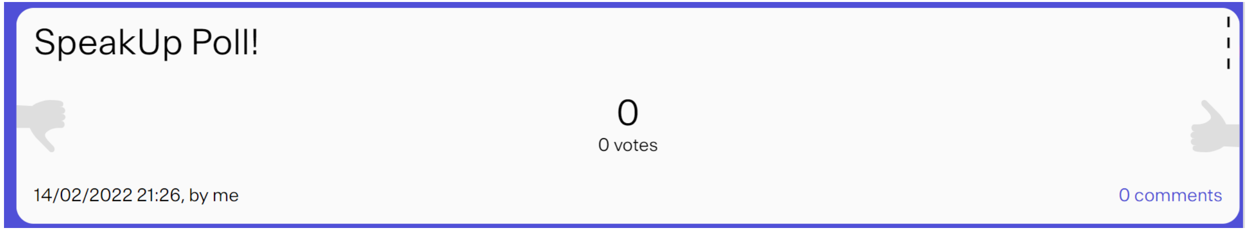


Short quiz about the past...

[KT] BKT can represent the relationships between different skills.

a) True

b) False

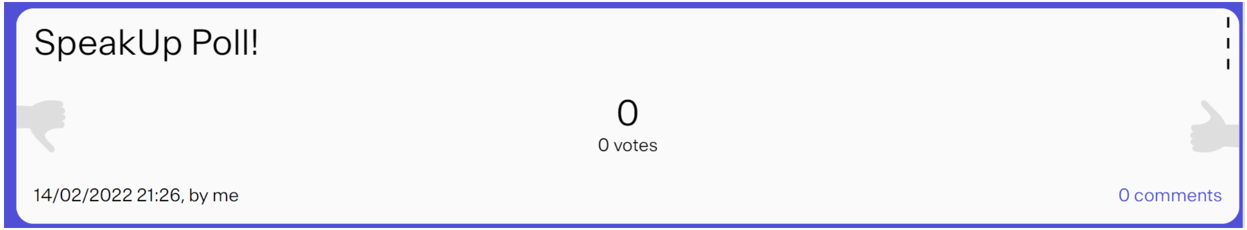


https://go.epfl.ch/mlbd-lecture

Short quiz about the past…

[Mixed Models] Mixed-effect models are useful when the samples in the data set are uncorrelated.

a) True
b) False



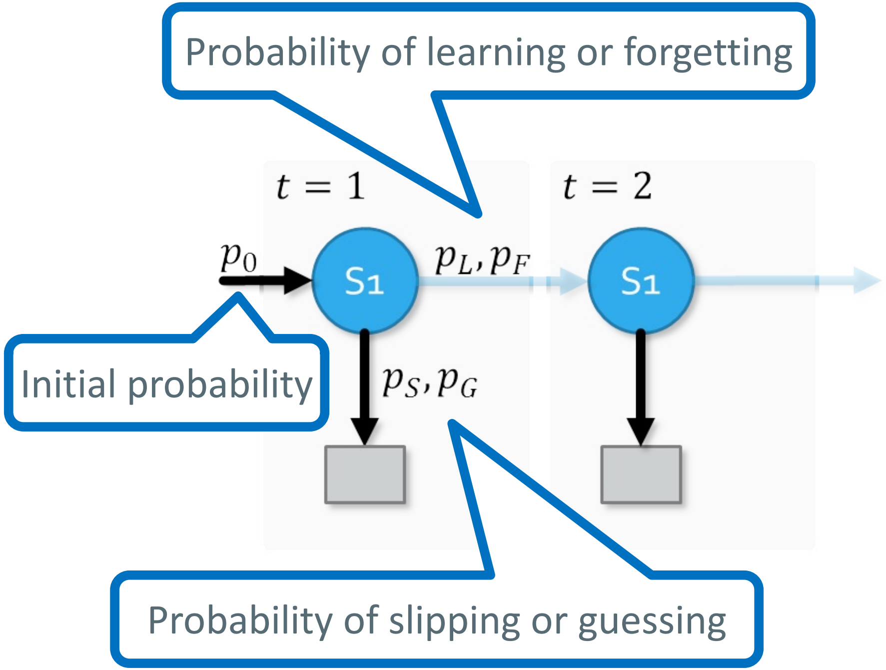
Today's Topic

Week	Lecture/Lab
1	Introduction
2	Data Exploration
3	Regression
4	Classification
5	Model Evaluation
6	Knowledge Tracing
7	**Knowledge Tracing**
8	Time Series Prediction

Supervised learning on time series:

- **Probabilistic graphical models**
- **Neural networks: LSTM, GRU, etc.**

Last Week: Bayesian Knowledge Tracing



- Predict $p(o_{i_{s1},t} | o_0, \dots, o_{t-1})$, the probability that the student will solve task i_{s1} correctly at time step t
- Predict $p(s_{1,t} | o_0, \dots, o_{t-1})$, the probability that the student has mastered skill s_1 at time step t

Assumptions behind BKT

- 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)

Today

- **Learning Curves**
- Alternative Models for Knowledge Tracing:
  - AFM/PFA
  - Deep Knowledge Tracing

Today's Use Case

- 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
- 4,151 middle-school students
- 525,534 observations

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)

