THE UNIVERSITY OF UTAH

Streaming Bayesian Deep Tensor Factorization

Shikai Fang, Zheng Wang, Zhimeng Pan, Ji Liu, Shandian Zhe

Presenter: Shikai Fang

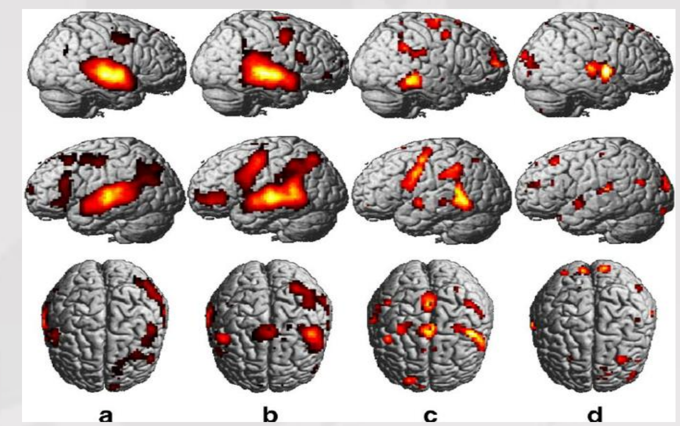
The University of Utah

For ICML 2021

Tensor Data: Widely Used High-Order Data Structures to Represent Interactions of Multiple Objects/Entities

(user, movie, episode)

(user, advertisement, page-section)



(subject, voxel, electrode)

(user, item, online-store)

(user, user, location, message-type)

(patient, gene, condition)

THE UNIVERSITY OF UTAH

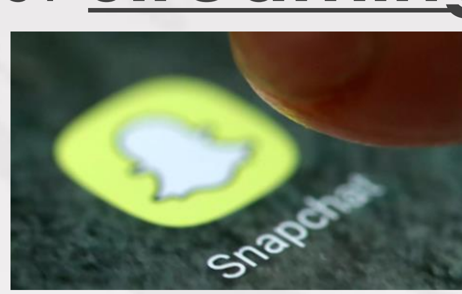
Tensor Decomposition

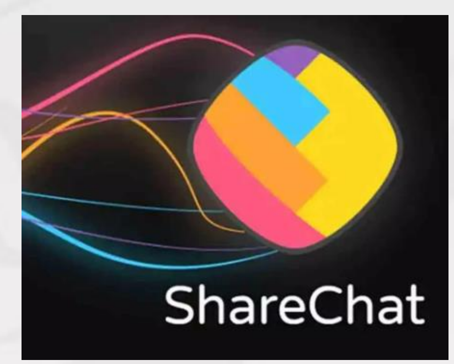
- Traditional methods: **Oversimplified multilinear assumptions**

$$\mathcal{Y}(i_1, i_2, i_3) = \sum_{j=1}^{r} \alpha_j \prod_k u^1_{i_1} u^2_{i_2} u^3_{i_3} \qquad \text{Tucker}$$

$$\mathcal{Y} = \mathcal{W} \times_1 \mathbf{U}_1 \times_2 \mathbf{U}_2 \times_3 \mathbf{U}_3 \qquad \text{CP}$$

- Hard to handle fast **streaming** data!

