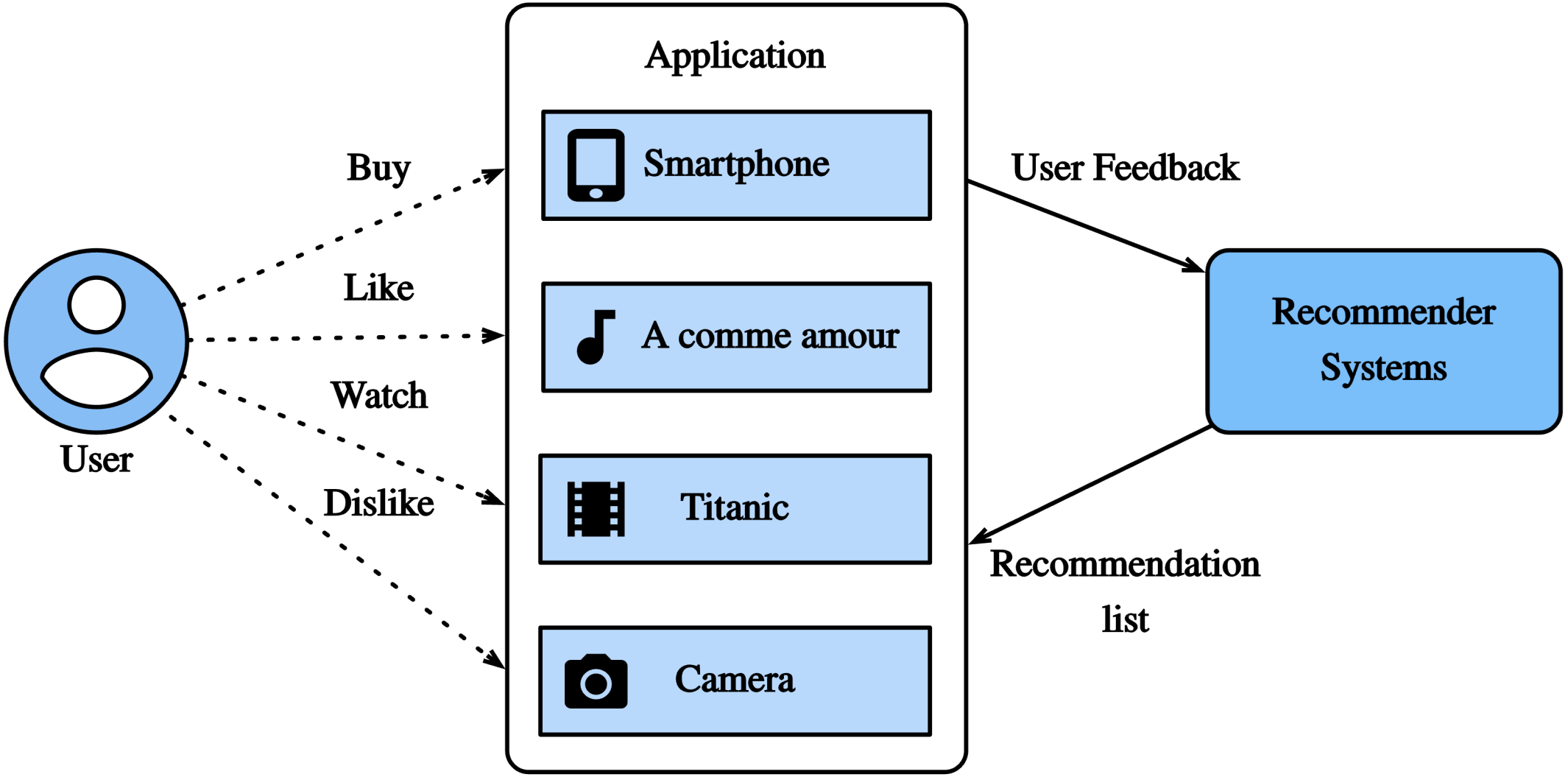
COMP 4332 / RMBI 4310
Big Data Mining (Spring 2024)

