Web/App Intelligence Part II: Unsupervised Machine Learning

Shan-Hung Wu
shwu@cs.nthu.edu.tw

Department of Computer Science,
National Tsing Hua University, Taiwan

Software Design & Studio

Outline

Outline

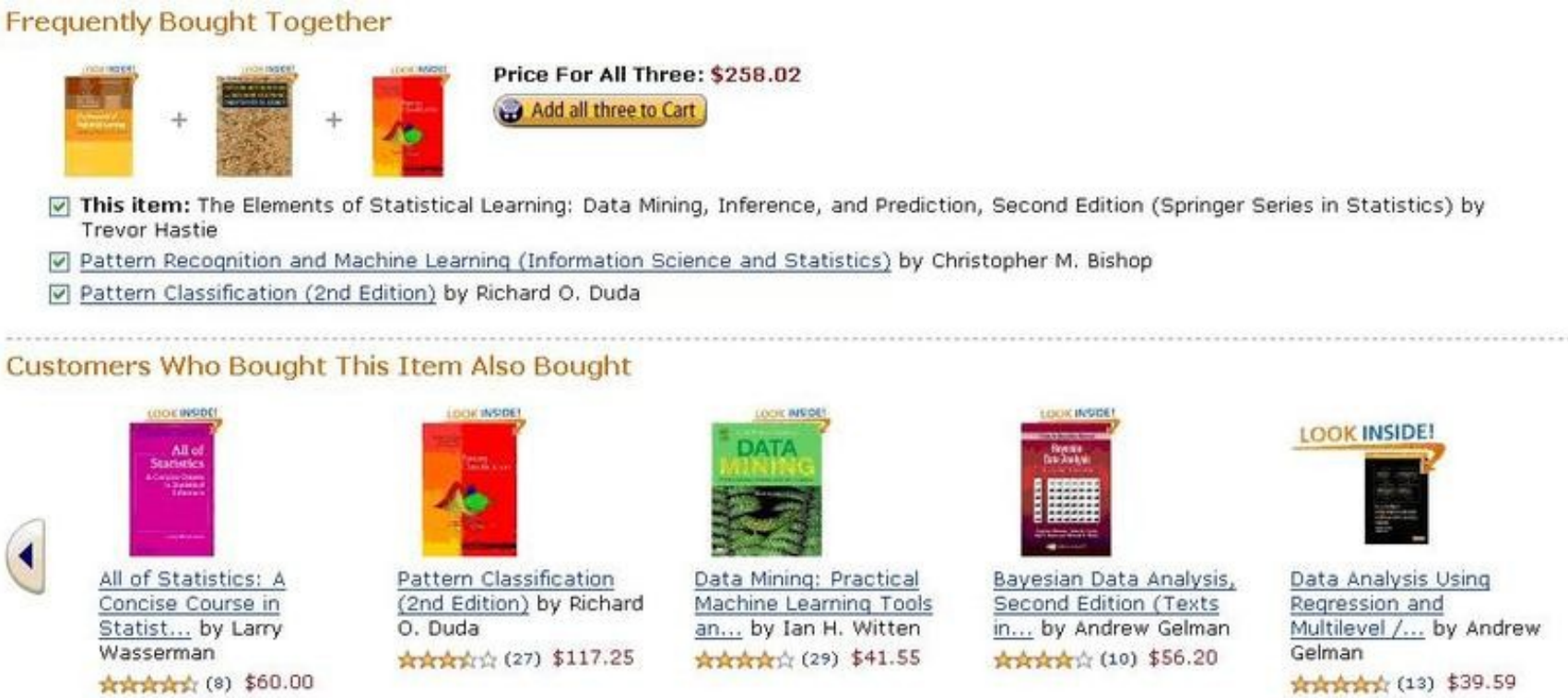
Unsupervised Learning

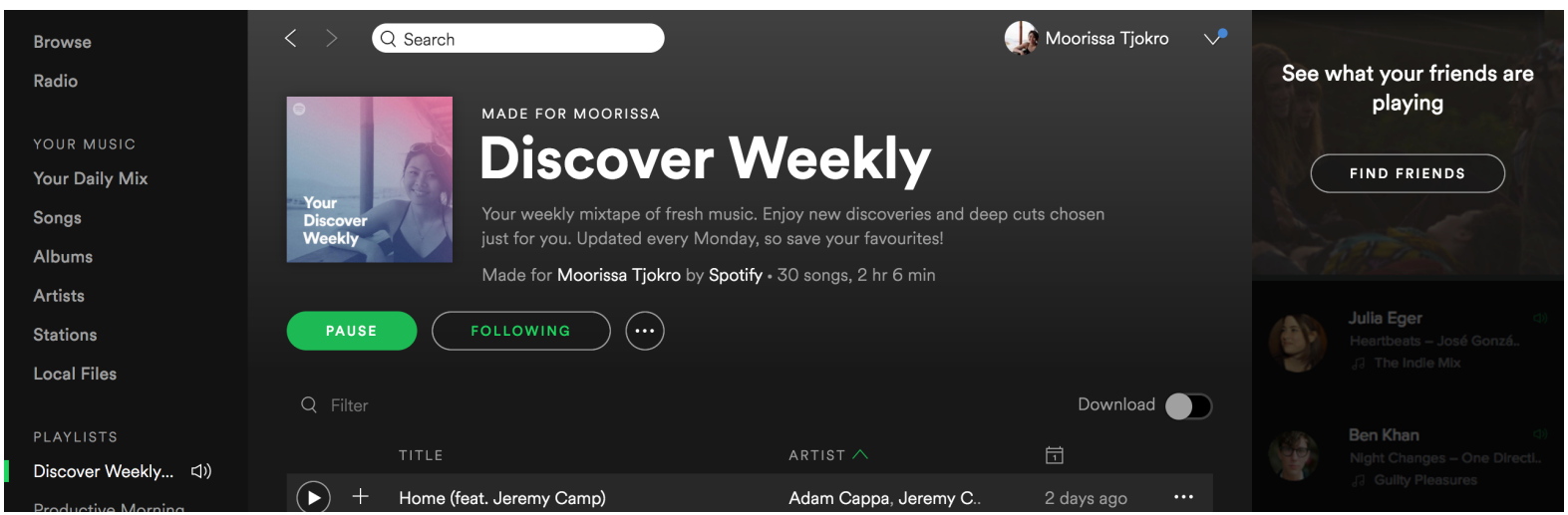
- Dataset: $\mathbb{X} = \{\boldsymbol{x}^{(i)}\}_i$
  - No supervision such as labels

Unsupervised Learning

- Dataset: $\mathbb{X} = \{\boldsymbol{x}^{(i)}\}_i$
  - No supervision such as labels
- What can we learn?

