\Orchestrating a brighter world NEC

Non-compensatory Knowledge Tracing with Local Variational Approximation

Hiroshi Tamano (NEC, The Graduate University for Advanced Studies)
Daichi Mochihashi (The Institute of Statistical Mathematics)

Outline

- Background and Purpose
- Proposed Method
  - Generative Model
  - Posterior and Parameter Inference
- Experiments
- Conclusion

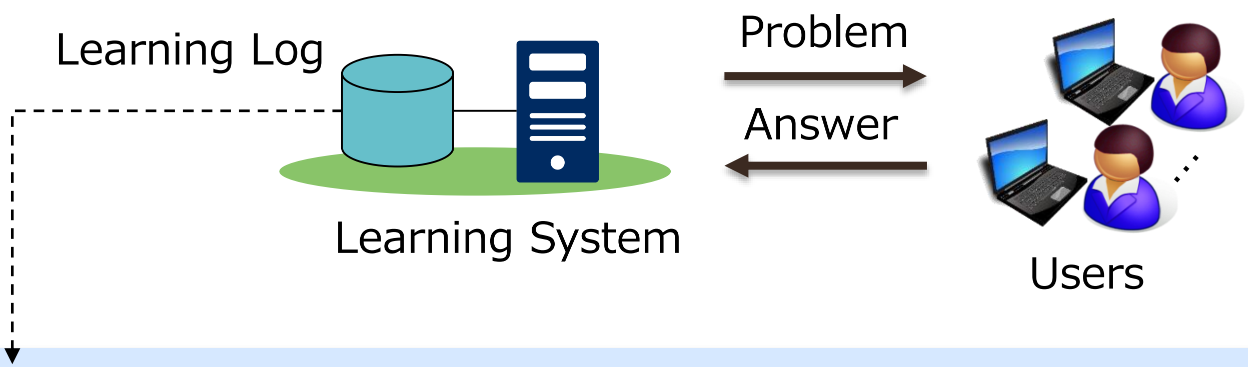
Outline

- Background and Purpose
- Proposed Method
  - Generative Model
  - Posterior and Parameter Inference
- Experiments
- Conclusion

Knowledge Tracing (KT)

Knowledge tracing is a ML task to trace learner's latent skill and predict whether he can solve a next problem or not

- KT can be used in personalized education
- KT has been studied in state space models, RNNs, factorizations



Train Data

User 1 [(Problem2, ○), (Problem4, ×), ⋯, (Problem56, ○)]

⋮

User N [(Problem3, ○), (Problem9,×), ⋯, (Problem34, ×)]

Test Data

New User [(Problem5, ○), (Problem9, ×), (Problem12, ?)]