Project 3 Rating Prediction

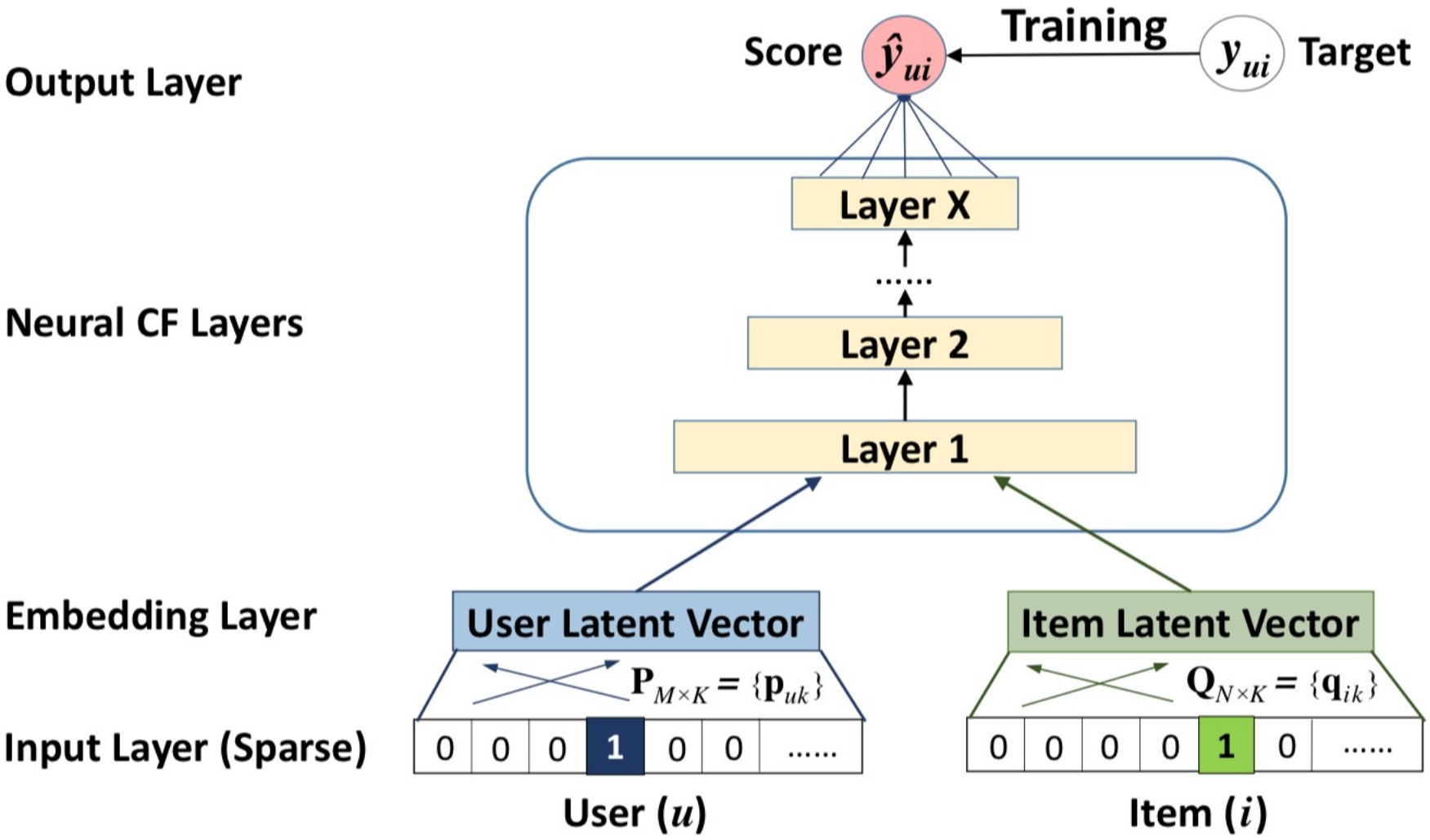
TA: DENG Zheye (zdengah@connect.ust.hk)

Recommendation Systems



In Previous Tutorial

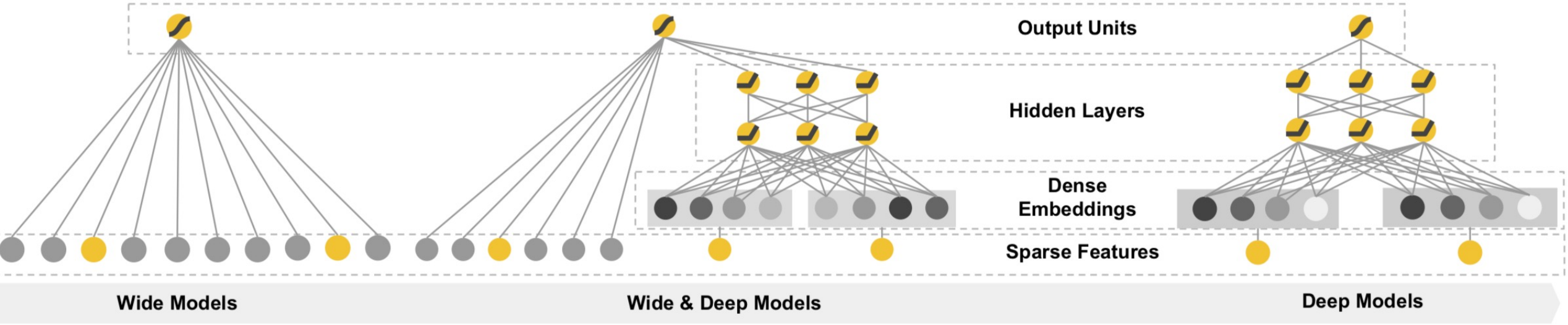
Neural CF



Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu and Tat-Seng Chua (2017). [Neural Collaborative Filtering.](#) In Proceedings of WWW '17, Perth, Australia, April 03-07, 2017.