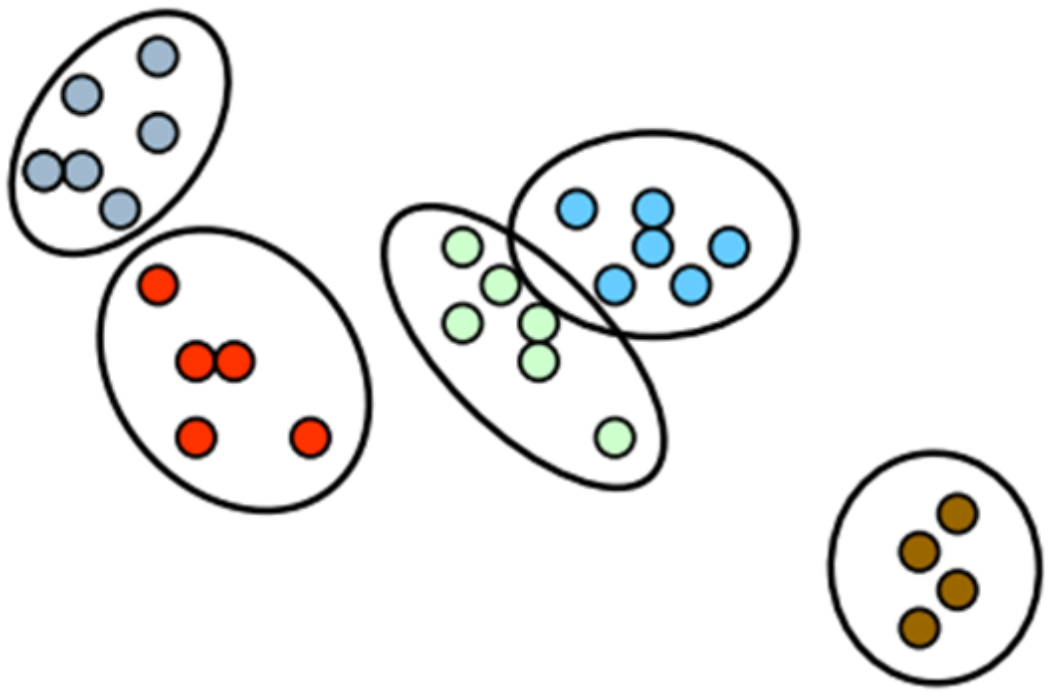
Product Recommendations





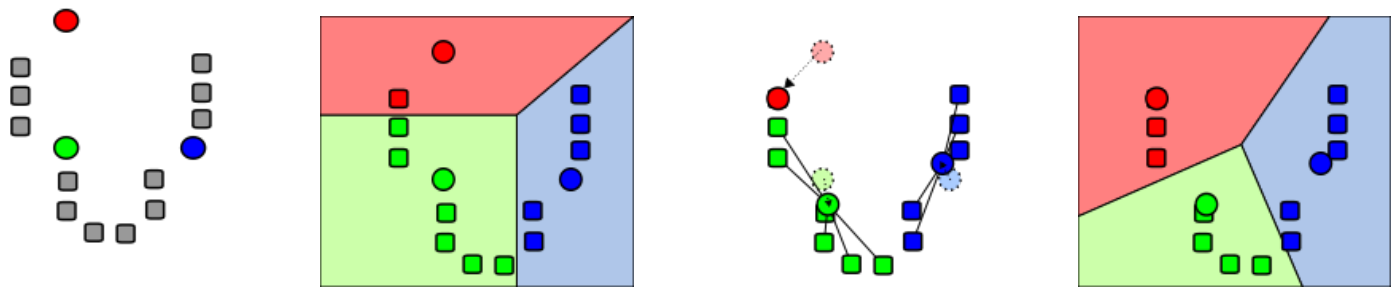
Clustering I

- Goal: to divide $\boldsymbol{x}^{(i)}$'s into K groups/***clusters***
- Based on some pairwise similarity/distance measure
  - E.g. $\cos(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)})$ or $\|\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)}\|$



Clustering II