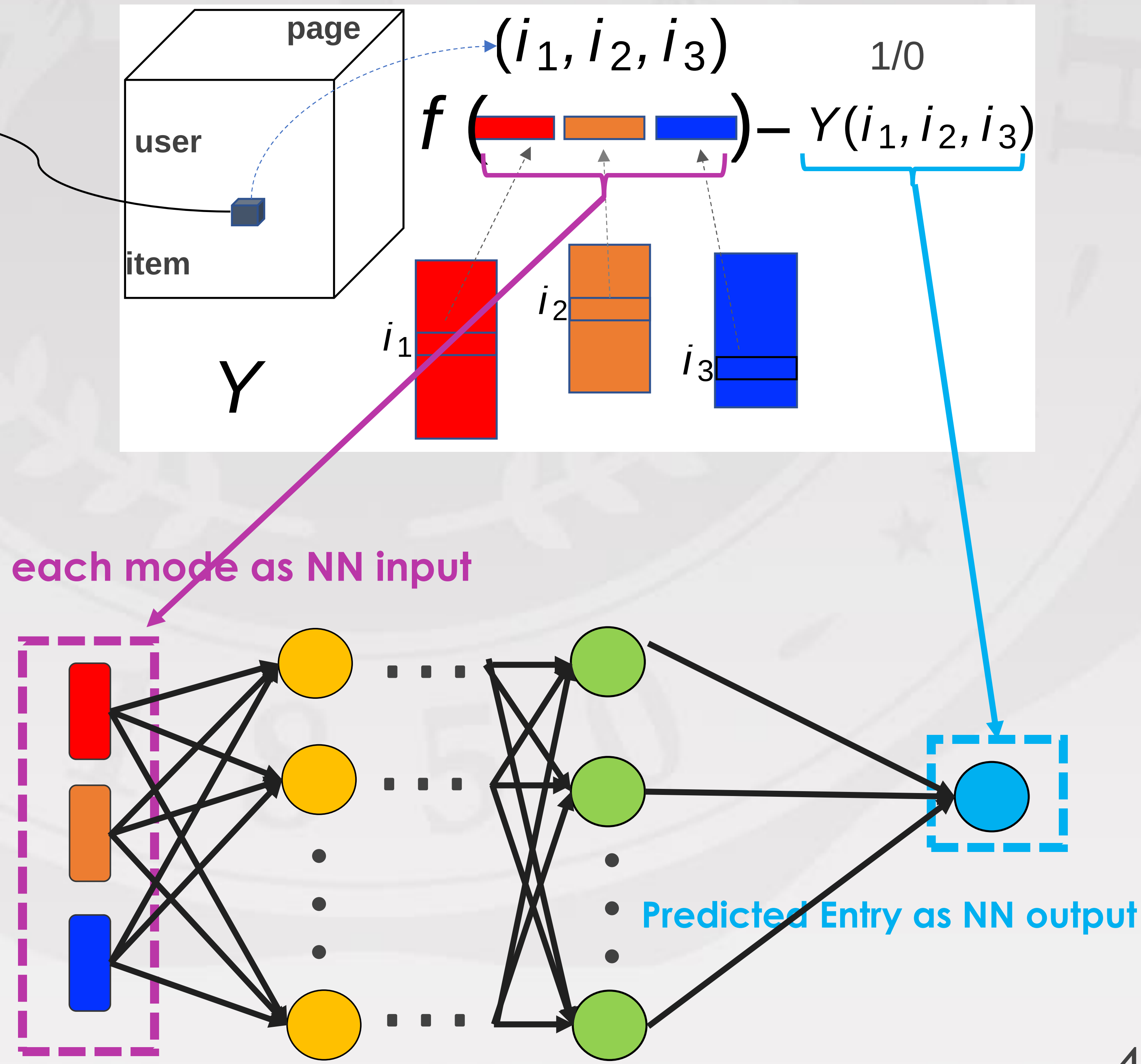




(privacy-demanding applications like Snapchat/Instagram, Data are **NOT** allowed to be stored/revisited)