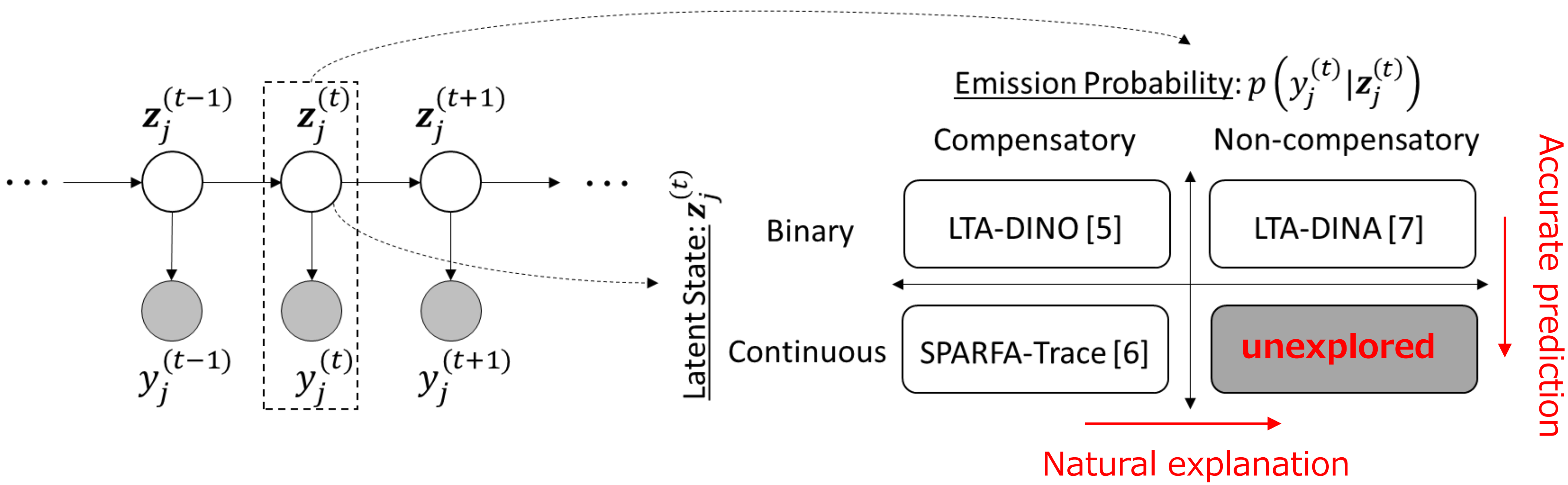
⋮

Predict

State space models in KT

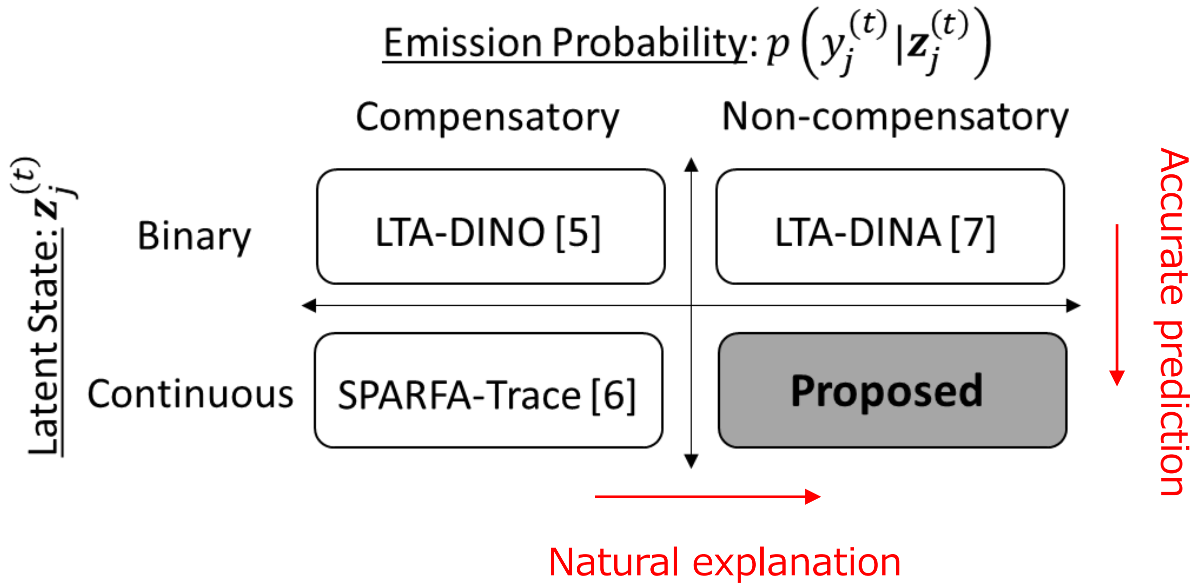
Four categories (of models handling problems requiring multiple skills)

- Latent State: binary or continuous vector
- Emission Probability: compensatory or non-compensatory
  - It means whether each skill can complement other skills or not
  - E.g., Problem: $1/5\,x+3/10=2x$
    Fraction and equation skills are required. Non-compensatory assumption is natural.



Purpose

- We explore the potential of state space model KT which has continuous skill state and non-compensatory emission
- This model is expected to give **accurate prediction** and **natural explanation**



Outline

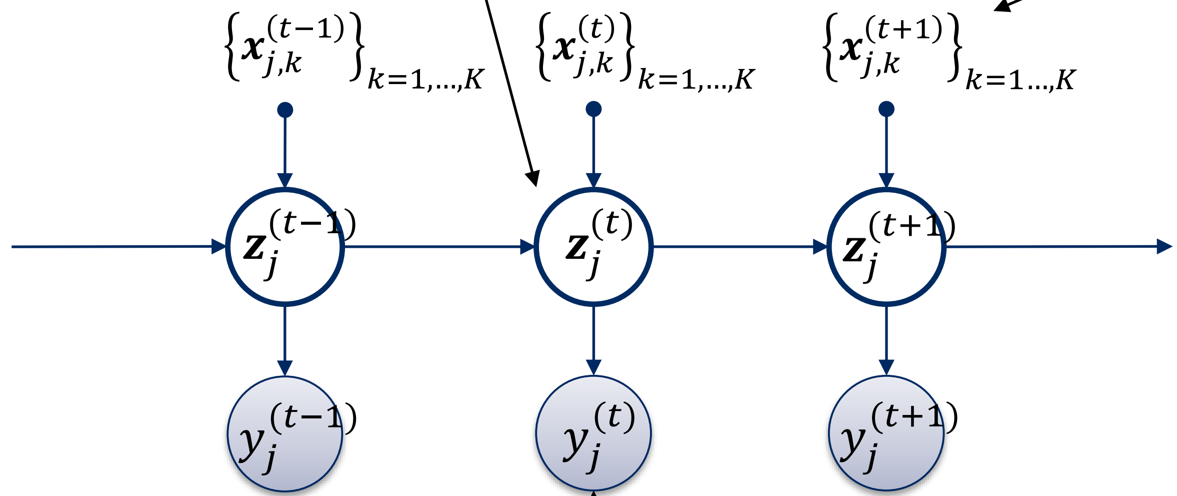
- Background and Purpose
- Proposed Method
  - Generative Model
  - Posterior and Parameter Inference
- Experiments
- Conclusion

Generative Model (1/4)