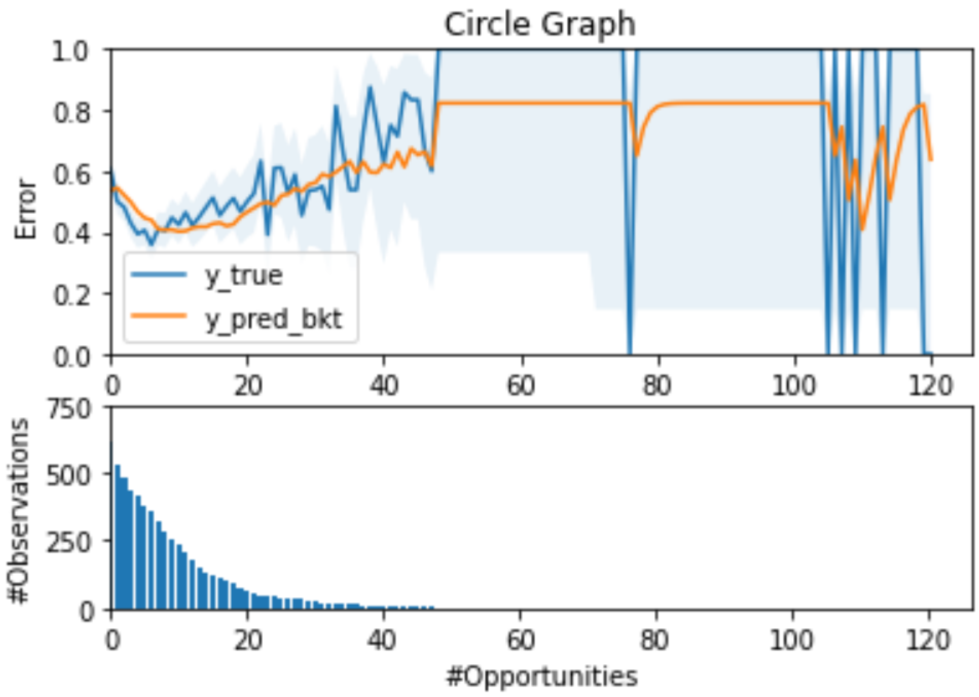
➡ Choose the next appropriate activity

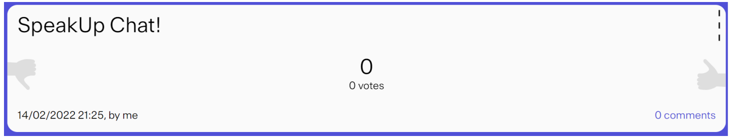
➡ Know which activities support learning

Building a learning curve for skill s

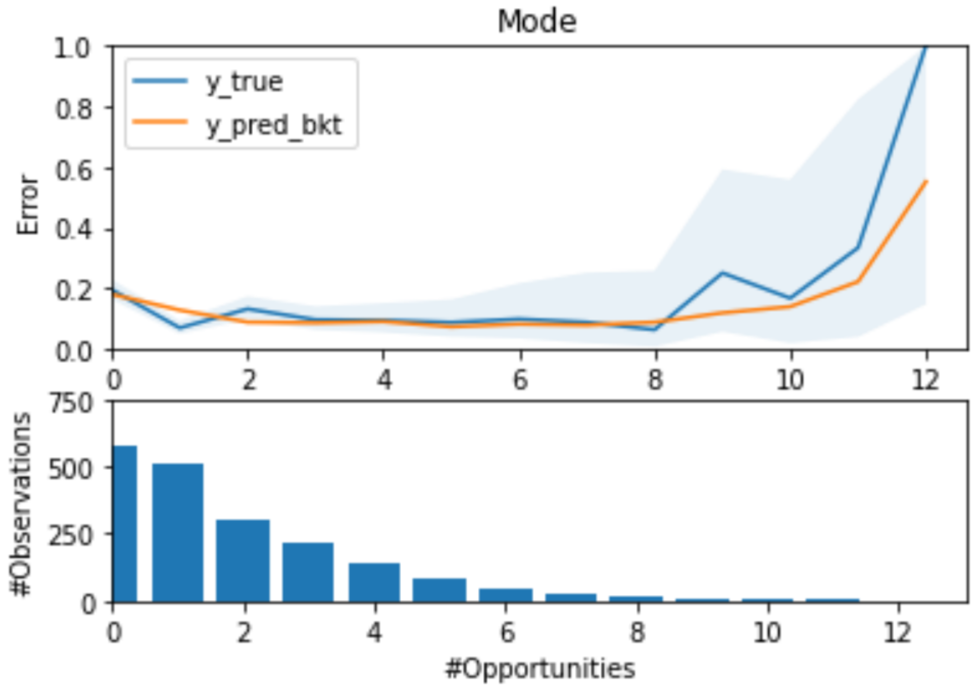
Student	Opportunity	y_true	y_pred
0	0	0	0.3
0	1	0	0.5
0	2	1	0.7
0	3	1	0.9
1	0	0	0.3
1	1	1	0.5
2	0	0	0.3
2	1	1	0.5
2	2	1	0.7
3	0	1	0.3
3	1	0	0.7
3	2	1	0.5
3	3	1	0.9

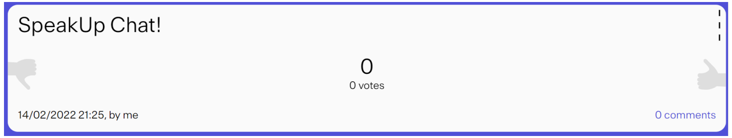
What could this curve indicate?





What could this curve indicate?



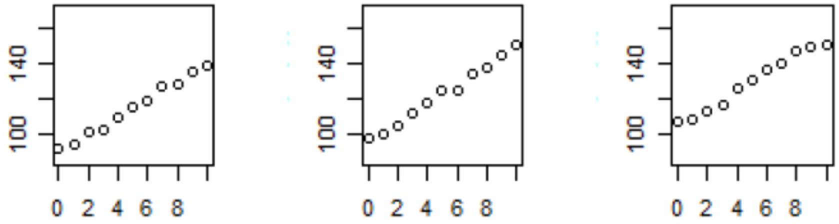


Today

- Learning Curves
- Alternative Models for Knowledge Tracing:
  - **AFM/PFA**
  - Deep Knowledge Tracing

Generalized Linear Mixed Effects Models revisited

- Example: strength gain by weight training
  - Each person has individual starting strength



$$y_n = \beta_0 + u_n + \beta_1 x_{n,1}$$

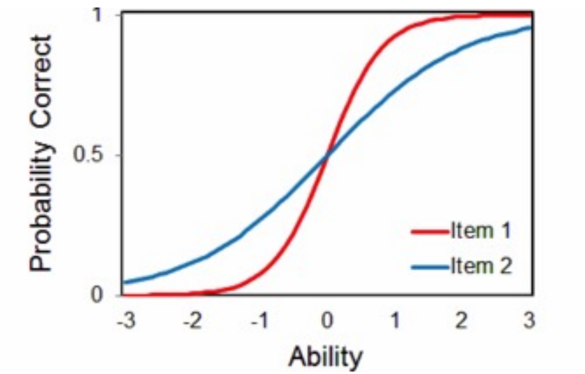
“Fixed” Effects

“Random” Effect

Rasch Model

$$\log\left(\frac{p_{i,n}}{1-p_{i,n}}\right) = \theta_n - b_i$$

Probability that student n will solve item i correctly.



θ_n: student ability

b_i: difficulty of item i

Additive Factors Model (AFM)

$$\boxed{p_{n,i} = \frac{1}{1 + e^{-\pi_{n,i}}}}$$

Probability that student n will solve task i correctly.