Interaction Records

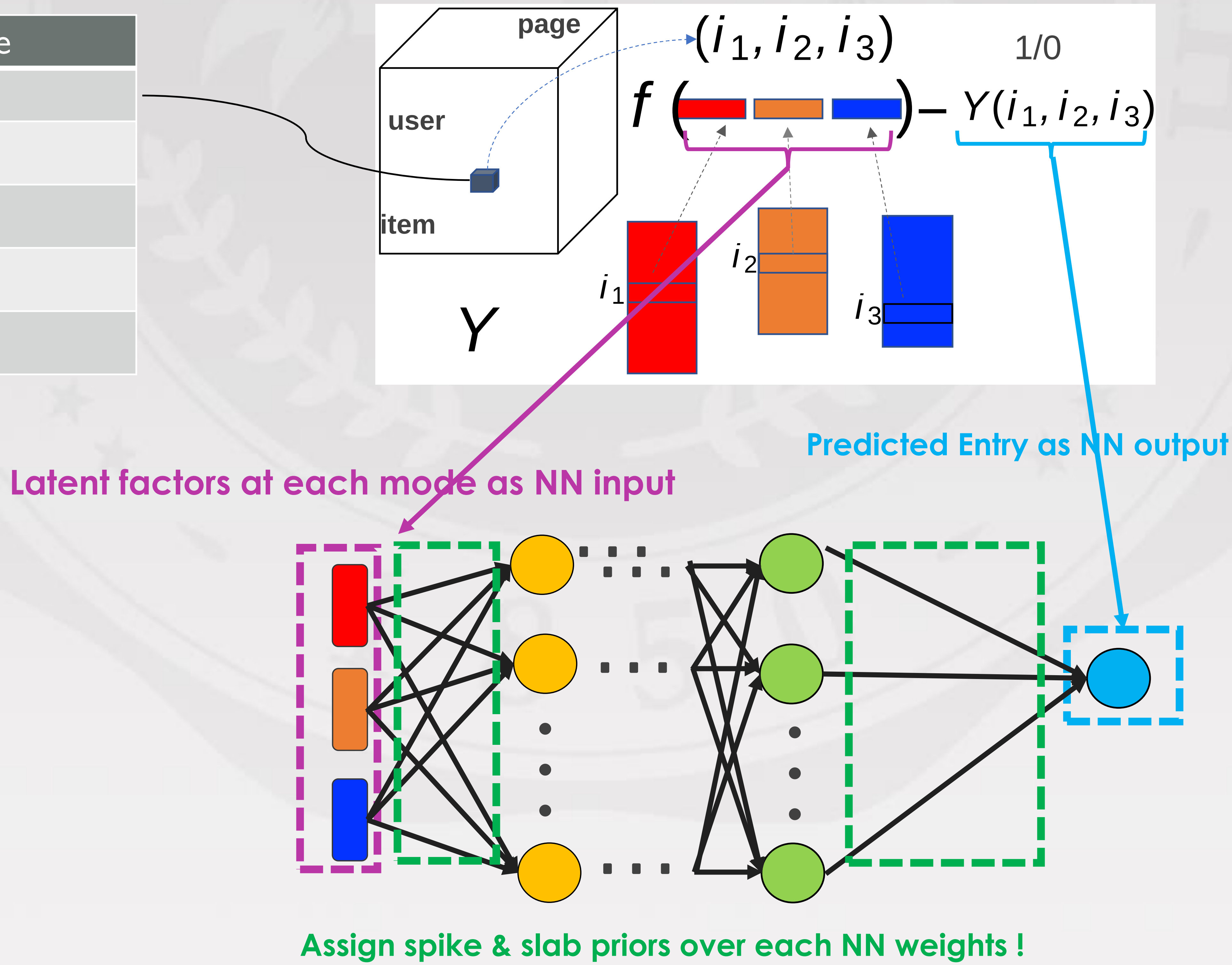
user	item	page	purchase
100	25	35	1
23	21	56	0
100	25	35	1
..	..		..
32	33	46	0



