

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

1 Unsupervised Learning

2 Generative Models

Outline

1 **Unsupervised Learning**

2 Generative Models

Unsupervised Learning

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

Unsupervised Learning

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

Product Recommendations

Frequently Bought Together



Price For All Three: \$258.02

Add all three to Cart

- ☒ **This item:** The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics) by Trevor Hastie
- ☒ [Pattern Recognition and Machine Learning \(Information Science and Statistics\)](#) by Christopher M. Bishop
- ☒ [Pattern Classification \(2nd Edition\)](#) by Richard O. Duda

Customers Who Bought This Item Also Bought



[All of Statistics: A Concise Course in Statist...](#) by Larry Wasserman
★★★★☆ (8) \$60.00



[Pattern Classification \(2nd Edition\)](#) by Richard O. Duda
★★★★☆ (27) \$117.25



[Data Mining: Practical Machine Learning Tools an...](#) by Ian H. Witten
★★★★☆ (29) \$41.55



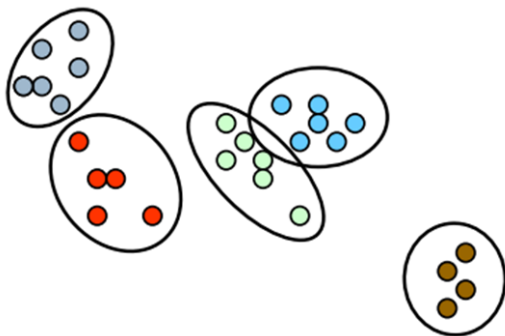
[Bayesian Data Analysis, Second Edition \(Texts in...](#) by Andrew Gelman
★★★★☆ (10) \$56.20



[Data Analysis Using Regression and Multilevel /...](#) by Andrew Gelman
★★★★☆ (13) \$39.59

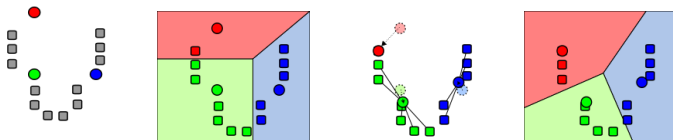
Clustering I

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



Clustering II

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

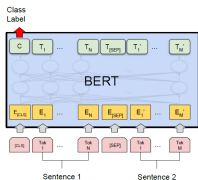


Problems

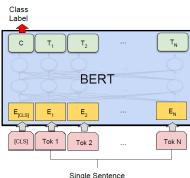
- It requires a way to transform items to vectors
- Also, the semantic similarity between items needs to be preserved
 - E.g, pop songs should be closer to R&B songs than classic music
- How?

Problems

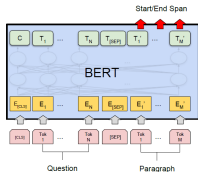
- It requires a way to transform items to vectors
- Also, the semantic similarity between items needs to be preserved
 - E.g, pop songs should be closer to R&B songs than classic music
- How? *Unsupervised deep machine learning* like BERT:



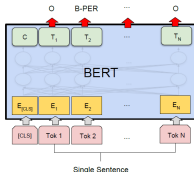
(a) Sentence Pair Classification Tasks:
MNLI, QQP, QNLI, STS-B, MRPC,
RTE, SWAG



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



(c) Question Answering Tasks:
SQuAD v1.1



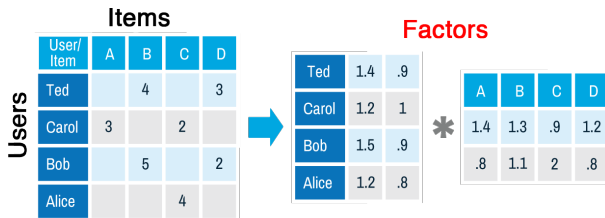
(d) Single Sentence Tagging Tasks:
CoNLL-2003 NER

Rating Matrix Factorization

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

Rating Matrix Factorization

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

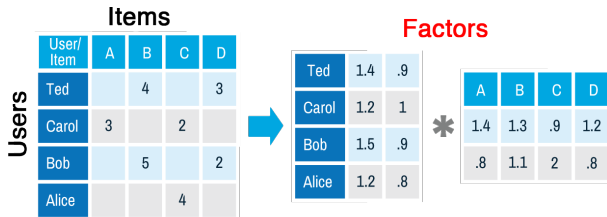


- Non-negative matrix factorization (NMF):

$$\arg \min_{\mathbf{W} \geq \mathbf{0}, \mathbf{H} \geq \mathbf{0}} \|\mathbf{X} - \mathbf{WH}\|_F$$

Rating Matrix Factorization

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



- Non-negative matrix factorization (NMF):

$$\arg \min_{\mathbf{W} \geq 0, \mathbf{H} \geq 0} \|\mathbf{X} - \mathbf{WH}\|_F$$