Additive Factors Model (AFM)

$$p_{n,i} = \frac{1}{1 + e^{-\pi_{n,i}}}$$

$$\pi_{n,i} = \theta_n + \sum_k q_{i,k} \cdot (\beta_k + \gamma_k \cdot T_{n,k})$$

[Cen et al., ITS 2006; Cen et al., IST 2008]

Additive Factors Model (AFM)

$$p_{n,i} = \frac{1}{1 + e^{-\pi_{n,i}}}$$

$$\pi_{n,i} = \theta_n + \sum_k q_{i,k} \cdot (\beta_k + \gamma_k \cdot T_{n,k})$$

Student proficiency

[Cen et al., ITS 2006; Cen et al., IST 2008]

Additive Factors Model (AFM)

$$p_{n,i} = \frac{1}{1 + \ e^{-\pi_{n,i}}}$$

$$\pi_{n,i} = \theta_n + \sum_k q_{i,k} \cdot (\beta_k + \gamma_k \cdot T_{n,k})$$

Student proficiency

$q_{ik} = 1,$ if item *i* uses skill *k*

[Cen et al., ITS 2006; Cen et al., IST 2008]

Additive Factors Model (AFM)

$$p_{n,i} = \frac{1}{1 + e^{-\pi_{n,i}}}$$

$$\pi_{n,i} = \theta_n + \sum_k q_{i,k} \cdot (\beta_k + \gamma_k \cdot T_{n,k})$$

Student proficiency

Difficulty of skill *k*

$q_{ik} = 1,$ if item *i* uses skill *k*

Additive Factors Model (AFM)

$$p_{n,i} = \frac{1}{1 + e^{-\pi_{n,i}}}$$

$$\pi_{n,i} = \theta_n + \sum_k q_{i,k} \cdot (\beta_k + \gamma_k \cdot T_{n,k})$$

Student proficiency

$q_{ik} = 1,$ if item *i* uses skill *k*

Difficulty of skill *k*

Learning rate at skill *k*

Number of practice opportunities student *n* had at skill *k*

[Cen et al., ITS 2006; Cen et al., IST 2008]

AFM - Assumptions

- Students may initially know more or less
- Students learn at the same rate
- Some skills are more likely to initially be known
- Some skills are easier to learn than others
- Students learn with each practice opportunity
- Each item belongs to one or more skills

Performance Factors Analysis (PFA)

$$\pi_{n,i} = \theta_n + \sum_k q_{i,k} \cdot (\beta_k + \boxed{\gamma_k \cdot T_{n,k}})$$

[Pavlik et al., AIED 2009]

Performance Factors Analysis (PFA)

$$\pi_{n,i} = \theta_n + \sum_k q_{i,k} \cdot (\beta_k + \boxed{\gamma_k \cdot s_{n,k} + \rho_k \cdot f_{n,k})}$$

Number of prior **successes** student *n* had at skill *k*

Number of prior **failures** student *n* had at skill *k*

[Pavlik et al., AIED 2009]

PFA - Assumptions

- Students may initially know more or less
- Students learn at the same rate
- Some skills are more likely to initially be known
- Some skills are easier to learn than others
- Students learning rate differs for correct and wrong practice opportunities
- Each item belongs to one or more skills

AFM/PFA in action…

- Jupyter Notebook

Today

- Learning Curves
- Alternative Models for Knowledge Tracing:
  - AFM/PFA
  - **Deep Knowledge Tracing**

Neural Networks

- We are given a sample data set:

$$T = \{(y_n, \boldsymbol{x}_n)\} \text{ with } n = 1, \ldots N,$$

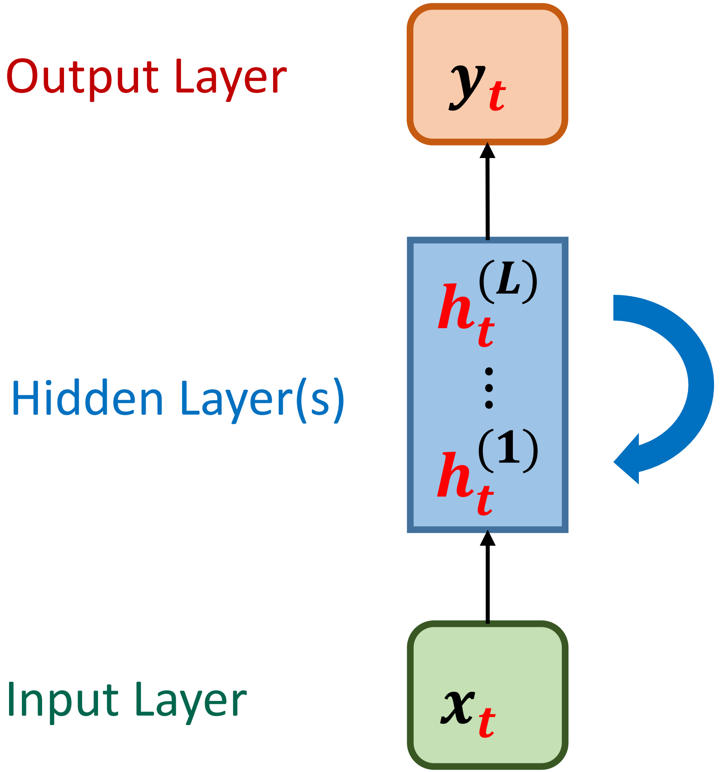
where y_n is the n'th output variable and $\boldsymbol{x}_n$ is a D-dimensional vector of input variables

- Goal: learn a model f such that $y_n \approx f(\boldsymbol{x}_n)$
- Linear regression/classification usually requires a lot of *feature engineering*

Neural Networks

- Neural networks are able to represent non-linear functions, i.e. $y_n \approx f(\boldsymbol{x}_n)$ can be non-linear
- Neural networks are able to *learn* the features and the weights (parameters) from the data
- Tutorial: http://neuralnetworksanddeeplearning.com/

Recurrent Neural Network



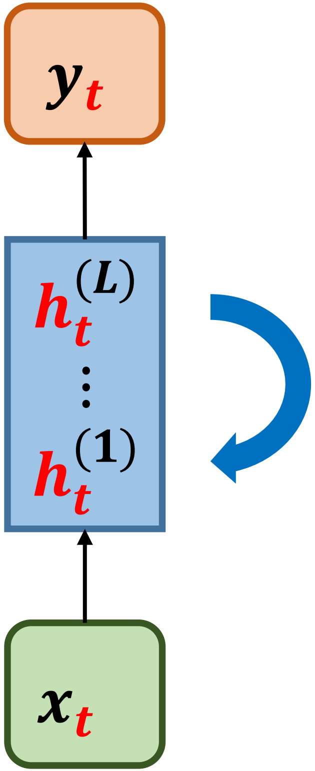
$$\boldsymbol{h}_t^{(1)} = \phi_h \left(W_{hx} \boldsymbol{x}_t + W_{h^{(1)}h^{(1)}}^{(1)} \boldsymbol{h}_{t-1}^{(1)} + \boldsymbol{b}^{(1)} \right)$$