Problem: Overfitting Risk

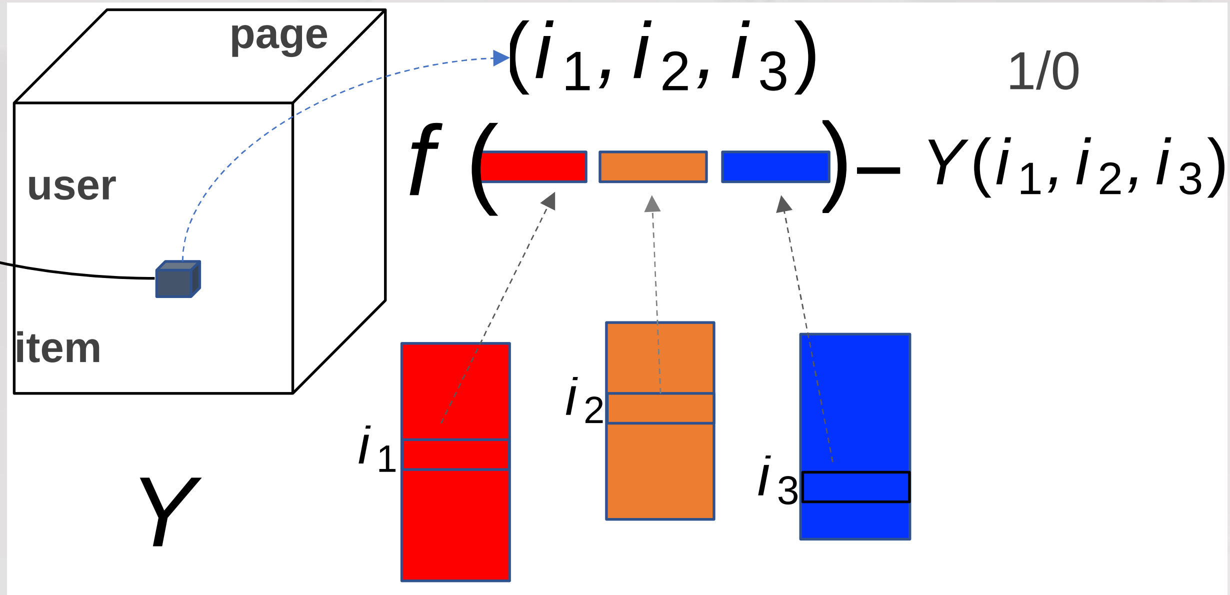
Interaction Records

user	item	page	purchase
100	25	35	1
23	21	56	0
100	25	35	1
..	..		..
32	33	46	0



Interaction Records

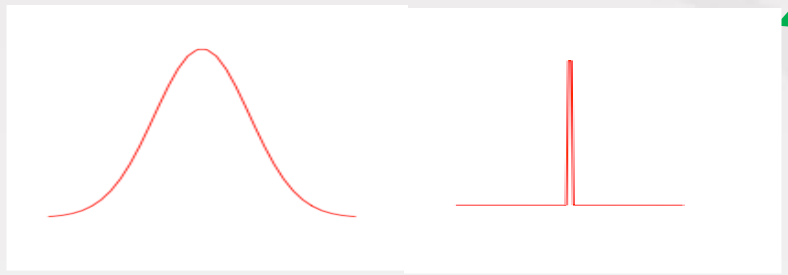
user	item	page	purchase
100	25	35	1
23	21	56	0
100	25	35	1
..	..		..
32	33	46	0



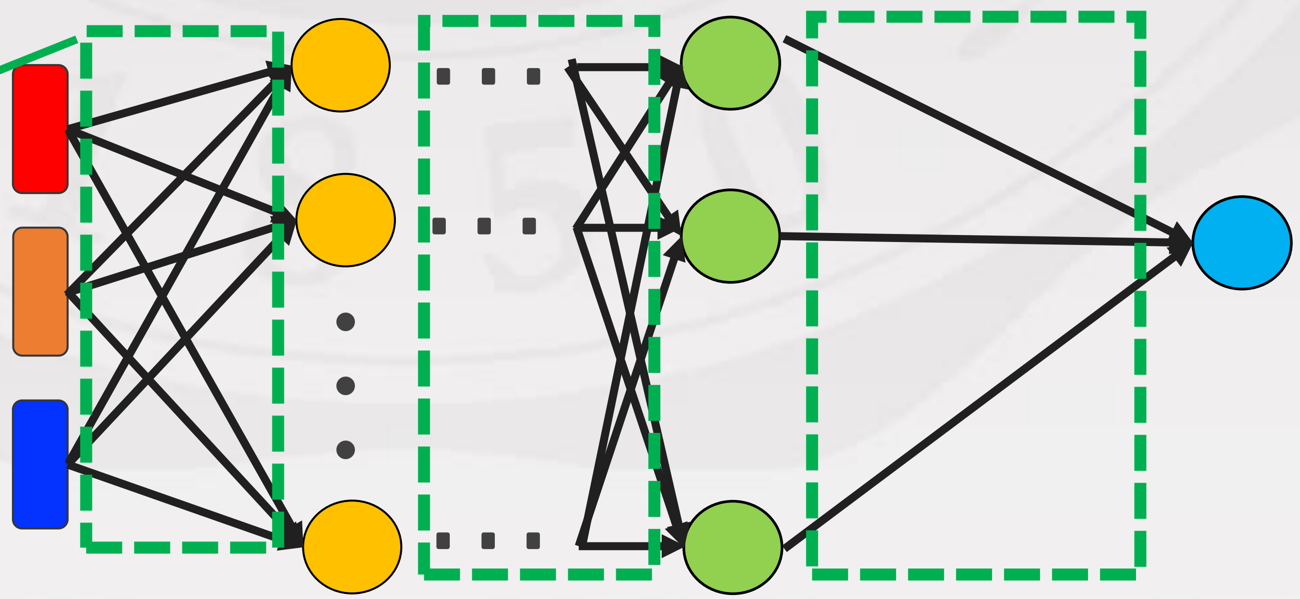
Solution: Build a sparse BNN!

w_{ij}^{n} Assign spike & slab priors over each NN weights !