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

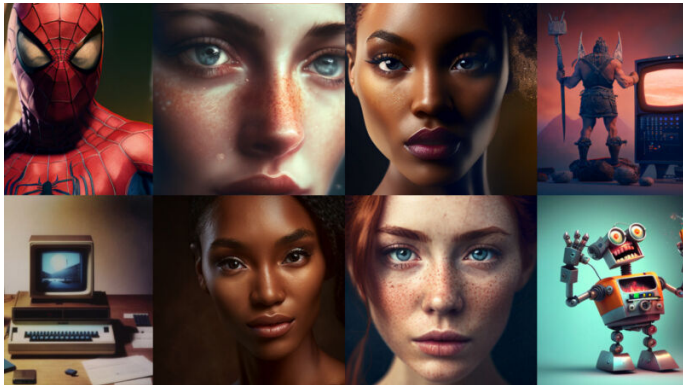
Outline

1 Unsupervised Learning

2 Generative Models

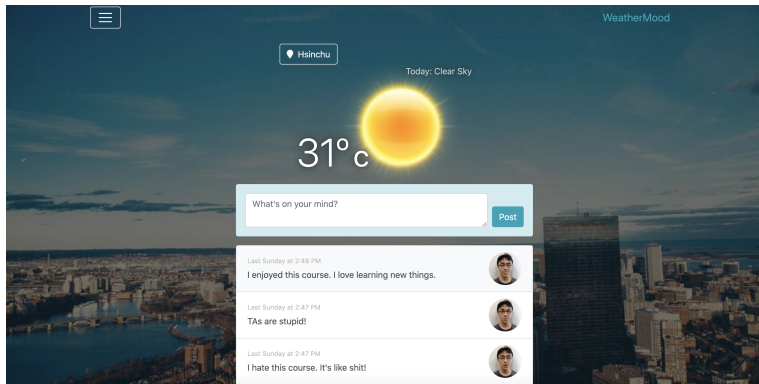
Generative Models

- 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
```



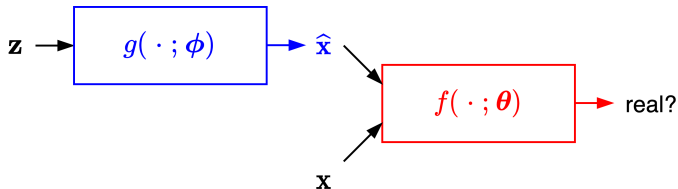
How to do it?

How to do it?

Generative Adversarial Networks

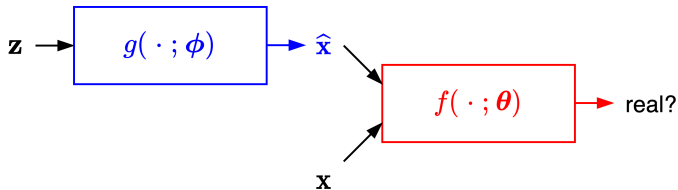
Basic Idea of GANs

- **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}



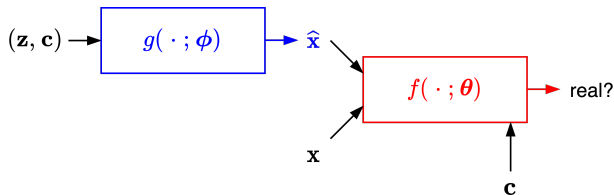
- After training: user g to generate images

Results



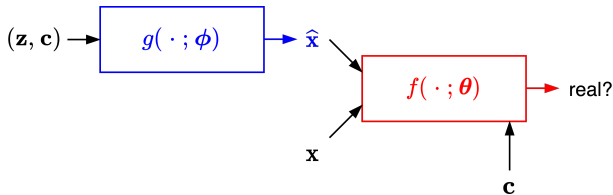
Conditional GAN

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



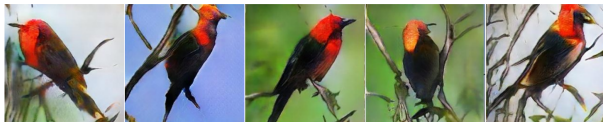
Conditional GAN

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

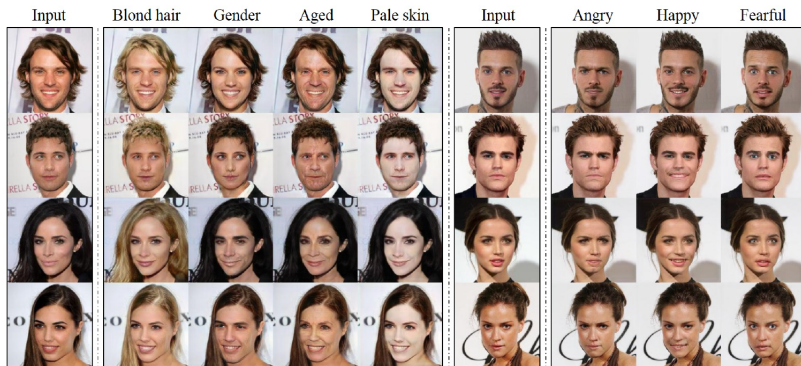


- Example: text as condition

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



StarGAN (for Face Generation)



Good Luck for Your Final Demo!