Linear Dynamical System + Non-compensatory Multi-dimensional IRT

latent skill state : $\boldsymbol{z}_j^{(t)} \in R^K$

covariates : $\boldsymbol{x}_{j,k}^{(t)} \in R^{F_k}$

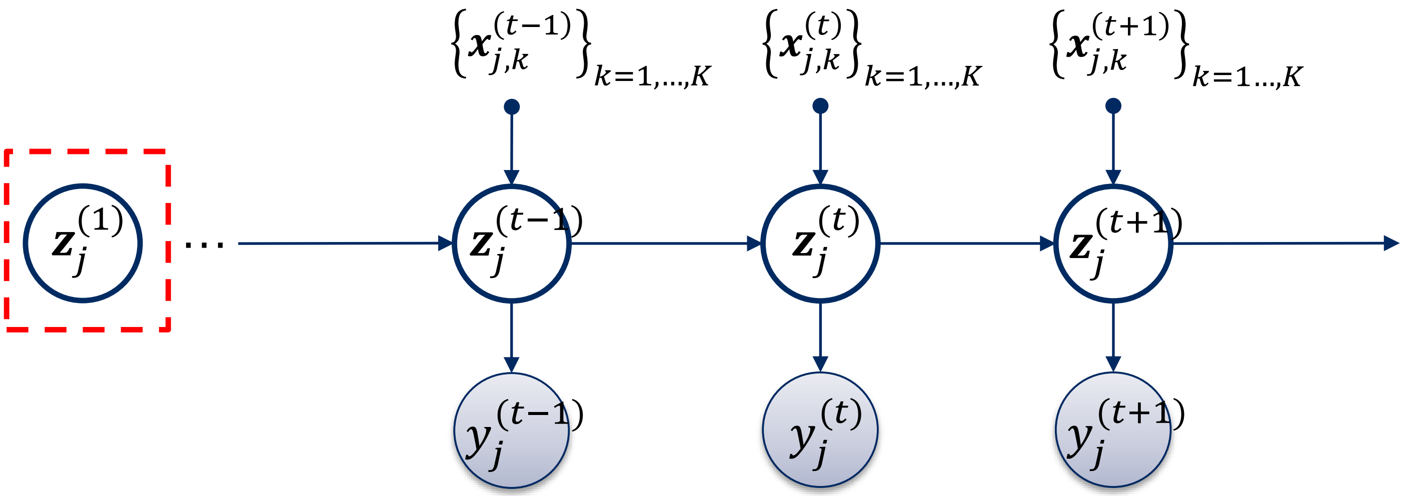


student response : $y_j^{(t)} \in \{0, 1\}$

Generative Model (2/4)

Initial State Probability

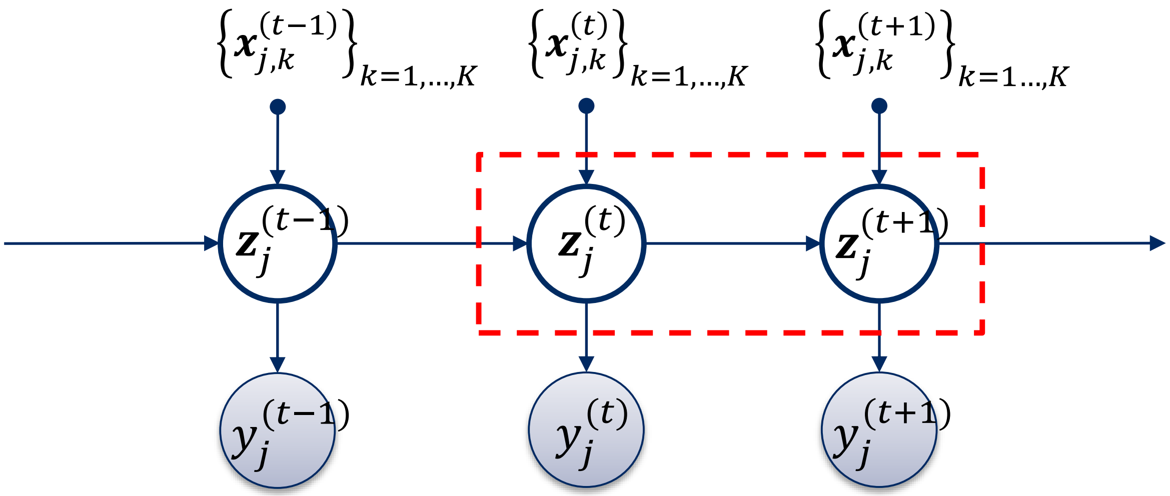
$$P\left(\boldsymbol{z}_j^{(1)}; \mu_0, P_0\right) = N\left(\boldsymbol{z}_j^{(1)} \middle| \mu_0, P_0\right)$$



Generative Model (3/4)