Recurrent Neural Network

Output Layer

Hidden Layer(s)

Input Layer



$$\boldsymbol{h}_t^{(l)} = \phi_h(W^{(l)}_{h^{(l)}h^{(l-1)}} \boldsymbol{h}_t^{(l-1)} + W^{(l)}_{h^{(l)}h^{(l)}} \boldsymbol{h}_{t-1}^{(l)} + \boldsymbol{b}^{(l)})$$

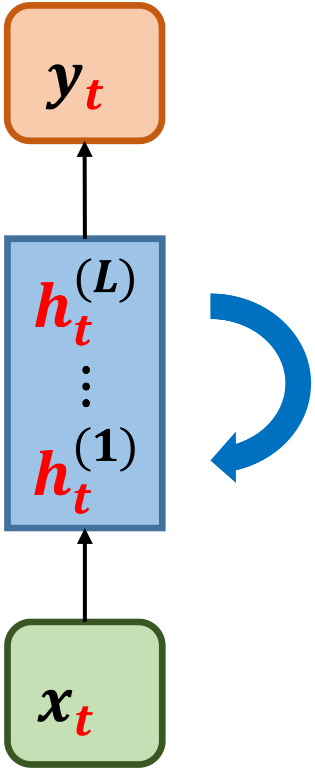
$$\boldsymbol{h}_t^{(1)} = \phi_h\left(W_{hx}\boldsymbol{x}_t + W^{(1)}_{h^{(1)}h^{(1)}} \boldsymbol{h}_{t-1}^{(1)} + \boldsymbol{b}^{(1)}\right)$$

Recurrent Neural Network

Output Layer

Hidden Layer(s)

Input Layer

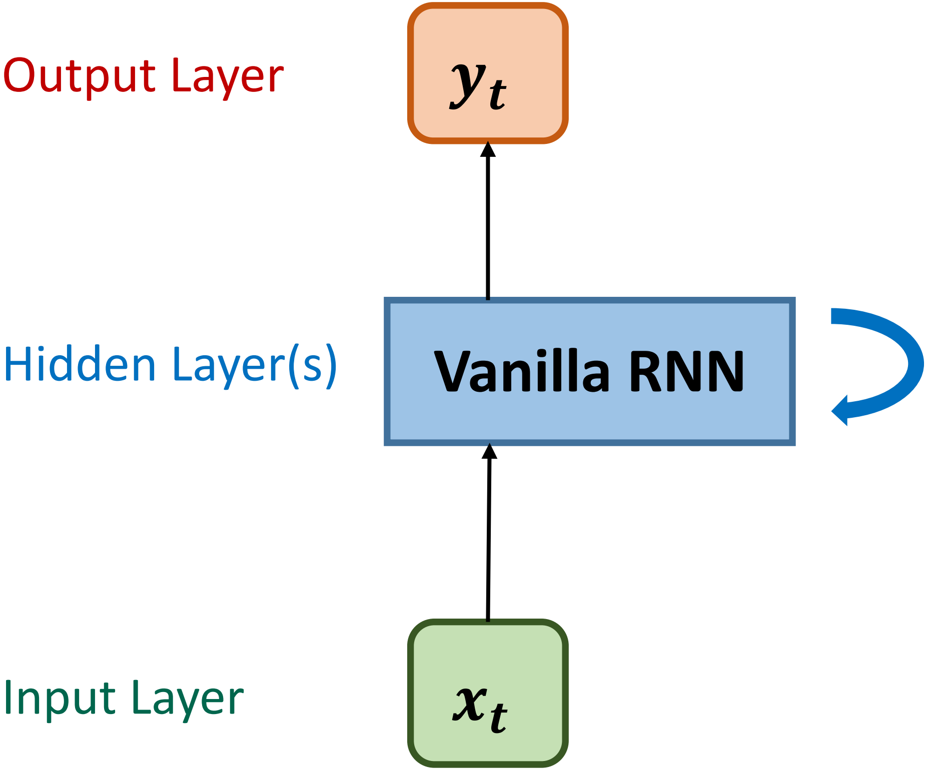


$$\boldsymbol{y}_t = \phi_y(W_{yh}\boldsymbol{h}_t^L + \boldsymbol{b}^{(y)})$$

$$\boldsymbol{h}_t^{(l)} = \phi_h(W_{h^{(l)}h^{(l-1)}}^{(l)}\boldsymbol{h}_t^{(l-1)} + W_{h^{(l)}h^{(l)}}^{(l)}\boldsymbol{h}_{t-1}^{(l)} + \boldsymbol{b}^{(l)})$$

