SEMINAR IN DATA SCIENCE AND INFORMATION TECHNOLOGY

ADVANCED LEARNING REPRESENTATIONS

LECTURE 1: OVERVIEW

WEEK 1-11

Instructor： Yang Li

SPRING 2024

LECTURE OUTLINE

- About this class
- A preview of the topics we will cover
- Break
- Discussion on *Representation Learning: A New Perspective*

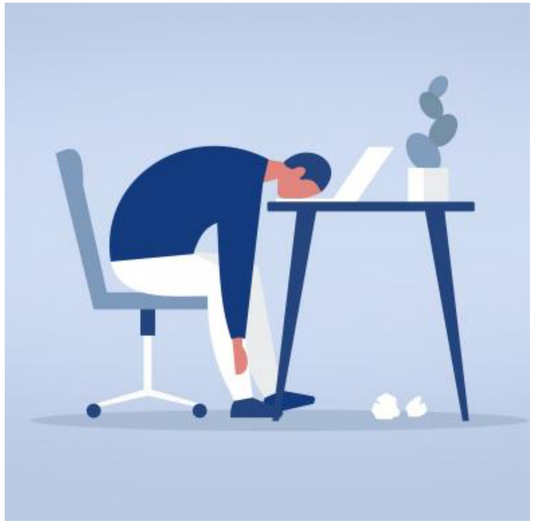
PURPOSE

- **What have we learned so far in Learning from Data:**
  - **Supervised Learning: discriminative & generative models, SVM, neural net**
  - **Unsupervised Learning: clustering & linear projection models (PCA/ICA/CCA)**
  - **Semi-Supervised Learning**
  - **Reinforcement Learning**

PURPOSE

- **What's next after Learning from Data?**

How to find "ideas" for machine learning research?



Given a new problem, how to find the best learning method?

PURPOSE

- **Learn from classic literatures**

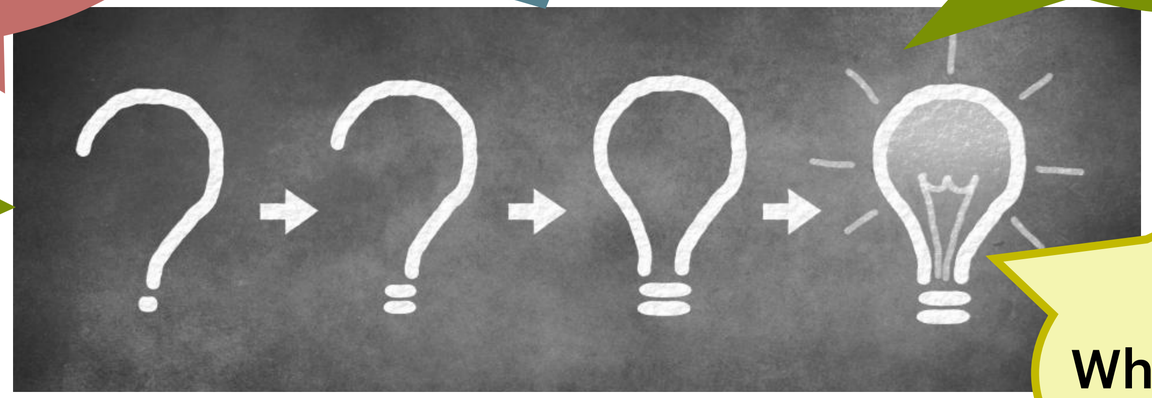
How was the problem formulated?

What priors or assumptions did it make?

How was the method derived?

How did it contribute to the field?

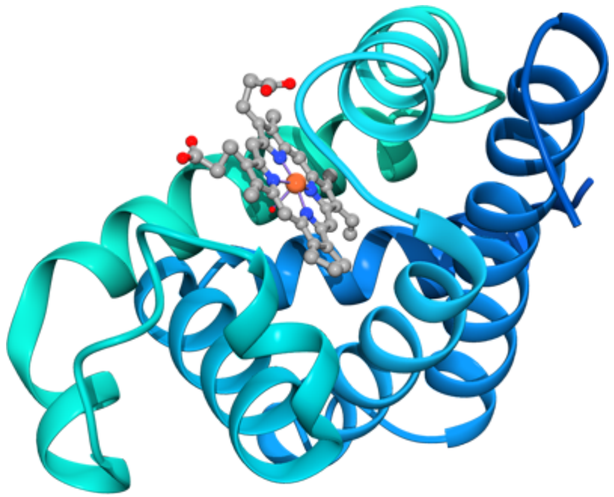
Where did it draw inspiration from?