- K-means algorithm (K fixed)
- Repeat until converge:
  - Decide K cluster heads
  - Partition points in $\mathbb{X}$ based on the similarity measure

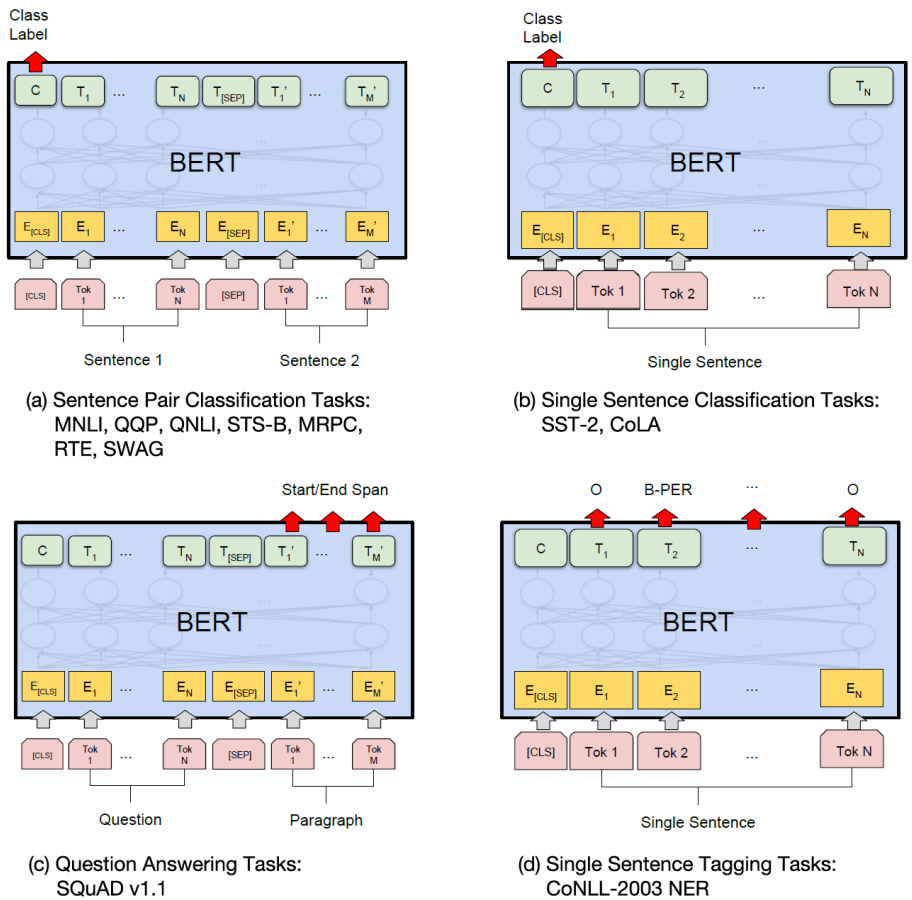


Problems

- It requires a way to transform items to vectors
- Also, the semantic similarity between items needs to be preserved
  - E..g, pop songs should be closer to R&B songs than classic music
- How?