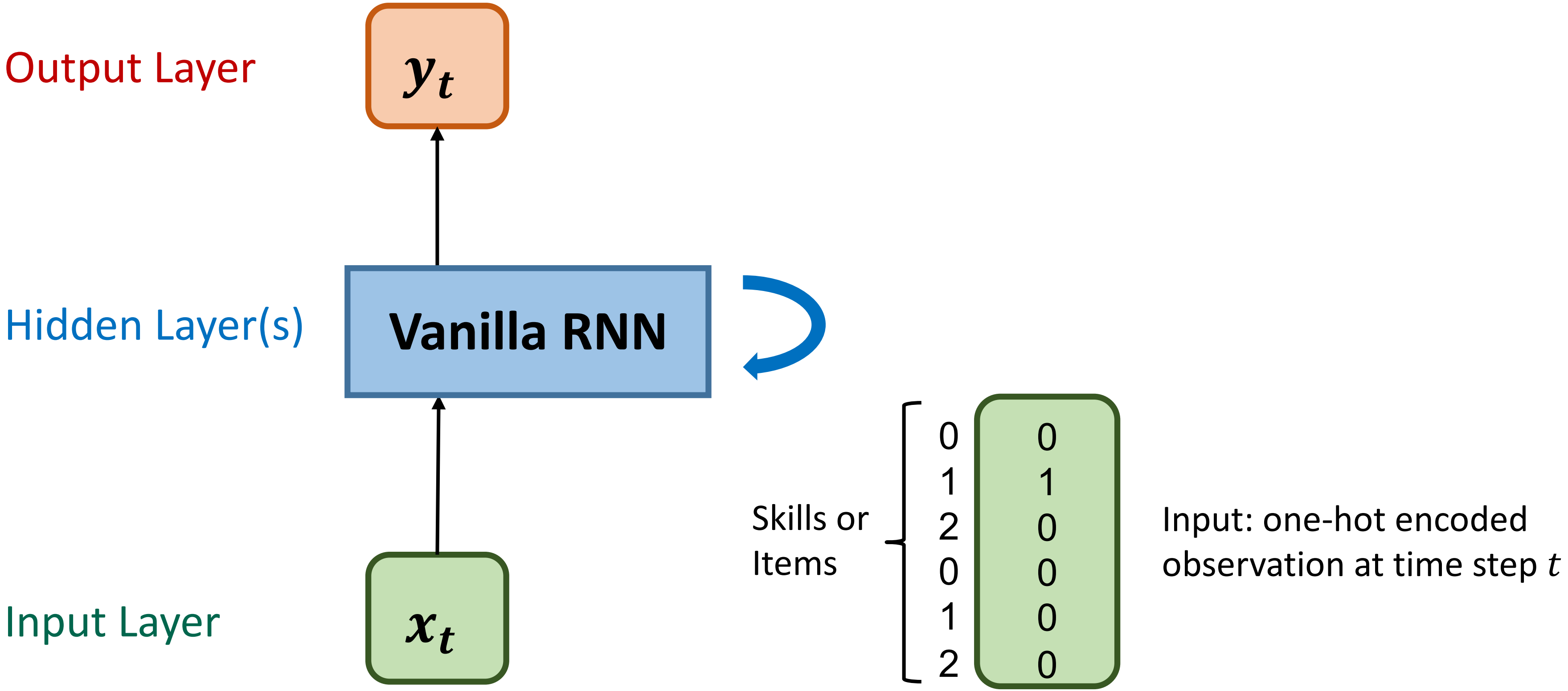
$$\boldsymbol{h}_t^{(1)} = \phi_h\,(W_{hx}\boldsymbol{x}_t + W_{h^{(1)}h^{(1)}}^{(1)}\boldsymbol{h}_{t-1}^{(1)} + \boldsymbol{b}^{(1)})$$

Deep Knowledge Tracing



[*Deep Knowledge Tracing*, Piech et al., NIPS 2015]

Deep Knowledge Tracing



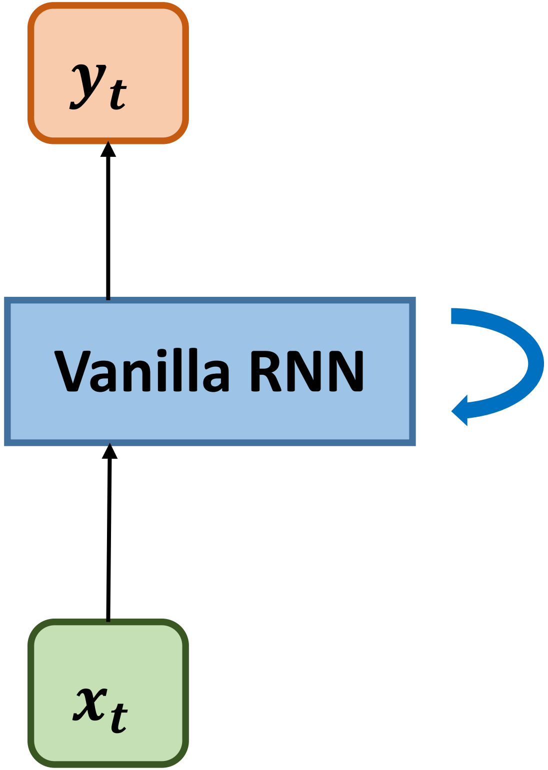
[*Deep Knowledge Tracing*, Piech et al., NIPS 2015]

Deep Knowledge Tracing

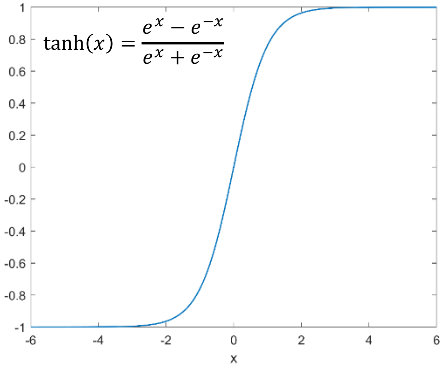
Output Layer

Hidden Layer(s)

Input Layer



$$\boldsymbol{h_t} = \tanh(W_{hx}\boldsymbol{x_t} + W_{hh}\boldsymbol{h_{t-1}} + \boldsymbol{b_h})$$

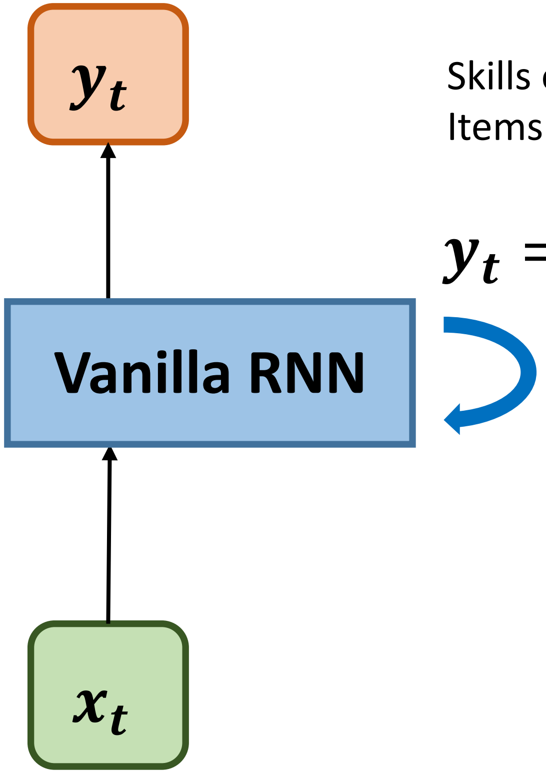


Deep Knowledge Tracing

Output Layer

Hidden Layer(s)