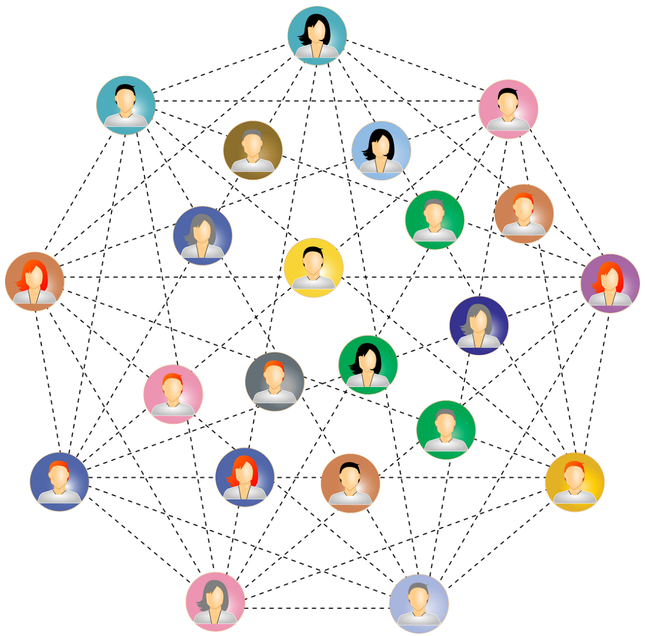


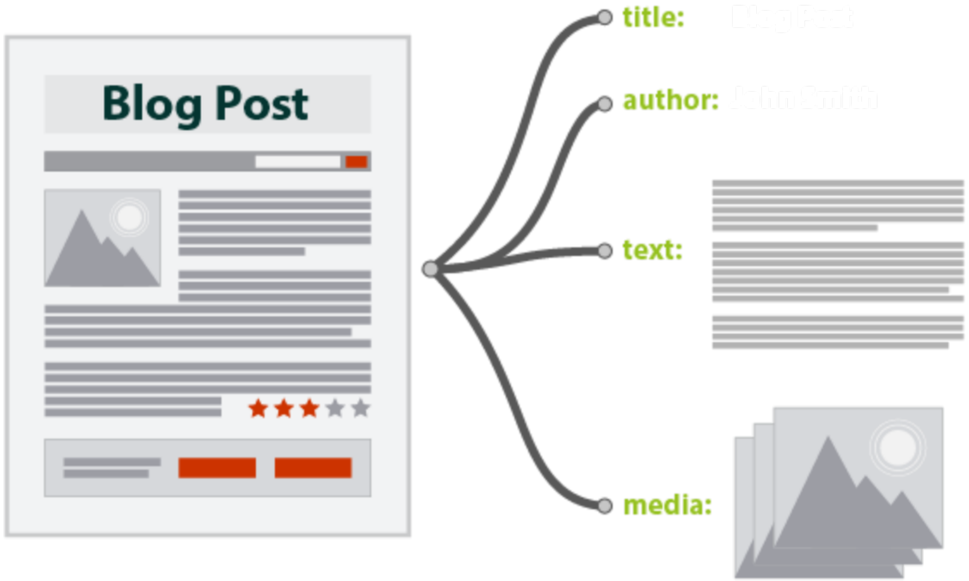
What can I learn from this work?

REPRESENTATION LEARNING

- **"Deep learning is about learning good representations"**



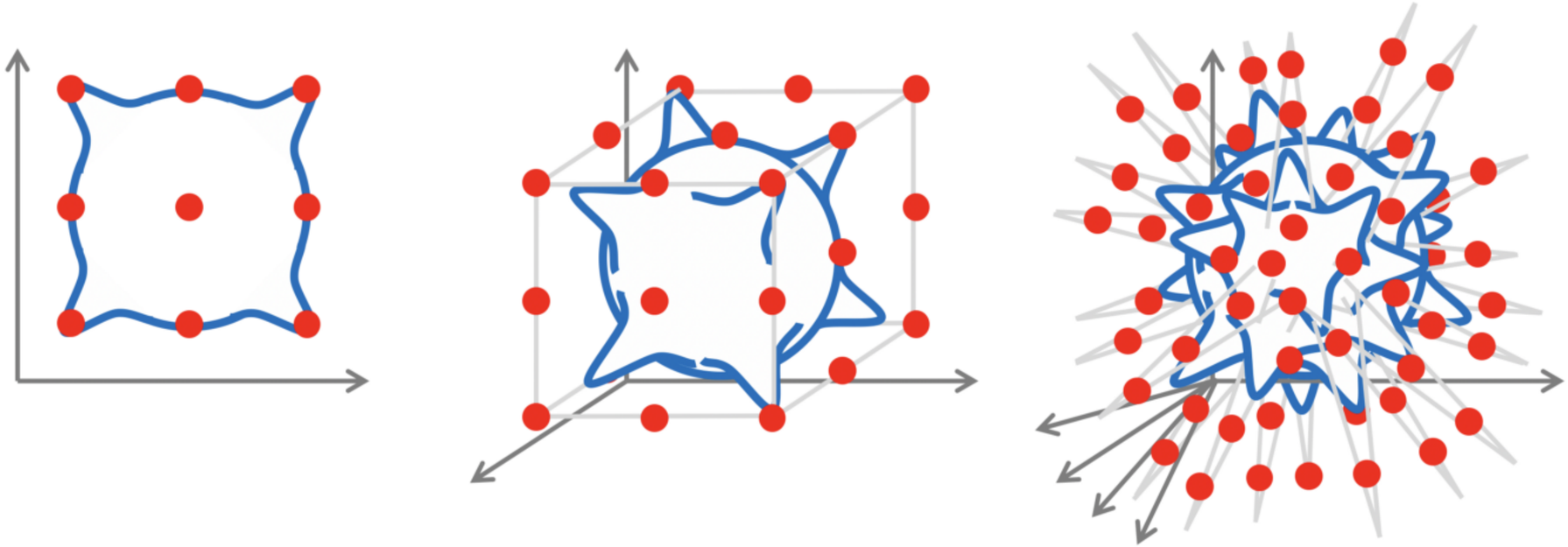




Examples of high dimensional, structured data

REPRESENTATION LEARNING

- **"Curse of dimensionality"**



Approximating a Lipschitz-continous function in R^d within error ε requires at least $\mathcal{O}(1/\varepsilon^d)$ samples

REPRESENTATION LEARNING

- This class focus on general paradigms for learning good representations
  - How to represent structured data in learning ? e.g. images, text, sequences, graphs
  - How to exploit the hidden structures and prior knowledge in data?