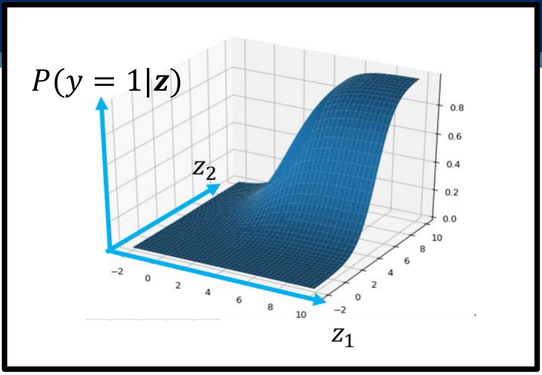
State Transition Probability

$$P\left(z_j^{(t+1)} \middle| z_j^{(t)}; D, \beta, \Gamma\right) = N\left(z_j^{(t+1)} \middle| D_{i(j,t)} z_j^{(t)} + \begin{bmatrix} \vdots \\ \beta_k^T x_{j,k}^{(t+1)} \\ \vdots \end{bmatrix}, \;\; \Gamma_{i(j,t+1)}\right)$$



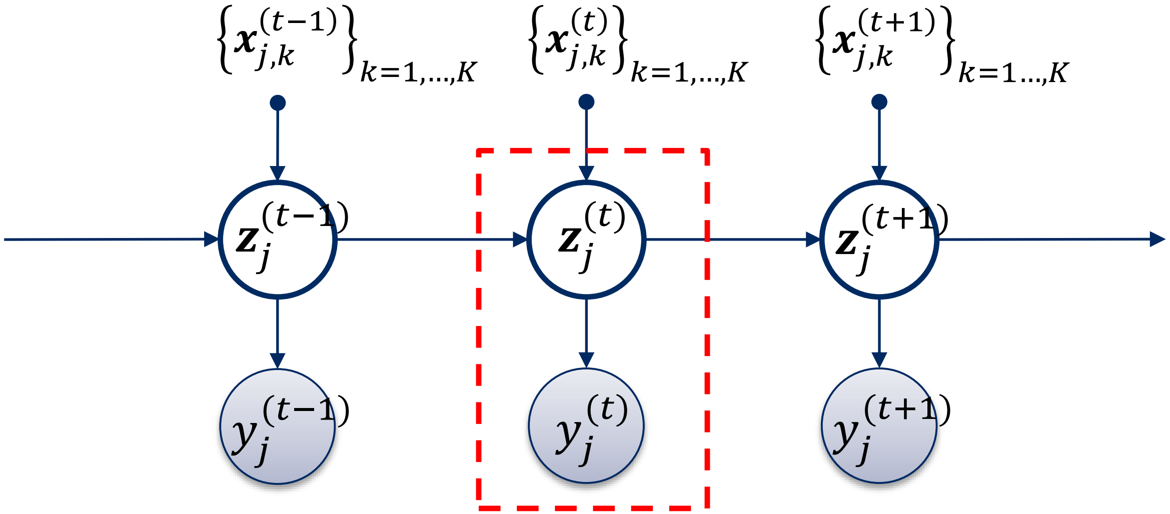
$i(j,t)$: question index user j answered at time t
(abbreviate i if it's clear from context)

Generative Model (4/4)



Emission Probability (=Non-compensatory model in MIRT)

$$P\left(y_j^{(t)} = 1 \middle| \boldsymbol{z}_j^{(t)}; \boldsymbol{a}, \boldsymbol{b}, Q\right) = \prod_k \sigma\left(a_{i,k}\left(z_{j,k}^{(t)} - b_{i,k}\right)\right)^{Q_{i,k}}$$



question-skill map: $Q_{i,k} \in \{0,1\}$

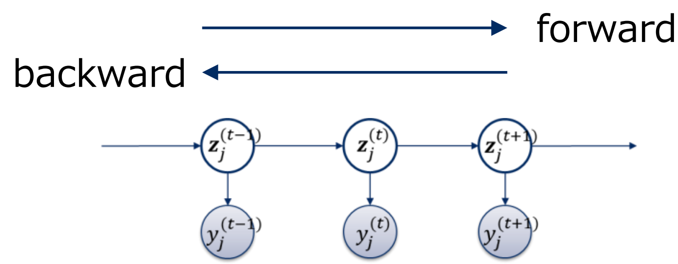
Outline

- Background and Purpose
- Proposed Method
  - Generative Model
  - Posterior and Parameter Inference
- Experiments
- Conclusion

Posterior and Parameter Inference

Linear Dynamical System

- Posterior Inference: forward backward algorithm
- Parameter Inference: EM algorithm



Non-compensatory KT

- Posterior Inference: forward backward algorithm **with Gaussian approximation**

<u>Forward message ($\hat{\alpha}$ message)</u>

$$\hat{\alpha}\left(z_j^{(t)}\right) = \underbrace{P\left(z_j^{(t)} \middle| y_j^{(1)}, \dots, y_j^{(t)}\right)}_{\text{posterior}} \propto \underbrace{P\left(y_j^{(t)} \middle| z_j^{(t)}\right)}_{\text{likelihood}} \underbrace{P\left(z_j^{(t)} \middle| y_j^{(1)}, \dots, y_j^{(t-1)}\right)}_{\text{prior}}$$