Problems

- It requires a way to transform items to vectors
- Also, the semantic similarity between items needs to be preserved
  - E..g, pop songs should be closer to R&B songs than classic music
- How? ***Unsupervised deep machine learning*** like BERT:



(a) Sentence Pair Classification Tasks: MNLI, QQP, QNLI, STS-B, MRPC, RTE, SWAG

(b) Single Sentence Classification Tasks: SST-2, CoLA

(c) Question Answering Tasks: SQuAD v1.1

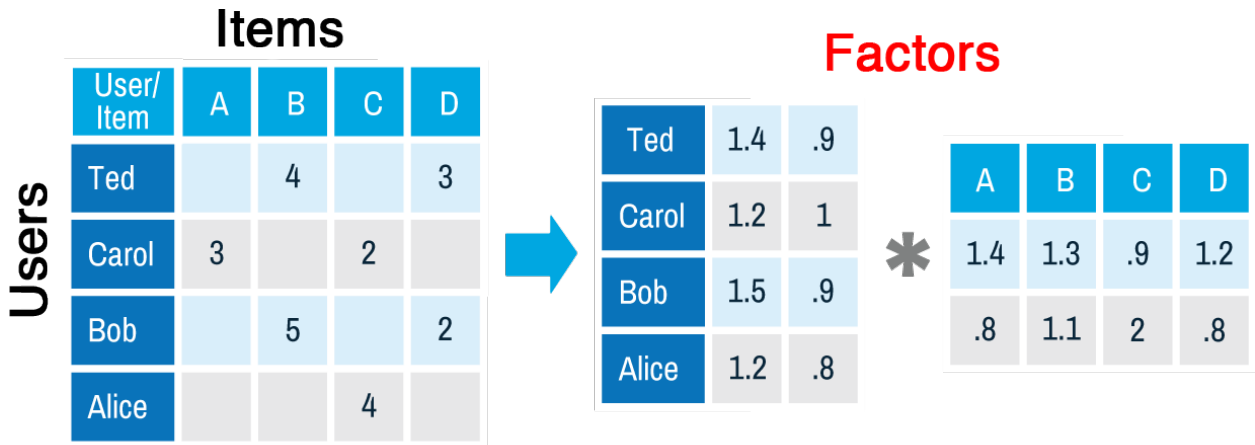
(d) Single Sentence Tagging Tasks: CoNLL-2003 NER

Rating Matrix Factorization

- Goal: to uncover the ***factors*** behind $\mathbb{X}$
  - Useful when each $\boldsymbol{x}^{(i)}$ represents item clicks

Rating Matrix Factorization

- Goal: to uncover the ***factors*** behind $\mathbb{X}$
  - Useful when each $\boldsymbol{x}^{(i)}$ represents item clicks
- Let $\boldsymbol{X}$, $\boldsymbol{X}_{i,:} = \boldsymbol{x}^{(i)}$, be a rating matrix

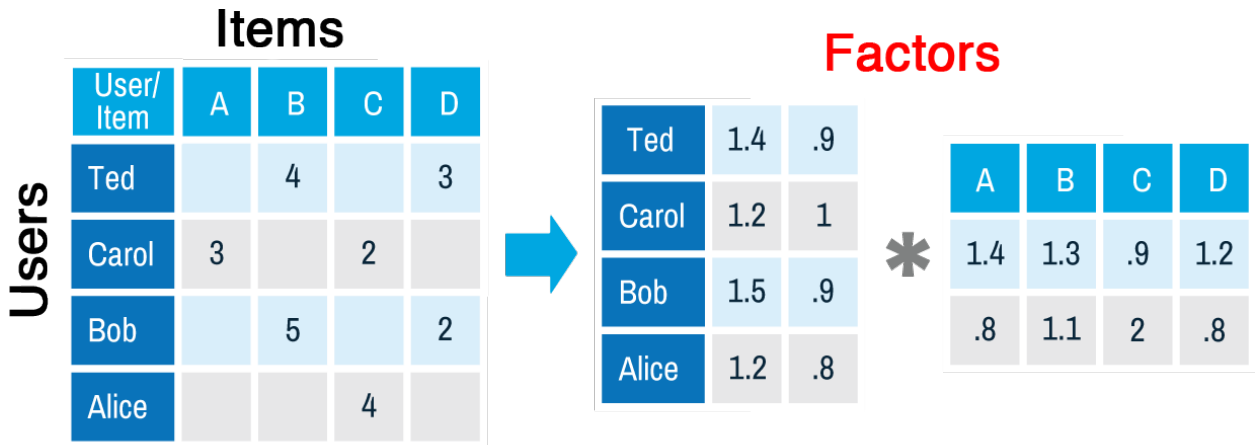


- Non-negative matrix factorization (NMF):