Learn from classic literature

SEMINAR FORMAT

- Read and discuss classic literature each week
- Role-Playing Discussion:
  - Form 8 groups
  - Each group will take a special role (e.g. discussion leader, reviewer, etc) during discussion
  - Roles rotate every week

ROLES

- **Discussion Leader:**
  - **Prepare a simple presentation to go over the paper (10 min)**
  - **Highlight its contribution and novelty**
- **Scientific Reviewer:**
  - **Imagining you are reviewing this article for a top machine learning/AI conference (e.g. ICLR/NIPS/AAAI/CVPR) etc**
  - **Score the paper (1-7) and give reasons**

ROLES

- **Archeologist:**
  - **Where this paper sits in the context of previous and subsequent work?**
  - **Discuss important prior works and new papers citing this work** (one per person)
- **Academic Researcher:**
  - **Propose an imaginary follow-up project, that is only possible due to the success of the current paper**

ROLES

- Industry Practitioner
  - As an application developer, why would you be paid to implement this method?
  - Discuss one positive /negative impact of this application
- Hacker
  - Create a demo： Reimplement a small part or simplified version on a small dataset, with documentation
  - Give credits to any code from existing implementation for “backbone” code
  - What are some implementation suggestions？

ROLES

- **Private Investigator**
  - **Background check on one of the paper's author and team**
  - **Research their past work, research interests and potential motivation**
- **Social Impact Assessor**
  - **Does the paper self-assess (discuss) its impact on the world?**
  - **What are some positive or negative social impact overlooked?**

CLASS PROJECT

- WikiBook Project on Representation Learning
  - Each group will write an online self-study guide for the presented topic
  - Be user-friendly： include background, examples, derivations and illustrations
  - Use creativity！
- Bonus programming exercises

GRADING

- **Homework: 60%**
- **Seminar participation: 20%**
- **Final presentation & wiki-book: 20%**

CLASS LOGISTICS

- Join class WeChat group
- Class website
- Contact TA if you need to use a GPU server for homework

http://yangli-feasibility.com/home/classes/dss2024spring/

Team

Yang Li
Instructor
yangli@sz.tsinghua.edu.cn

Tao Shi
TA
shitao21@mails.tsinghua.edu.cn

Office hours

Name	Time	Location
Yang	Friday 2:00-4:00pm	Info Building 1108a
Tao	Monday 2:00-4:00pm	Info Building, 1111A

You can also make appointments outside office hours.

LECTURE OUTLINE

- About this class
- A preview of the topics we will cover
- Break
- Discussion on *Representation Learning: A New Perspective*

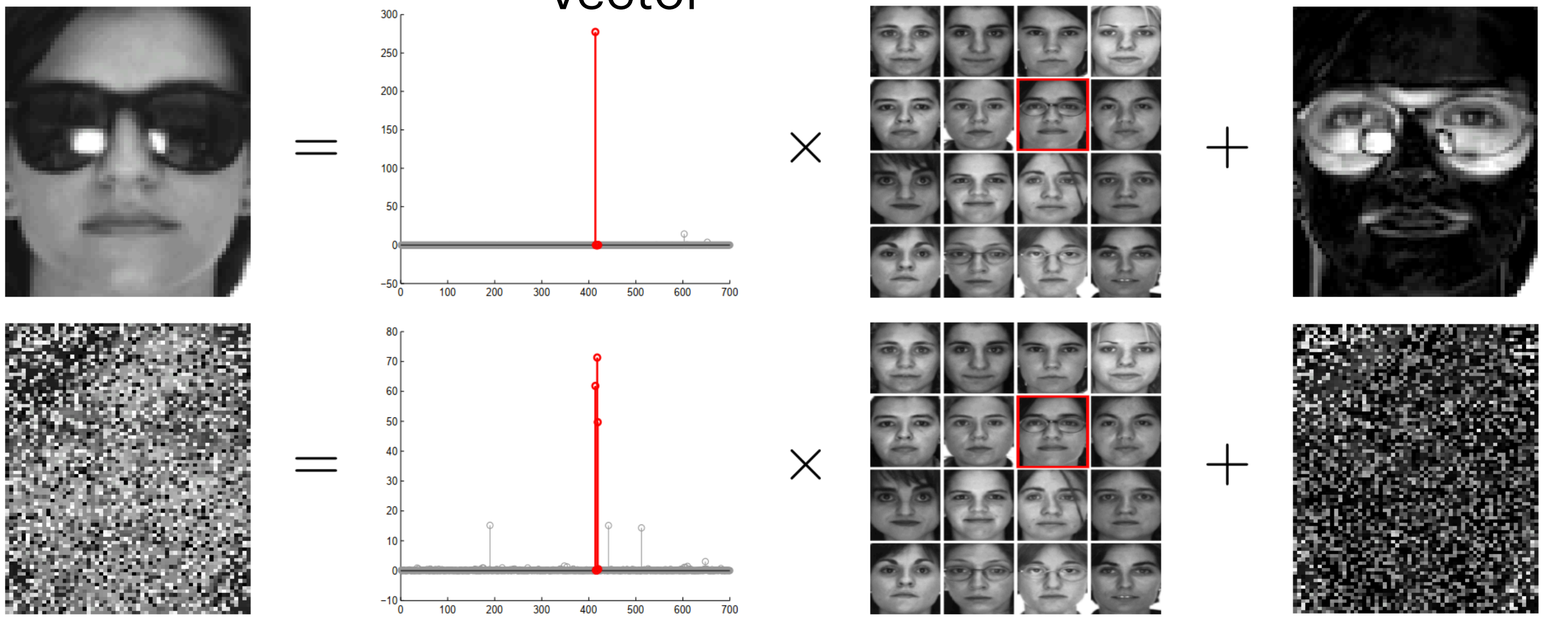
MAIN TOPICS

1. **Manifold learning**
2. **Generative modeling**
3. **Metric learning**
4. **Attention mechanism**
5. **Graph representation learning**

1. SPARSE REPRESENTATION

- **High dimensional data (e.g. images, sequences) contains regular structure**
  - **can be represented using a sparse representation given a basis**