Input Layer

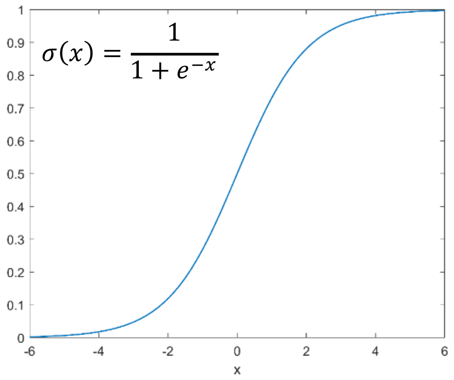


Skills or Items

0	0.3
1	0.5
2	0.7

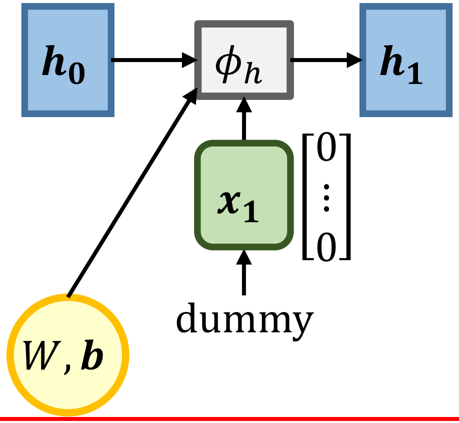
Output: probability for answering skill (or item) correct at time step t

$$\boldsymbol{y_t} = \sigma(W_{yh}\boldsymbol{h_t} + \boldsymbol{b_y})$$

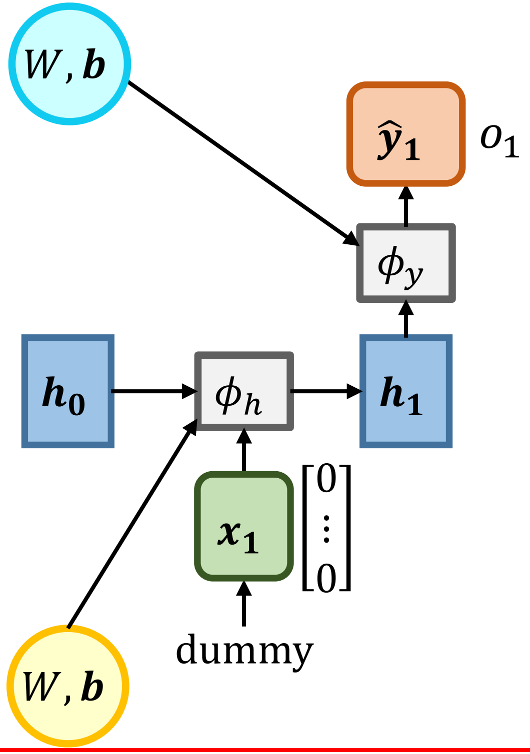


[*Deep Knowledge Tracing*, Piech et al., NIPS 2015]

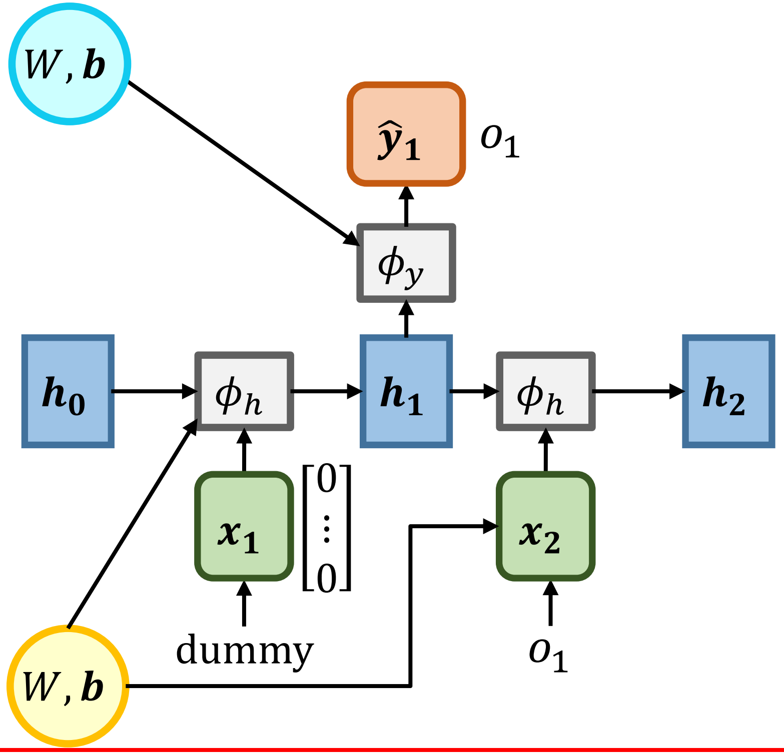
Deep Knowledge Tracing – Computational Graph



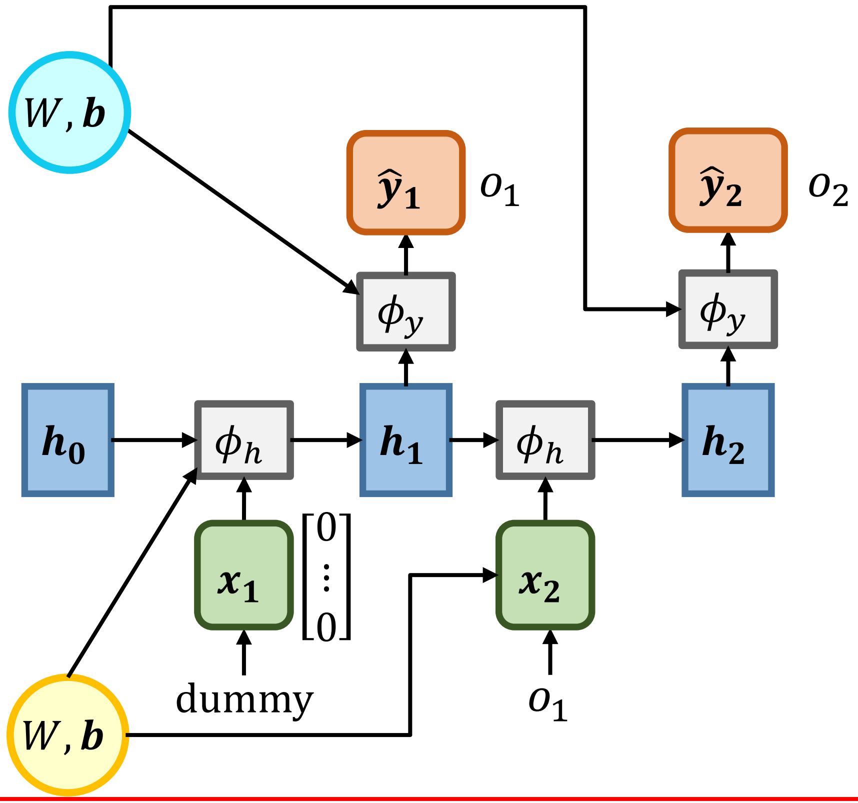
Deep Knowledge Tracing – Computational Graph