$$\arg \min_{\boldsymbol{W} \geq \boldsymbol{O}, \boldsymbol{H} \geq \boldsymbol{O}} \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}\|_F$$

Rating Matrix Factorization

- Goal: to uncover the ***factors*** behind $\mathbb{X}$
  - Useful when each $\boldsymbol{x}^{(i)}$ represents item clicks
- Let $\boldsymbol{X}$, $\boldsymbol{X}_{i,:} = \boldsymbol{x}^{(i)}$, be a rating matrix



- Non-negative matrix factorization (NMF):

$$\arg \min_{\boldsymbol{W} \geq \boldsymbol{O}, \boldsymbol{H} \geq \boldsymbol{O}} \|\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}\|_F$$

- $\boldsymbol{X}^* = \boldsymbol{W}^*\boldsymbol{H}^*$ a dense matrix and can be used to predict user interests

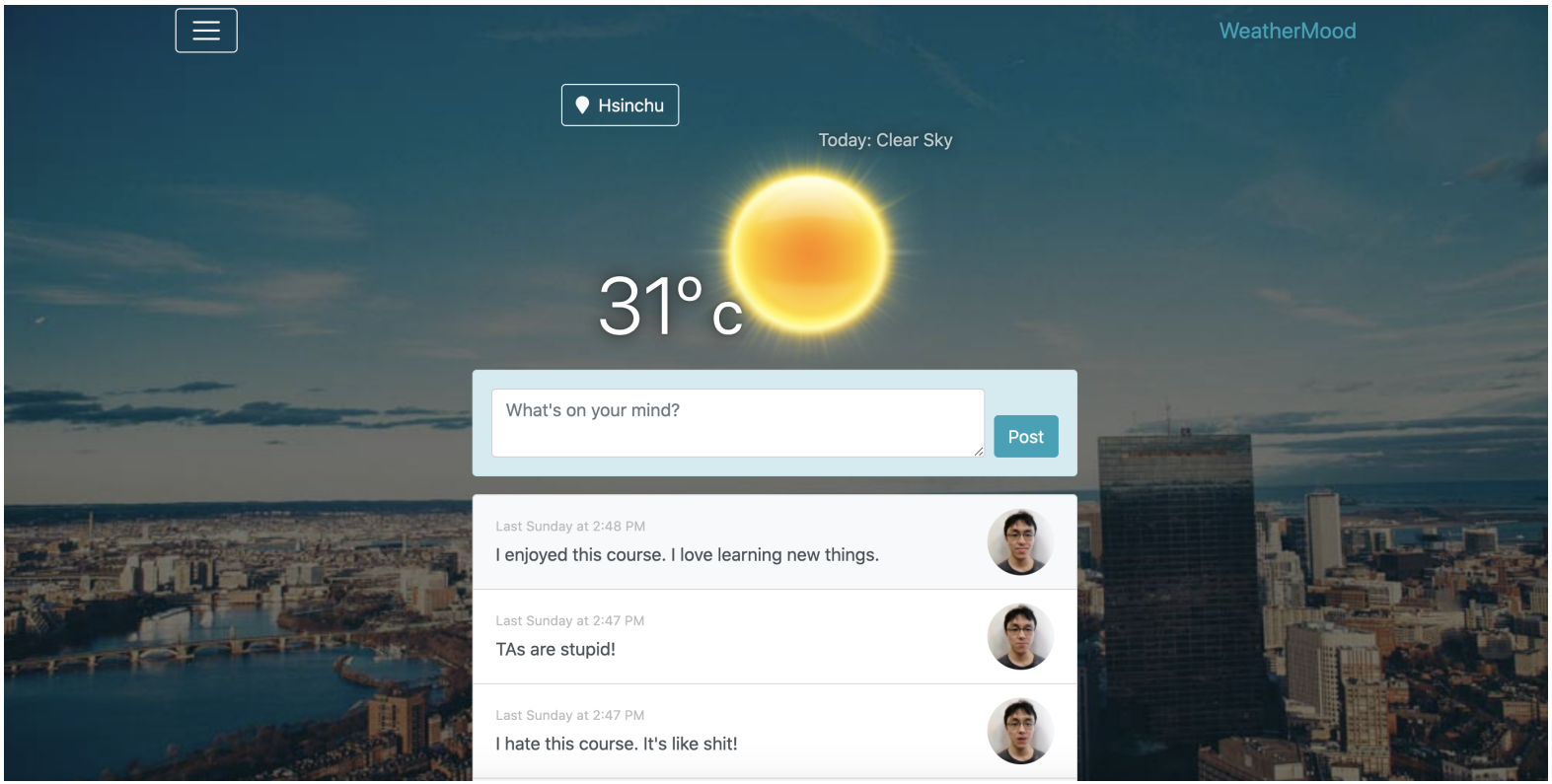
Outline

Generative Models

- Identify the patterns/structures within existing data to generate new content
- Common models:
  - ChatGPT
  - Midjourney

What Can It Do to WeatherMood?

```
$ git clone weathermood-stargan
$ npm install
$ npm run start
```



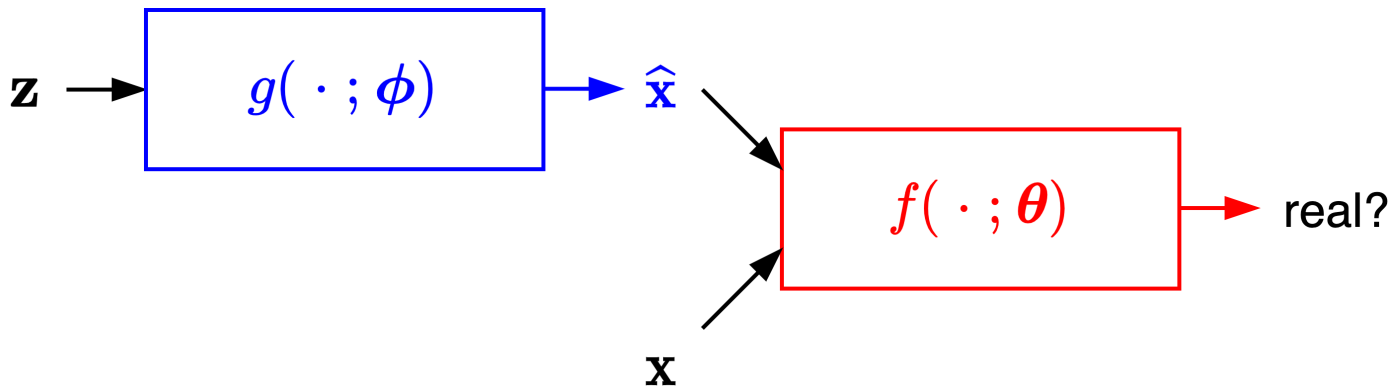
How to do it?

How to do it?

Generative Adversarial Networks

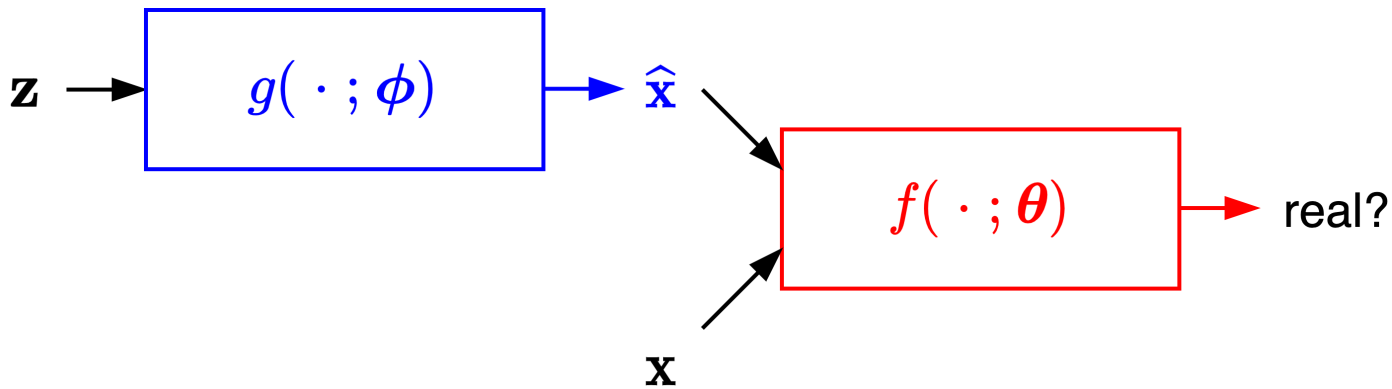
Basic Idea of GANs

- ***Generator*** g: to generate data points from random codes
- ***Discriminator*** f: to distinguish generated points from real ones in $\mathbb{X}$