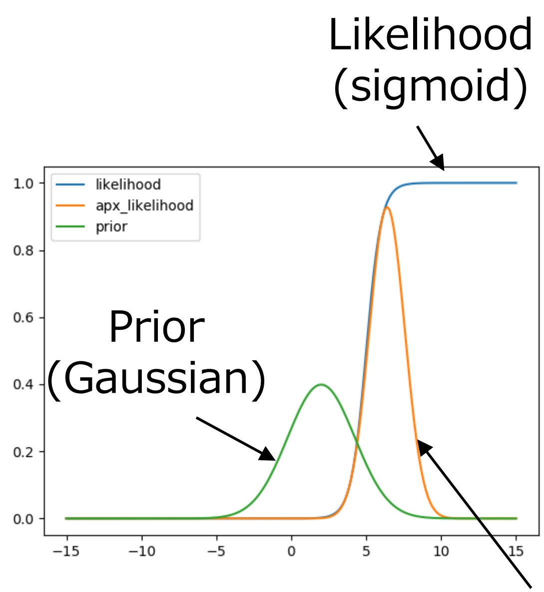
We approximate this as Gaussian using local variational method

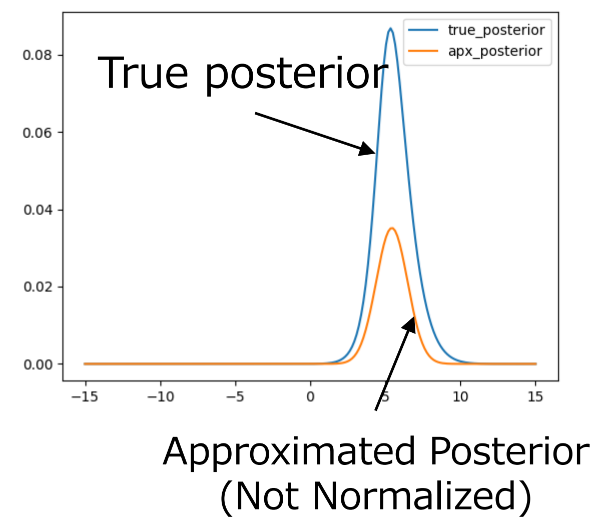
- Parameter Inference: (Monte Carlo) EM algorithm

Local Variational Method [Jaakkola & Jordan, 2000]

Locally approximating a likelihood around a prior gives accurate posterior approximation

Posterior $\propto$ Likelihood (Sigmoid) x Prior (Gauss)



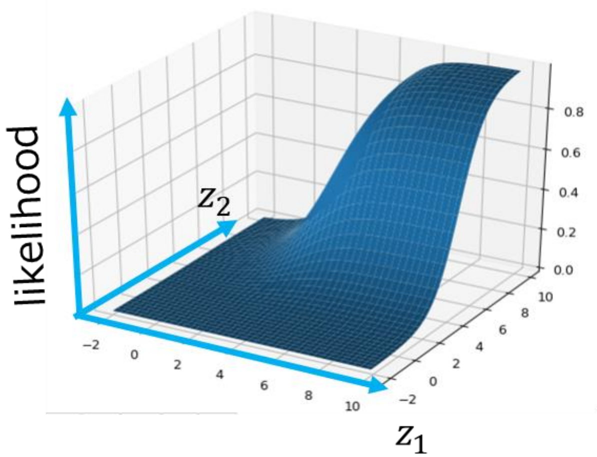


variational lower bound in an exponential quadratic form

Our likelihood function to be approximated

Correct Answer

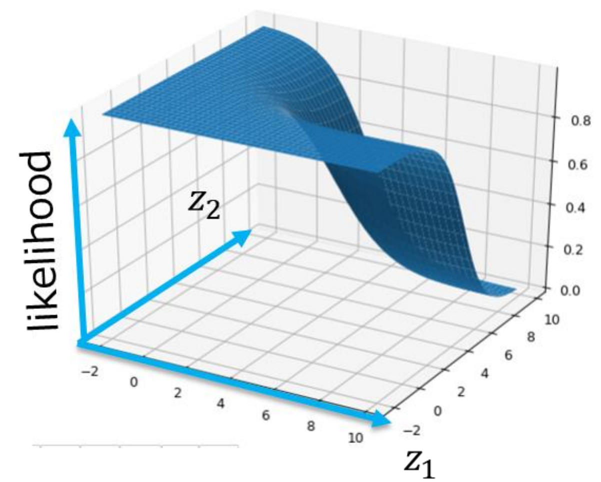
$$p(y=1|z)=\prod_{k=1}^{K_i}\sigma\big(a_{i,k}(z_k-b_{i,k})\big)$$



apply Jaakkloa's method
for each sigmoid

Incorrect Answer

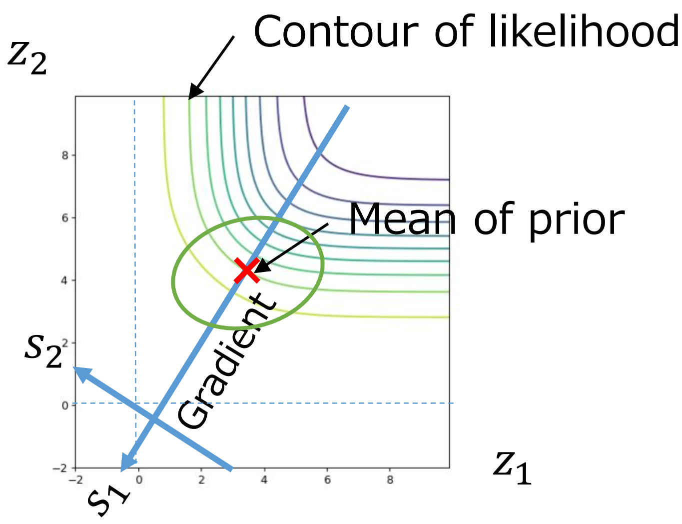
$$p(y=0|z)=1-\prod_{k=1}^{K_i}\sigma\big(a_{i,k}(z_k-b_{i,k})\big)$$