$s_{\mathbf{i,j}}^{n} = 1$ $\qquad$ $s_{\mathbf{i,j}}^{n} = 0$



$s_{\mathbf{i,j}}^{n} \mathcal{N}\left(w_{\mathbf{i,j}}^{n} \mid 0, \sigma_0^2\right)$ $\qquad$ $\left(1 - s_{\mathbf{i,j}}^{n}\right) \delta\left(w_{\mathbf{i,j}}^{n}\right)$



THE UNIVERSITY OF UTAH

Online moment-match for Streaming inference

Classical ADF:

- streaming update for **BNN weights & involved factors**
- integrating entries one by one via **moment matching**

model evidence

Closed form
Posterior update
(mean and var)

$$\begin{cases} \mu^* = \mu + v\dfrac{\partial \log Z_n}{\partial \mu} \\ v^* = v - v^2\left[\left(\dfrac{\partial \log Z_n}{\partial \mu}\right)^2 - 2\dfrac{\partial \log Z_n}{\partial v}\right] \end{cases}$$

intractable

$$\boxed{Z_n} = \int q_{\text{cur}}(\mathcal{W}, \mathcal{U}, \mathcal{S}) \Phi\left((2y_{\mathbf{i}_n} - 1)\, f_{\mathcal{W}}\left(\mathbf{x}_{\mathbf{i}_n}\right)\right) \mathrm{d}\mathcal{W} \mathrm{d}\mathcal{U} \mathrm{d}\mathcal{S}$$

For tractable model evidence/analytic posterior update:

- we use **delta method**

I. Expand the BNN output at the mean of U and W (Taylor approx.)

$$f_{\mathcal{W}}\left(\mathbf{x}_{\mathbf{i}_n}\right) \approx f_{\mathbb{E}[\mathcal{W}]}\left(\mathbb{E}\left[\mathbf{x}_{\mathbf{i}_n}\right]\right)+\boxed{\mathbf{g}_n^{\top}\left(\boldsymbol{\eta}_n-\mathbb{E}\left[\boldsymbol{\eta}_n\right]\right.}\Big)$$

Eliminate when take expectation

II. Get approx. of **first & second moments** of BNN output

$$\alpha_n=\mathbb{E}_{q_{\text{cur}}}\left[f_{\mathcal{W}}\left(\mathrm{x}_{\mathrm{i}_n}\right)\right] \approx f_{\mathbb{E}[\mathcal{W}]}\left(\mathbb{E}\left[\mathrm{x}_{\mathrm{i}_n}\right]\right)$$
$$\beta_n=\operatorname{Var}_{q_{\text{cur}}}\left(f_{\mathcal{W}}\left(\mathrm{x}_{\mathrm{i}_n}\right)\right) \approx \mathrm{g}_n^{\top} \operatorname{diag}\left(\gamma_n\right) \mathrm{g}_n$$

$$\xrightarrow{\text{Moment matching}} \quad q_{\text{cur}}(\cdot)=\mathcal{N}(\cdot|\alpha_n, \beta_n)$$

III. Tractable model evidence & **analytic form update** with ADF

$$\begin{aligned} Z_n &= \mathbb{E}_{q_{\text{cur}}(\mathcal{W}, \mathcal{U}, \mathcal{S})}\left[\Phi\left(\left(2 y_{\mathrm{i}_n}-1\right) f_{\mathcal{W}}\left(\mathrm{x}_{\mathrm{i}_n}\right)\right)\right] \\ &\approx \Phi\left(\frac{\left(2 y_{\mathrm{i}_n}-1\right) \alpha_n}{\sqrt{1+\beta_n}}\right) \end{aligned}$$

THE UNIVERSITY OF UTAH

<u>**Spike & Slab priors:**</u> repeated approx. & refinement

- Approx. in exponential family: **Gaussian + Bernoulli**

$$p\left(w_{mjt} \mid s_{mjt}\right) \propto A\left(w_{mjt}, s_{mjt}\right)$$
$$= \mathrm{Bern}\left(s_{mjt} \mid \boxed{c\left(\rho_{mjt}\right)}\right) \mathcal{N}\left(w_{mjt} \mid \mu^{0}_{mjt}, v^{0}_{mjt}\right)$$

Select posterior prob of each NN weight
<0.5: unselected <=> sparse

- Update with <u>**standard EP**</u> : analytic form

- Refine it after processing all entries in a coming batch to impose the **sparse inducing effect**

THE UNIVERSITY OF UTAH

Thanks for attention

Q&A time

Presenter's email: shikai.fang@utah.edu