Basic Idea of GANs

- ***Generator*** g: to generate data points from random codes
- ***Discriminator*** f: to distinguish generated points from real ones in $\mathbb{X}$

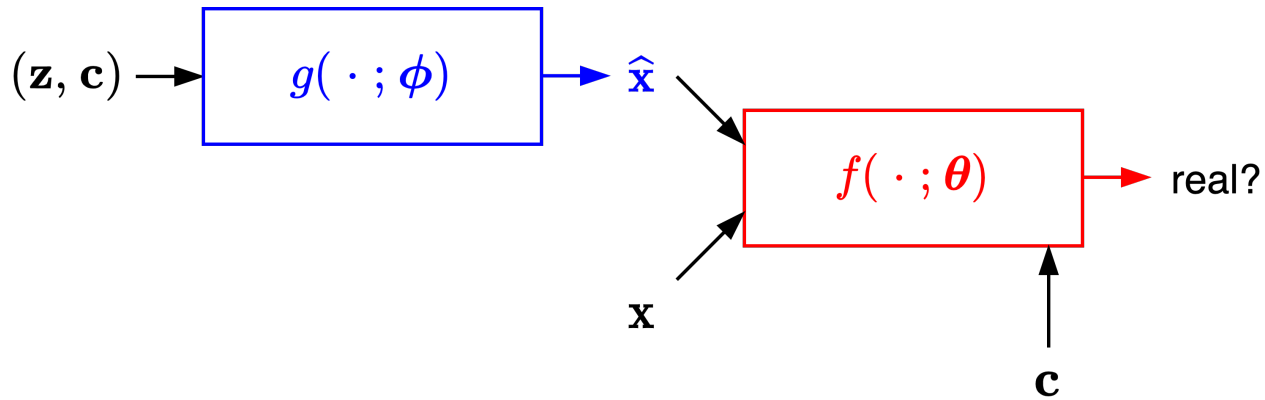


- After training: user g to generate images

Results

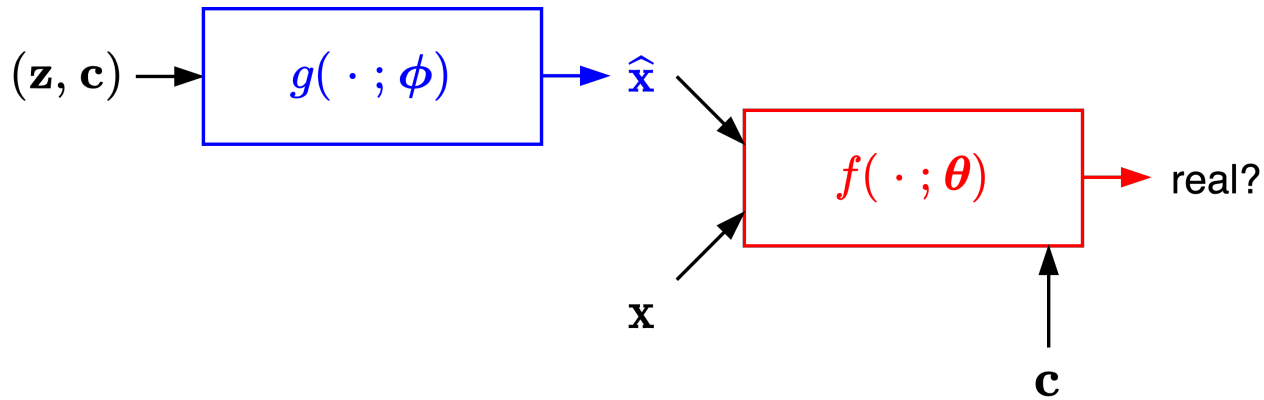
Conditional GAN