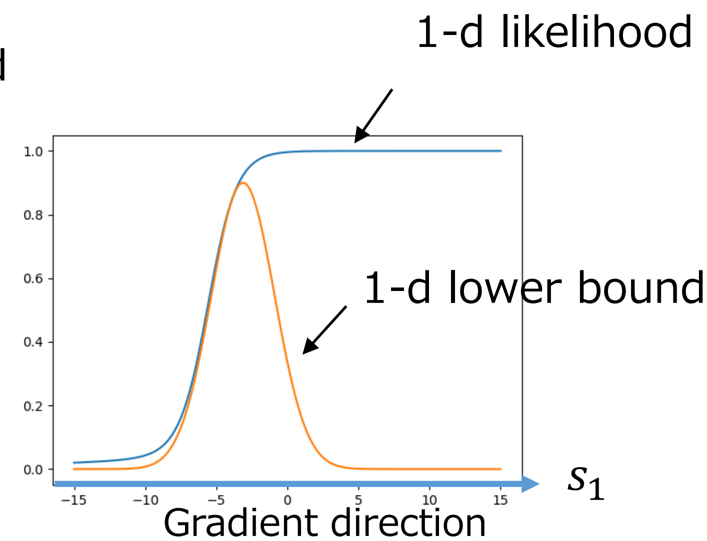


need more steps

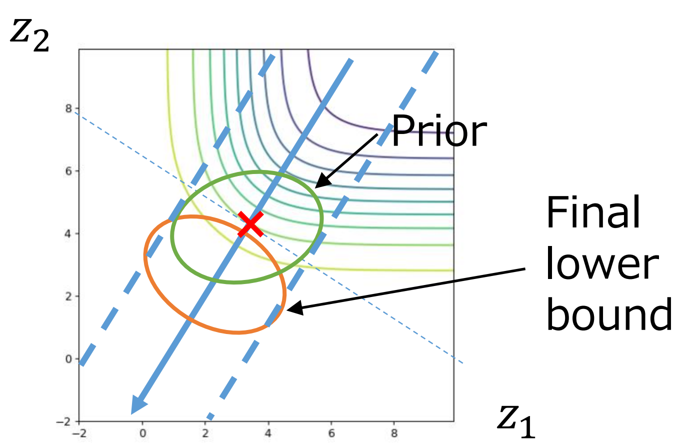
Rough image to approximate likelihood for incorrect answer

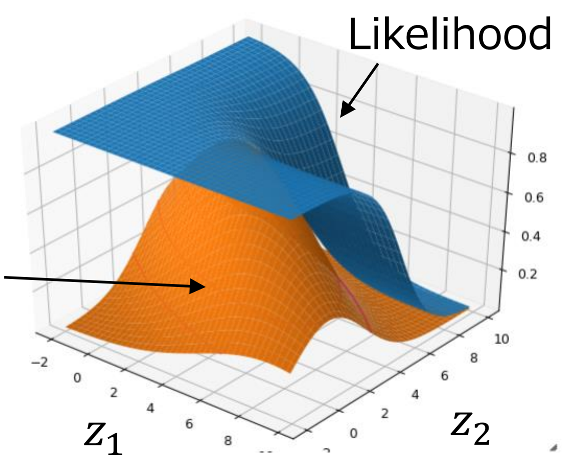
Step1





Step2





Outline

- Background and Purpose
- Proposed Method
  - Generative Model
  - Posterior and Parameter Inference
- Experiments
- Conclusion