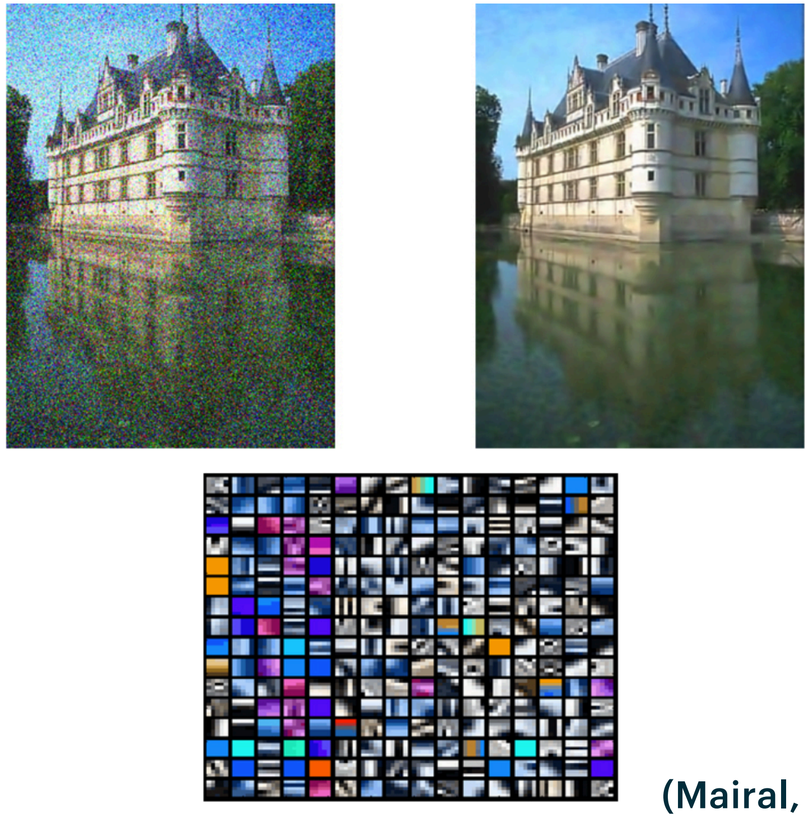
occluded/ corrupted images | sparse vector | basis | noise



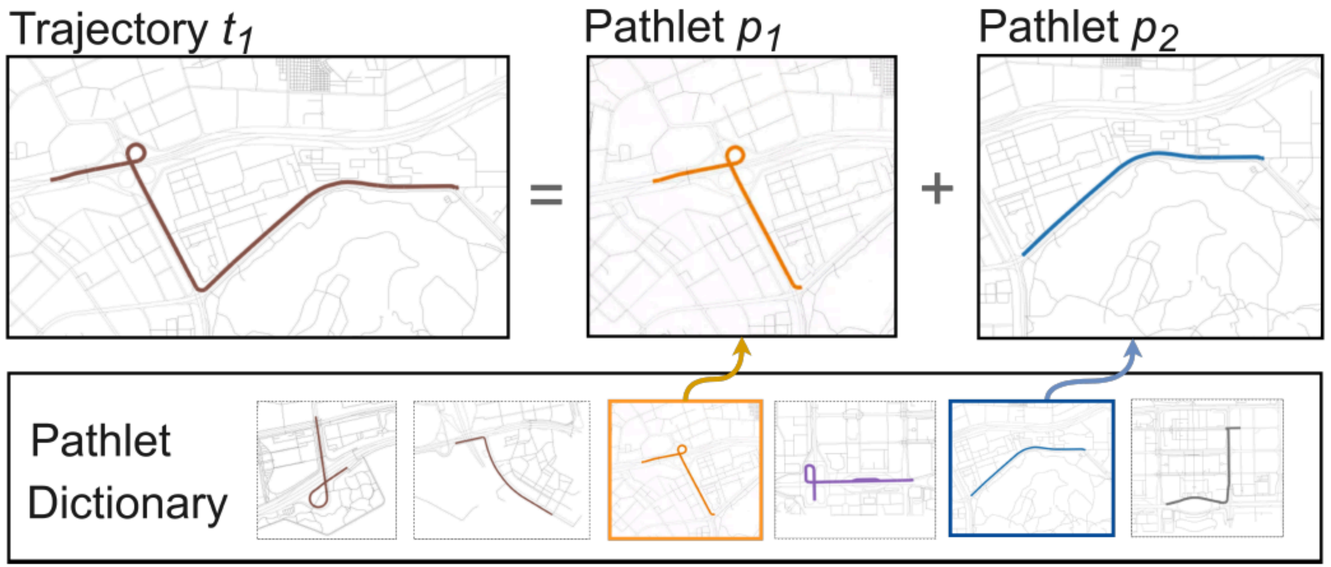
Sparse representation: naturally encode the semantic information

1. SPARSE REPRESENTATION

- **Dictionary Learning: find an optimal basis (basis) for a given data application**



(Mairal, Elad & Sapiro, 2008)



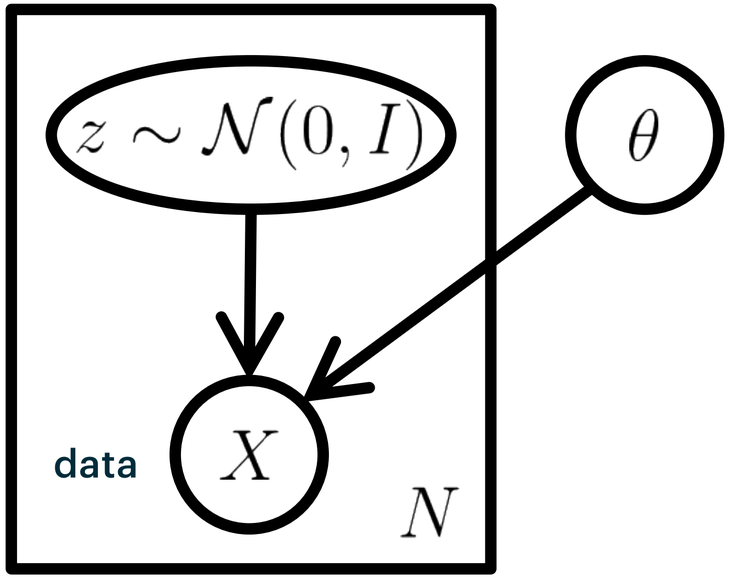
(Tang, Peng & Li. 2023)