- Let g and f take a ***condition vector*** as extra input



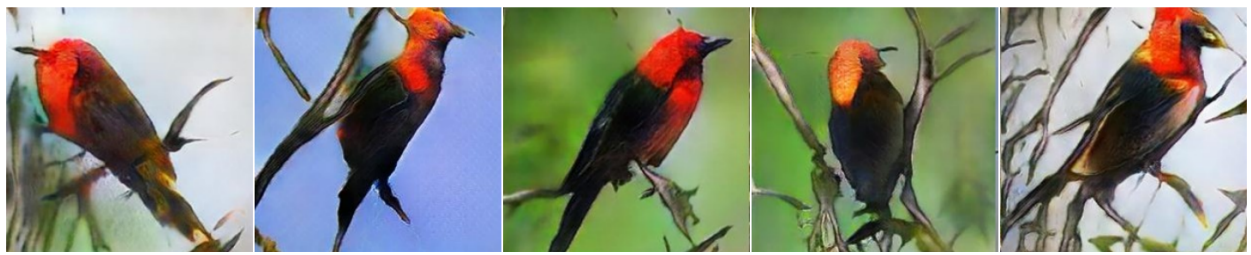
Conditional GAN

- Let g and f take a ***condition vector*** as extra input

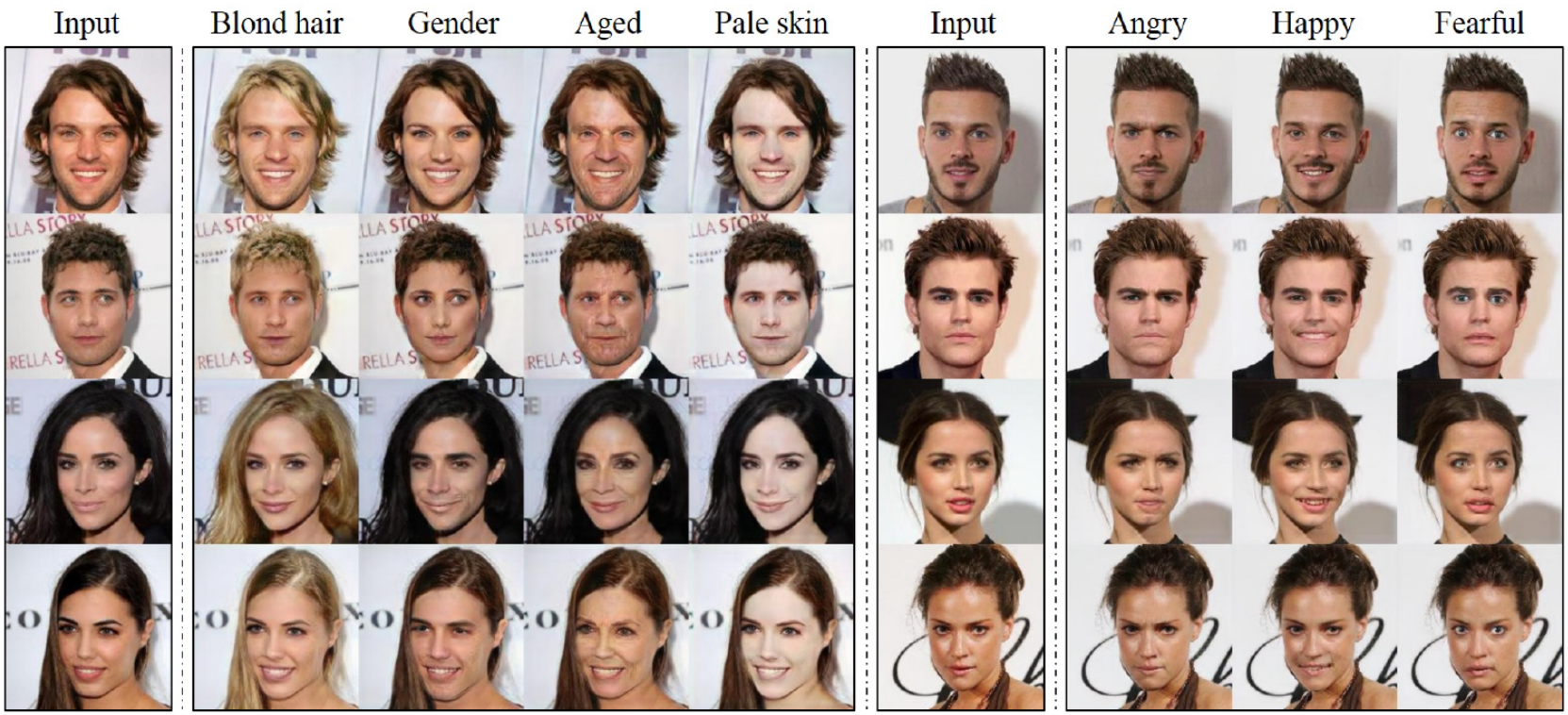


- Example: text as condition

  "*This bird is completely red with black wings and pointy beak.*"



StarGAN (for Face Generation)



Good Luck for Your Final Demo!