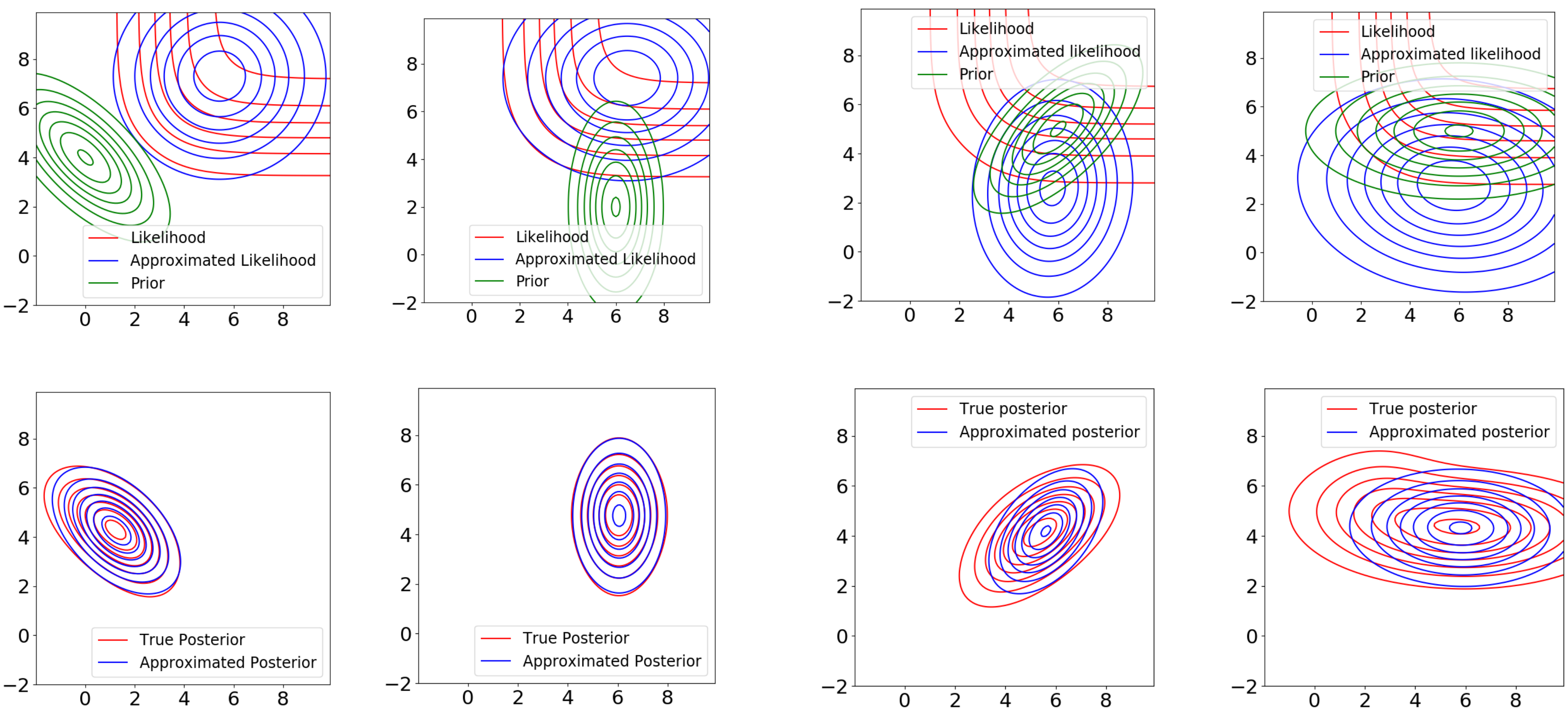
Experiments

- Approximation of $\hat{\alpha}$ message on artificial data
- Prediction performance on Assistment 2009-2010
- Visualization of explanation

Approximation of $\hat{\alpha}$ message

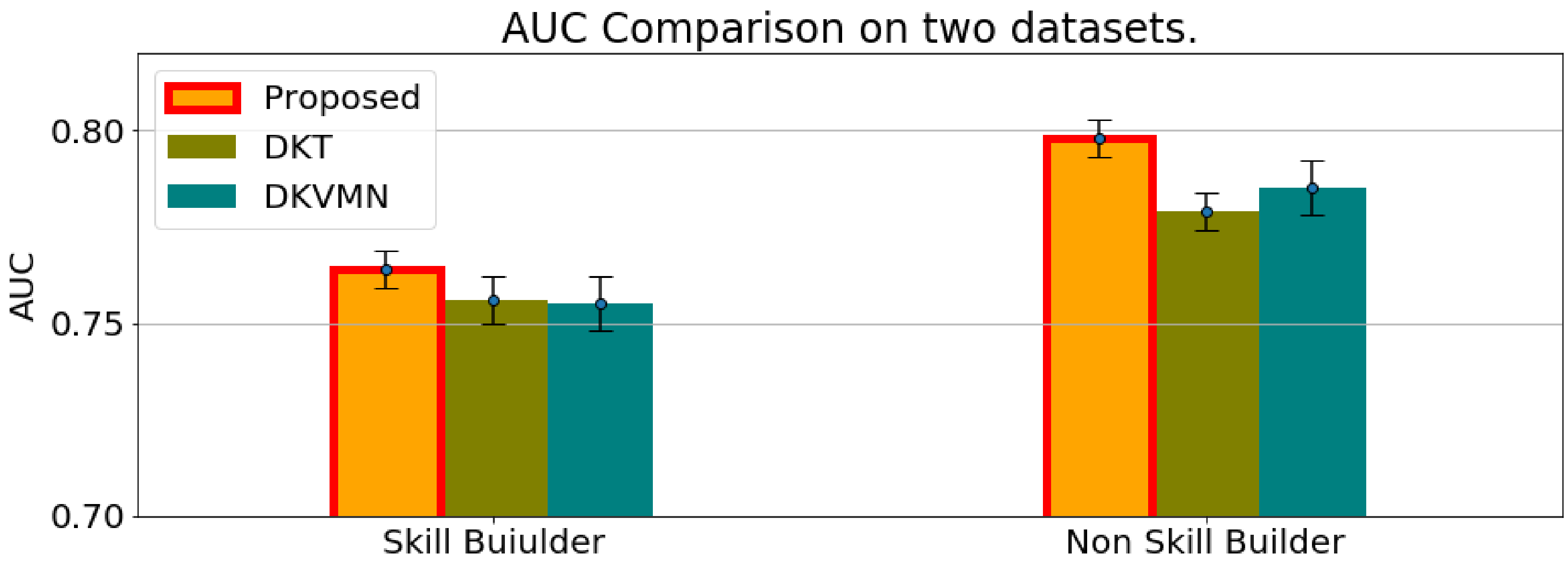
Correct Answer | Incorrect Answer

(a) (b) (c) (d)