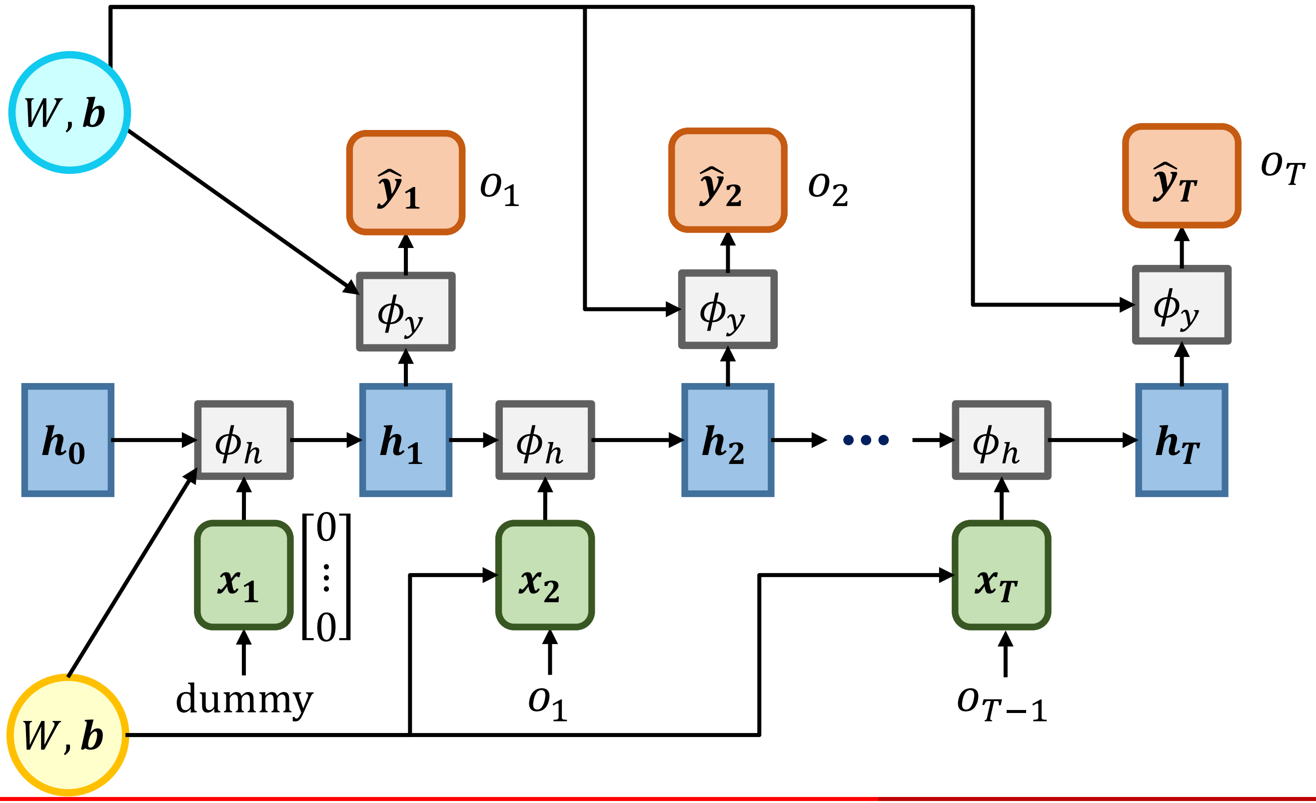
Deep Knowledge Tracing – Computational Graph



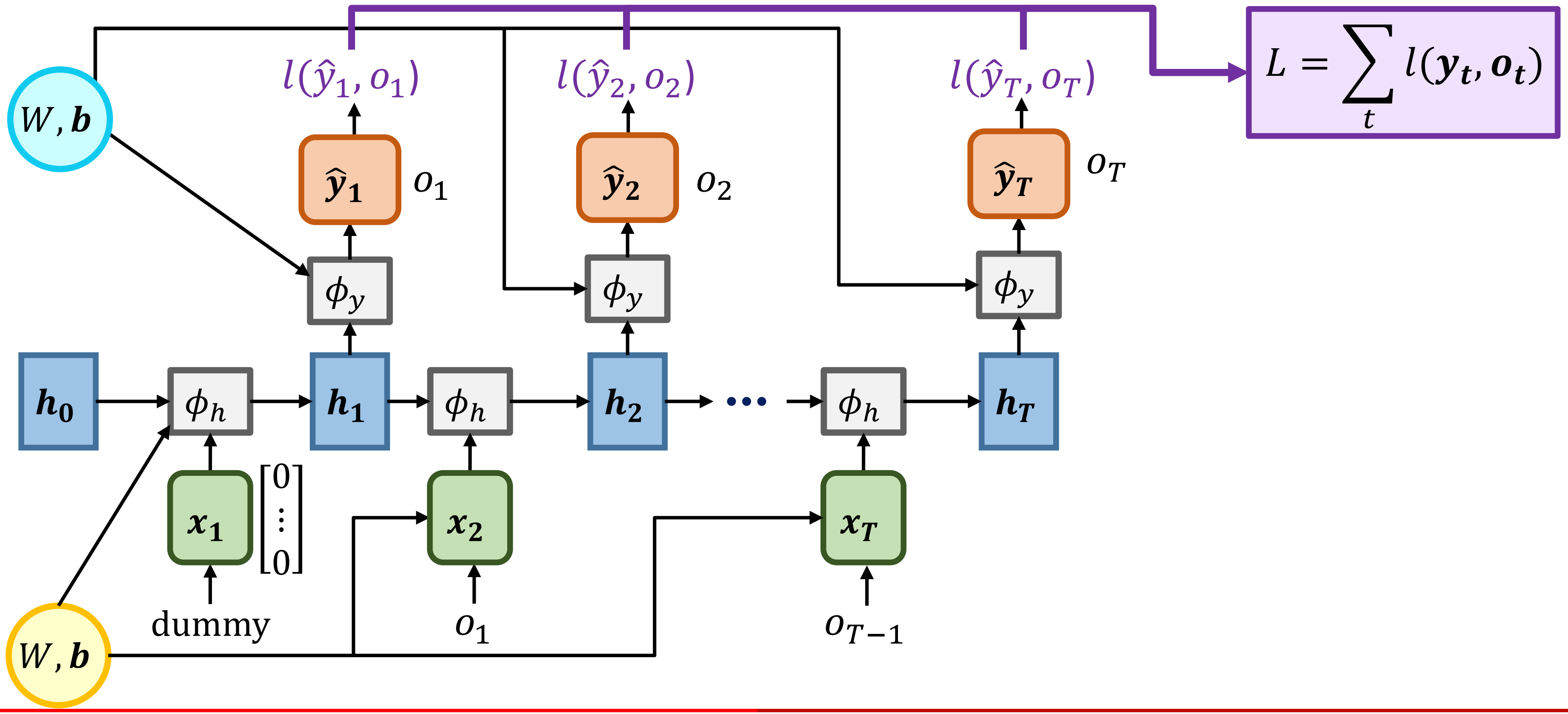
Deep Knowledge Tracing – Computational Graph