2. GENERATIVE MODELS

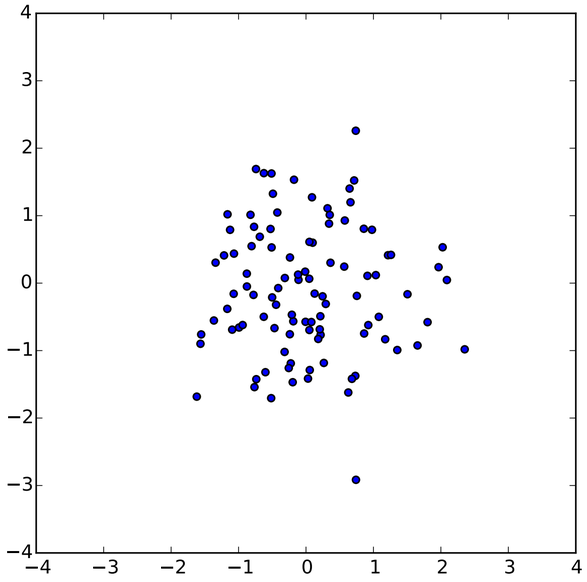
Variational Auto-Encoder (VAE)

A generative approach to non-linear dimension reduction

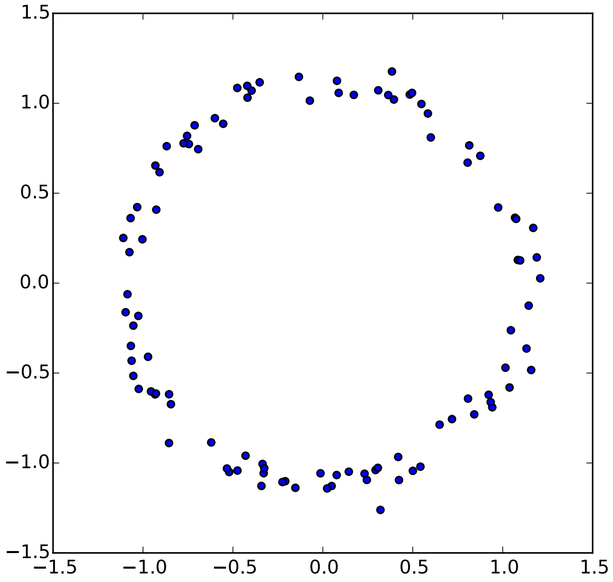
latent variable



$$P(X) = \int P(X|z;\theta)P(z)dz.$$



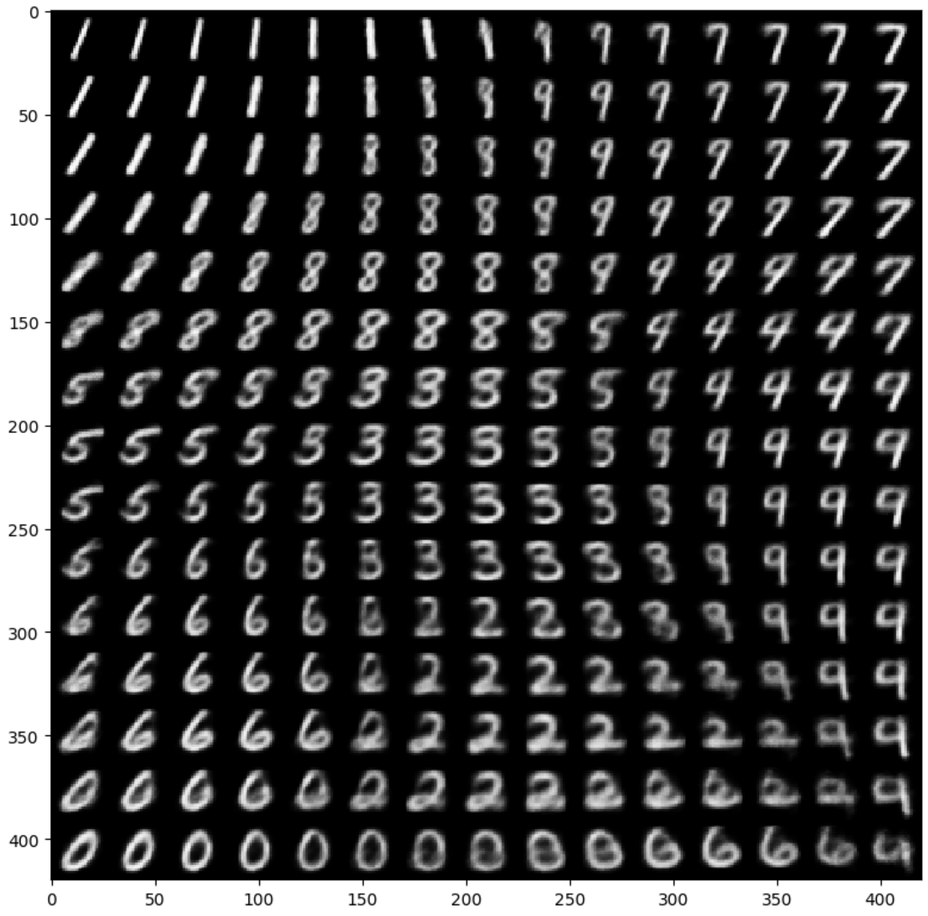
z



$x = z/10 + z/\|z\|$

2. GENERATIVE MODELS

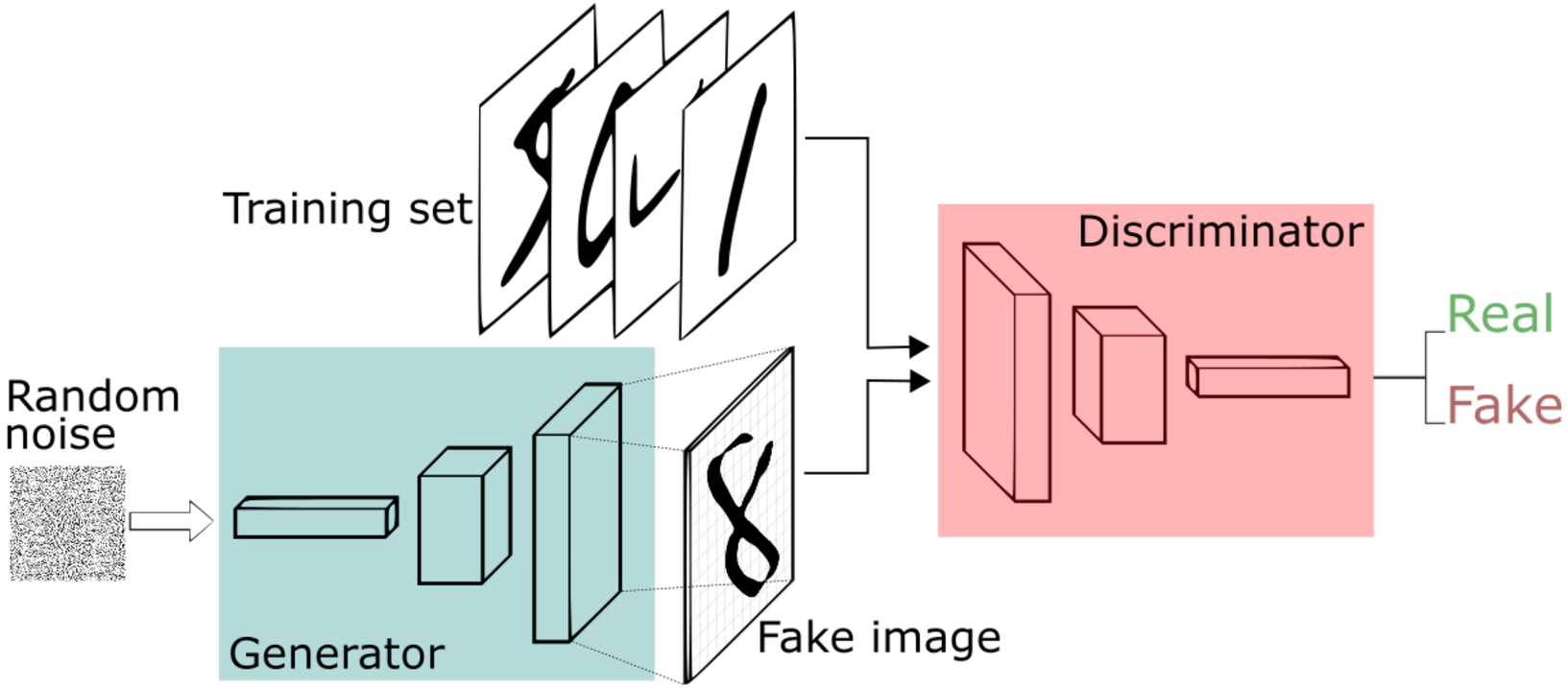