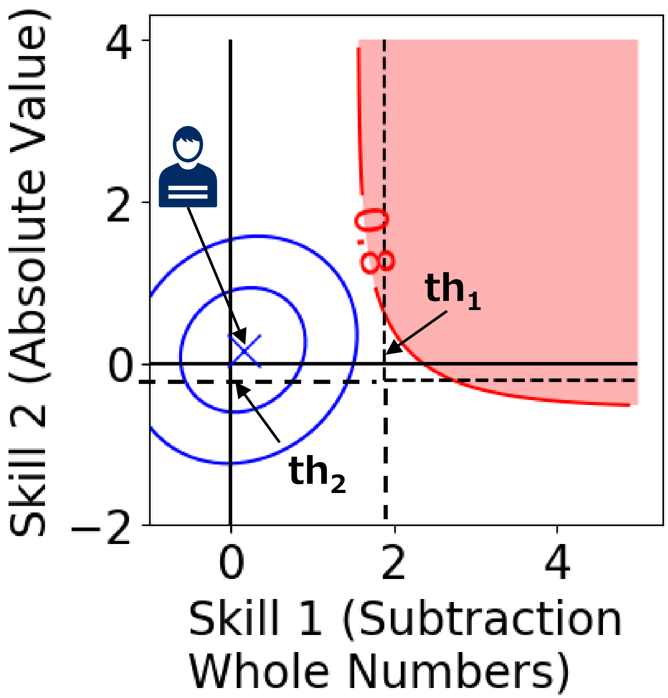
Prediction Performance

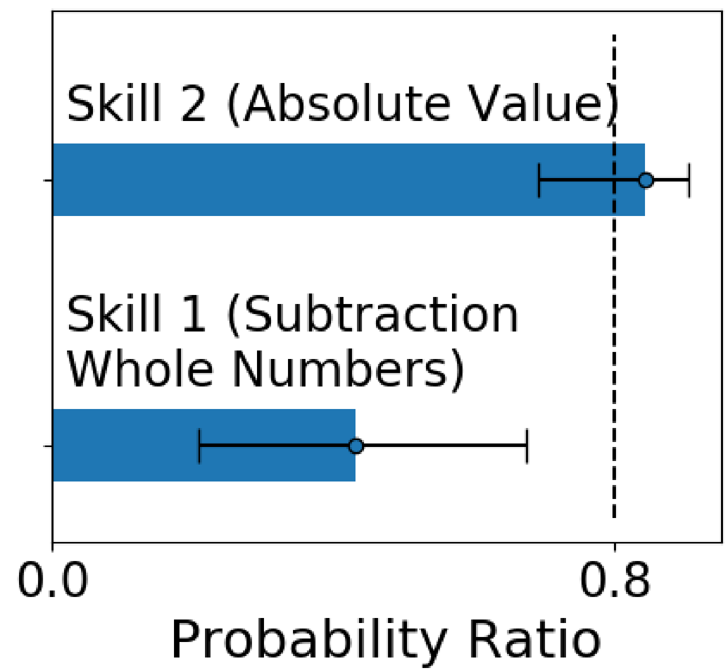
- Assistments 2009-2010 Dataset
- Main (i.e., non-scaffolding) questions answered by learners ≥ 10
- Skill Builder: 4,106 learners; 8,112 questions; 106 skills
- Non Skill Builder: 8,068 learners; 2,554 questions; 194 skills
- Compared DKT and DKVMN
  - Created joint skills for questions requiring multiple skills to prevent leakage



Visualization: explaining why a learner cannot solve a problem

Non-compensatory model can explain which skills and how much the skills are missing when he can not solve a problem





Inferred learner's state and item response function

Output explanation

Outline

- Background and Purpose
- Proposed Method
  - Generative Model
  - Posterior and Parameter Inference
- Experiments
- **Conclusion**

Conclusion

Purpose

- We explore the potential of state space model KT which has continuous skill state and non-compensatory emission
  - Accurate prediction and natural explanation

Method

- Generative model: LDS + Non-compensatory MIRT
- We proposed local Gaussian approximation to the likelihood function of non-compensatory emission by local variational method
- Posterior and parameter inference can be processed almost like LDS

Experiment

- Our variational posterior adequately approximates the true posterior
- Our model achieves better prediction accuracy compared to two popular deep learning-based methods on Assistments 2009-2010
- Visualization explaining why a learner cannot solve a problem

\Orchestrating a brighter world

NEC