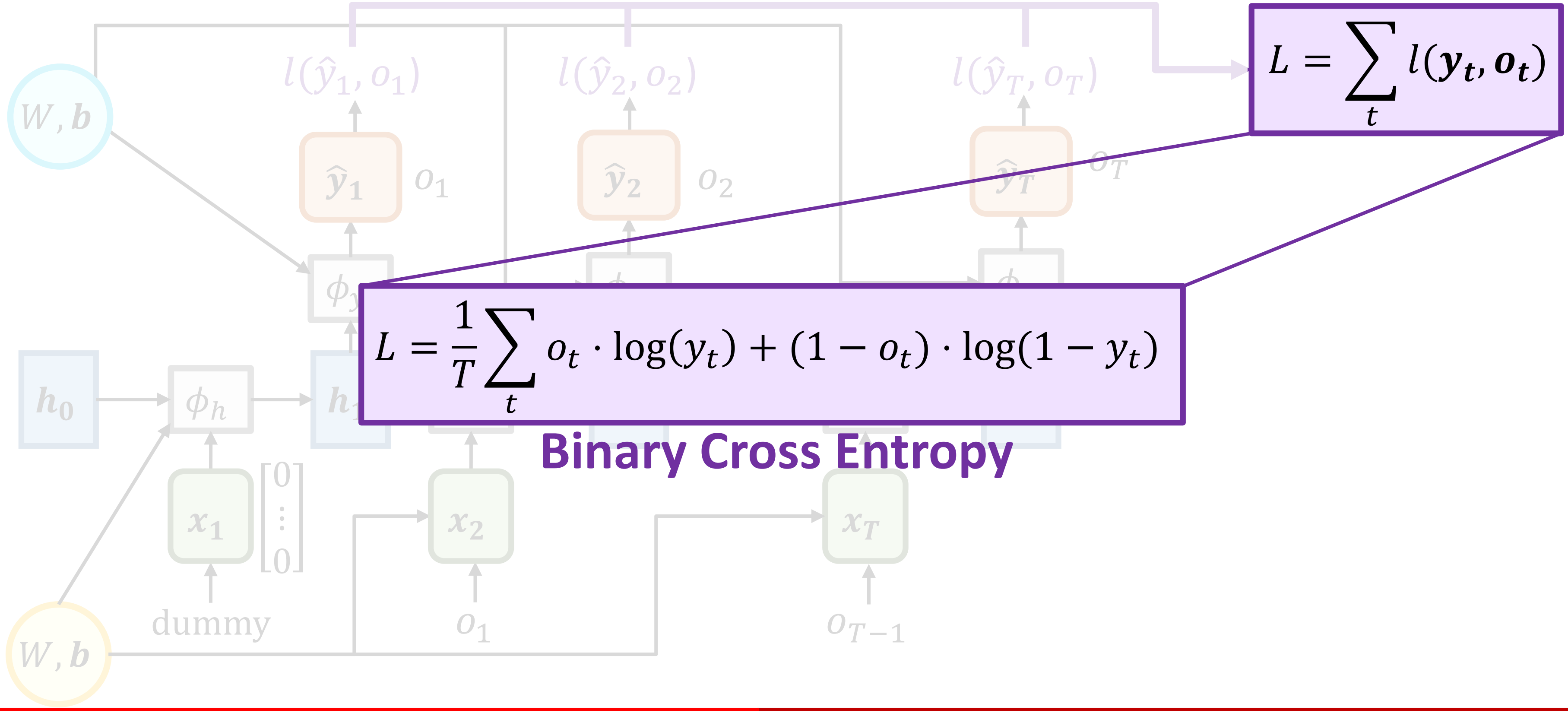
Deep Knowledge Tracing – Computational Graph



Deep Knowledge Tracing – Computational Graph



Training a DKT model: Binary Crossentropy Loss



$$L = \sum_t l(\boldsymbol{y}_t, \boldsymbol{o}_t)$$

$$L = \frac{1}{T} \sum_t o_t \cdot \log(y_t) + (1 - o_t) \cdot \log(1 - y_t)$$

Binary Cross Entropy

Your Turn: Comparing Models

- We have evaluated AFM, PFA, BKT, and DKT on a subset of six skills. Your task:
  - Visualize the overall RMSE and AUC of the models such that it can easily be compared
  - Discuss the obtained results
- We have also evaluated BKT and DKT on the full data set and plotted the results. Your task:
  - Do you see a difference between the results on the full model and on the subset? If so, why?

Summary

- Learning Curves
- Alternative Models for Knowledge Tracing:
  - AFM/PFA
  - Deep Knowledge Tracing

Final Project Presentations

- Poster Session
- May 30, 15.15-17.00 (location: EPFL campus)
- **Mandatory** presence of all team members
- There will be prices and snacks/drinks…

Don’t forget – M3 is due today

Week	Project Hours	Milestones
1	Detailed project presentation	-
2	Introduction to tasks for M2	*M1: Preferences on team members and data sets (Feb 28, 23:59)*
3	Office hours	
4	Office hours	
5	Introductions to tasks for M4	*M2: **Individual** exploration of selected data set (March 21, 23:59)*
6	Office hours	
7	Individual discussion with teams	*M3: selection of research question and approach (April 4, 23:59)*
8	Office hours	

- Submission for M3 is **mandatory**, one submission per team
- Meetings with TAs on April 6 are **mandatory**