- Variational Auto-Encoder (VAE)
  - Has weak assumptions and small approximation error
  - Can be trained efficiently with backpropagation



MNIST Digits Generated by VAE visualized in 2D

2. GENERATIVE MODELS

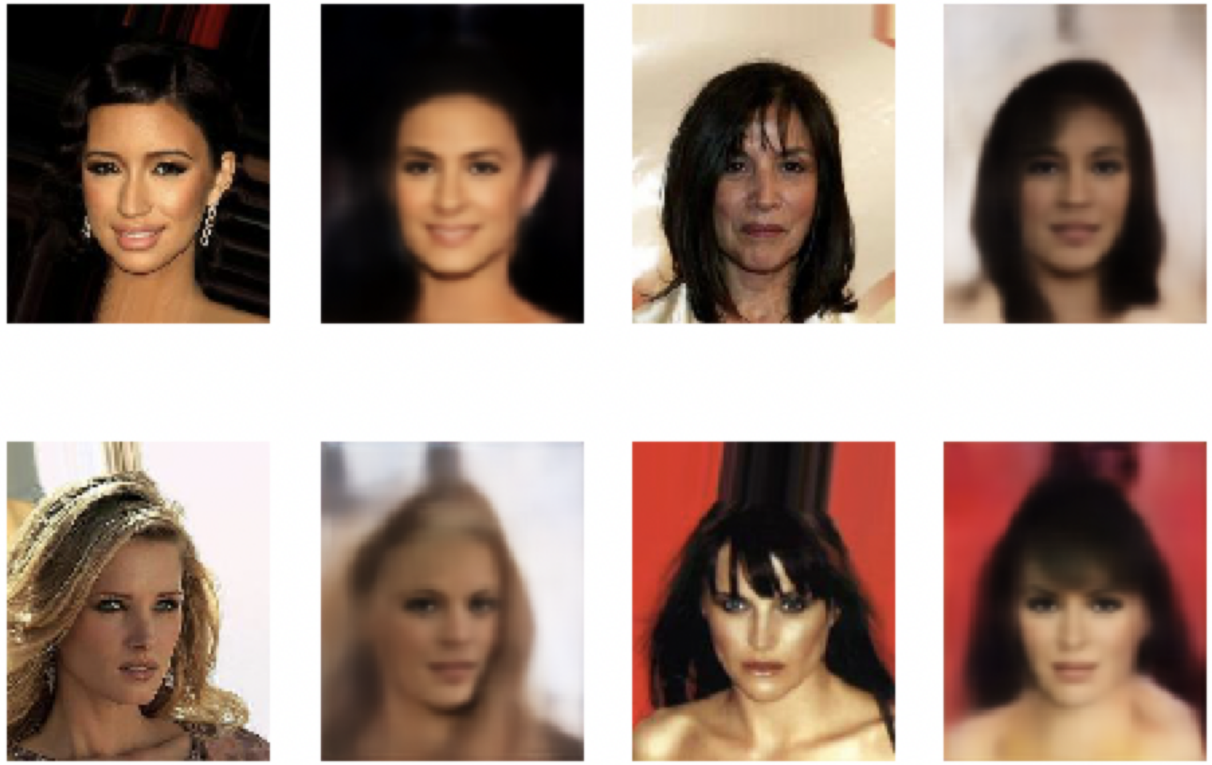
- **Generative models are difficult to approximate**
- **GAN: Generative modeling + Discriminative adversary**