In Previous Tutorial

Wide & Deep Learning



Memorization **Generalization**

Heng-Tze Cheng,Levent Koc,Jeremiah Harmsen,Tal Shaked,Tushar Chandra, Hrishi Aradhye, Glen Anderson, Greg Corrado, Wei Chai, Mustafa Ispir, et al. 2016. Wide & deep learning for recommender systems. In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems. ACM, 7–10.

Rating Prediction

- Predict users' ratings on items given some known ratings. The prediction would be evaluated by Root Mean Squared Error (RMSE)

	i_1	i_2	i_3	i_4	i_5	i_6
U1	4	?	3	?	5	?
U2	?	2	?	?	4	1
U3	?	?	1	?	2	5
U4	?	?	3	?	?	1
U5	1	4	?	?	2	5
U6	5	?	2	1	?	4
U7	?	2	3	?	4	5

Dataset

- User ratings
- User reviews
- Extra product information

User Reviews & Ratings

ReviewerID	ProductID	Text	Summary	Star
A1K4S4MWXI9E9M	B000FC27TA	Purchased more out of curiosity than any real ...	Not my favorite, but...	3.0
A3LF914GG87TWP	B000FC27TA	I actually received this text as an ebook, sin...	An interesting read	4.0
A1CNQTCRQ35IMM	B000FCKPG2	REVIEWER'S OPINION:\nThis was labeled as roman...	This was labeled romance but there was less ro...	2.0
ACVNKHUOX3QWU	B000GCFWXW	I liked this story although its probably not o...	Different	4.0
AU510CVD9XDG	B000GCFWXW	I have been saving the Argeneau novels for awh...	Science Fiction not Paranormal Romance	2.0

Extra product information

```python
with open("product.json", "r") as f:
    datas = json.load(f)
```

```python
datas
```

```
[{'category': ['Kindle Store', 'Kindle eBooks', 'Science Fiction & Fantasy'],
  'tech1': '',
  'description': [],
  'fit': '',
  'title': '',
  'tech2': '',
  'brand': "Visit Amazon's Elizabeth Moon Page",
  'feature': [],
  'rank': '88,963 Paid in Kindle Store (',
  'details': {'File Size:': '2199 KB',
   'Print Length:': '314 pages',
   'Publisher:': 'Del Rey (September 30, 2003)',
   'Publication Date:': 'September 30, 2003',
   'Language:': 'English',
   'ASIN:': 'B000FBJBA4',
   'Word Wise:': 'Enabled',
   'Lending:': 'Not Enabled'},
  'main_cat': 'Buy a Kindle',
  'similar_item': '',
  'date': '',
  'price': '',
  'imageURL': [],
  'imageURLHighRes': [],
  'ProductID': 'B000FBJBA4'}
```

We provide:

- Rating & Review data (rating scale is 1.0-5.0) :
  - ‘review.csv’ : 52,697 ratings & reviews
  - ‘validation.csv’ : 6,596 (reviewer, product, rating) triples
  - ‘prediction.csv’ : 6,597 (reviewer, product) pairs
    - (entries of ‘Star’ column in ‘prediction.csv’ are all set to 0.0)

- Product Information:
  - ‘product.json’: 6,734 products
    - (not guaranteed to include all products)

- Code for evaluating predictions: ‘evaluate.py’

Submission

- Predict on **test data** and fill in the results into the "Star" column (please make sure you can successfully evaluate your validation predictions on the validation data with the help of evaluate.py)
- Report (1~2 pages)
- Code (Frameworks and even programming languages are not restricted.)
- **DDL: 11:59 pm, May 26, 2023**
- Submission: Each **team leader** is required to submit the groupNo.zip file that contains prediction.csv and your team's code on canvas.
- we will check your report with your code and the RMSE.

Grading Rule

Grade	Model (80%)	Report (20%)	Baseline (RMSE on test set)
60%		submission	1.00
80%	an easy baseline that most students can outperform	detailed explanation	0.90
90%	a competitive baseline that about half students can surpass	detailed explanation and analysis	0.88
100%	a very competitive baseline	excellent visualization and analysis	0.86

Thank You