2. GENERATIVE MODELS

GAN can generate more realistic images

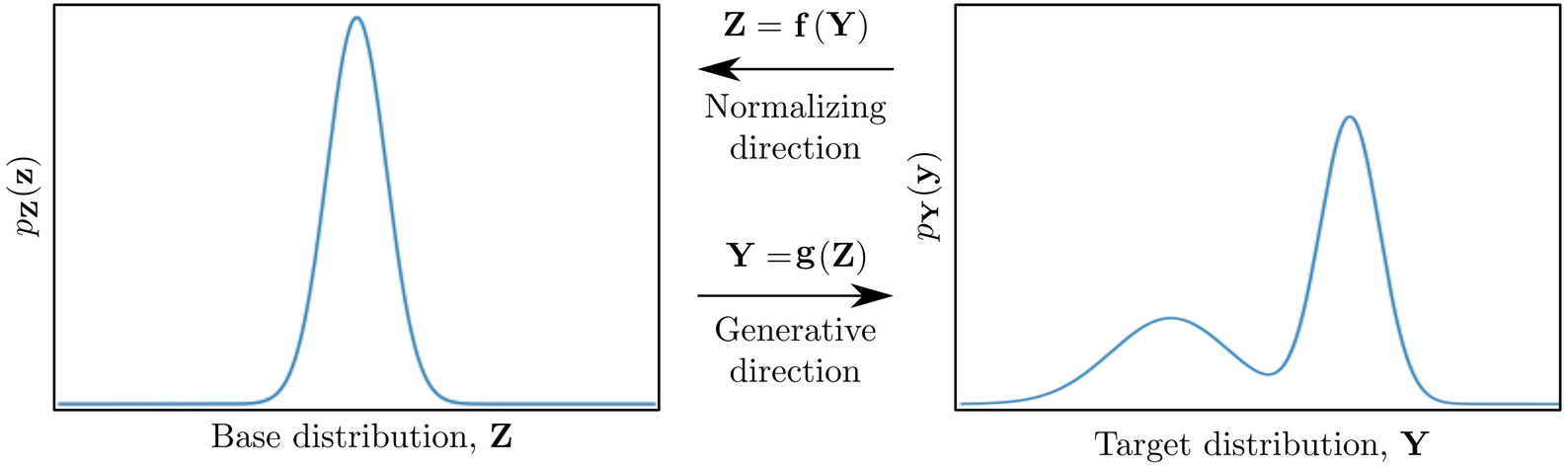
GAN (left) vs VAE (right)



GAN does not provide tractable estimation of density $P(x)$

2. GENERATIVE MODELS

- **Normalizing Flows： learning complex data distribution by transforming a simple density through an invertible map**
  - **allow both sampling and density evaluation.**
  - **Good for representing correlations in output structures**

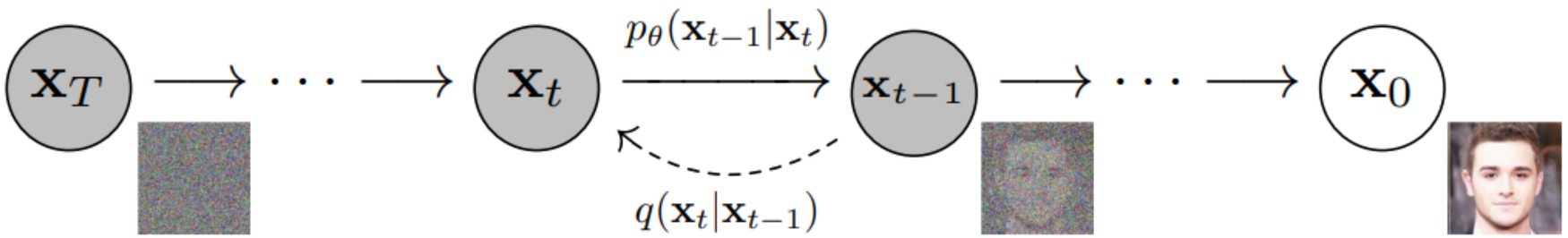


GAN (left) vs VAE (right)

Danilo Jimenez Rezende, Shakir Mohamed, Variational Inference with Normalizing Flows, 2015

2. GENERATIVE MODELS

- **Denoising Diffusion Model： learn a diffusion process that generates a probability distribution for a given dataset**
- **State-of-the-art image generation quality**

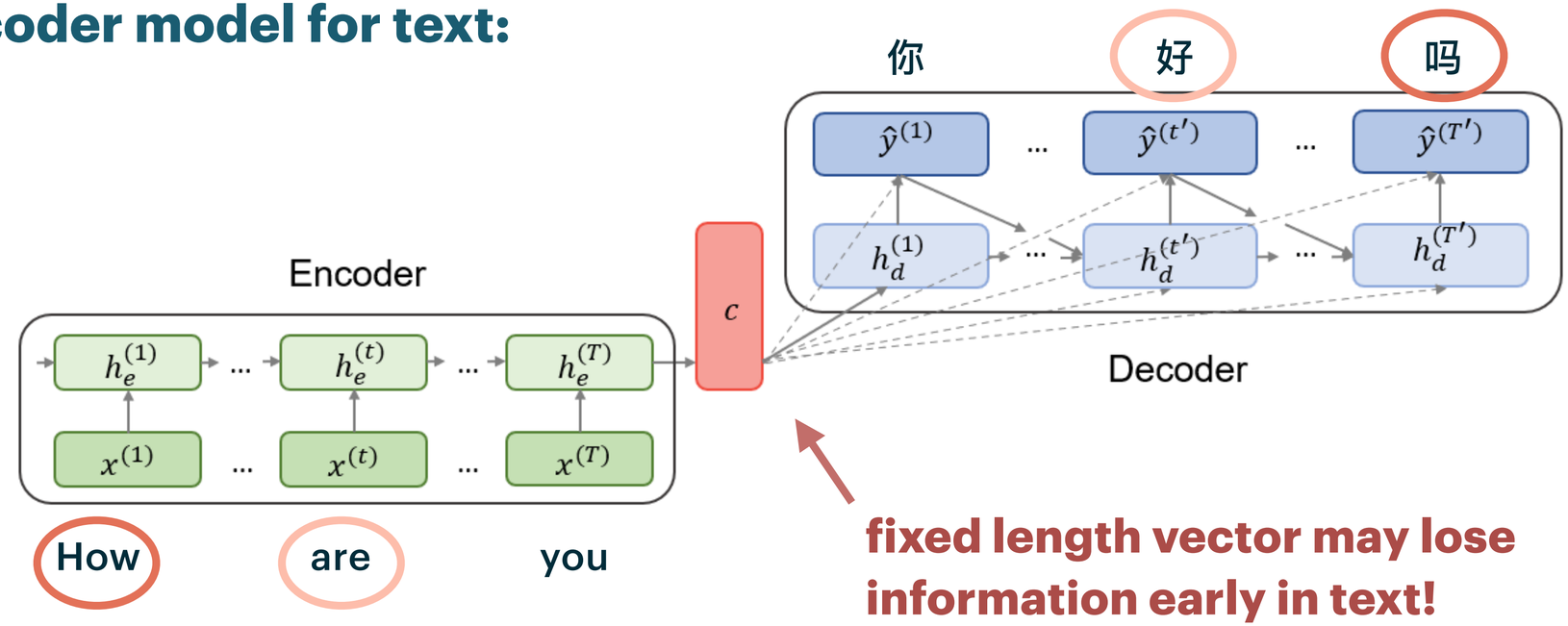


learning to recover the data by reversing the noising process

Jonathan Ho, Ajay Jain, Pieter Abbeel, Denoising Diffusion Probabilistic Models, 2020

3. ATTENTION MECHANISM

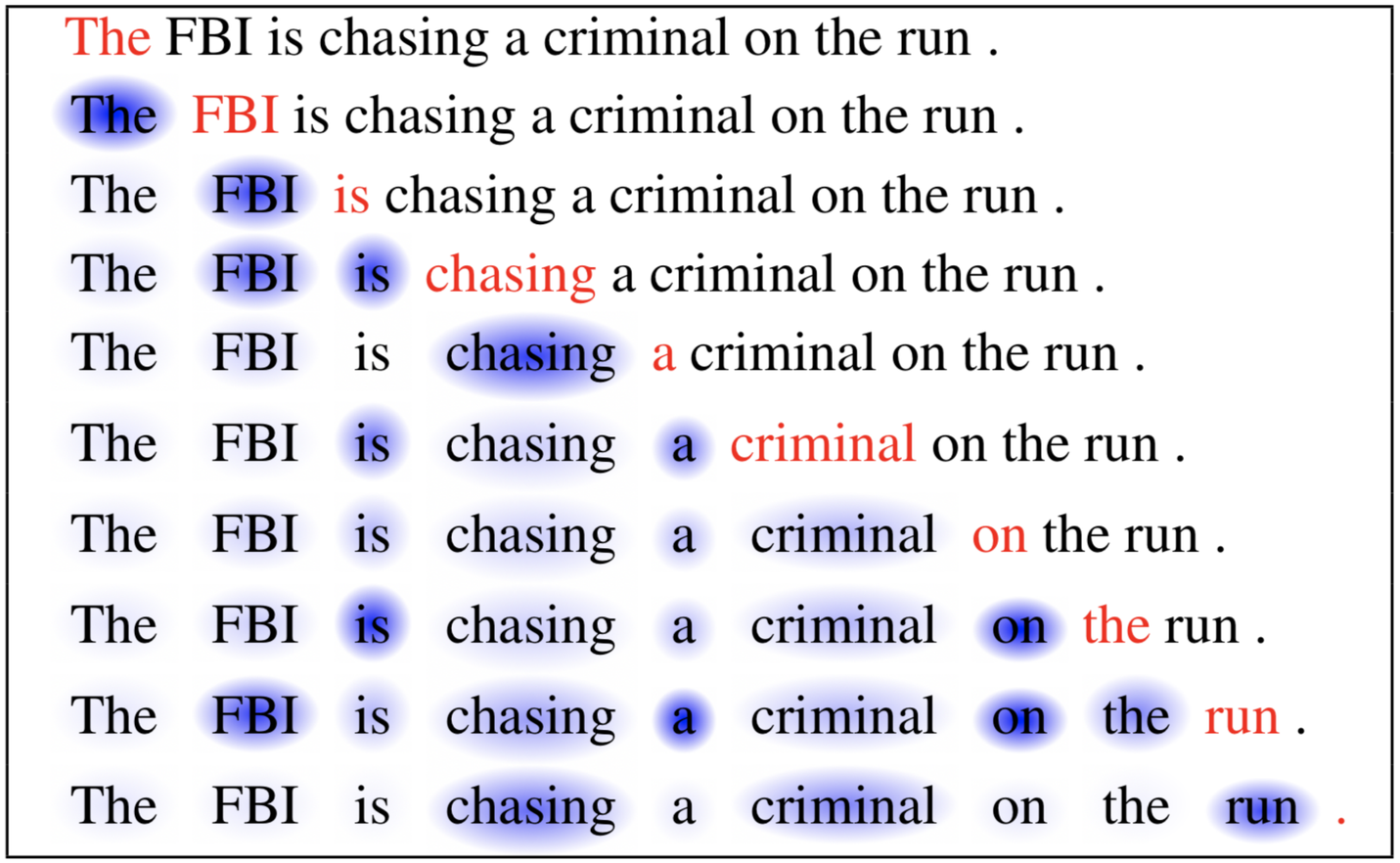
- **Sequence-to-Sequence Problem: map variable-length input sequences to variable-length output sequences (e.g. machine translation)**
- **Encoder-decoder model for text:**



fixed length vector may lose information early in text!

3. ATTENTION MECHANISM

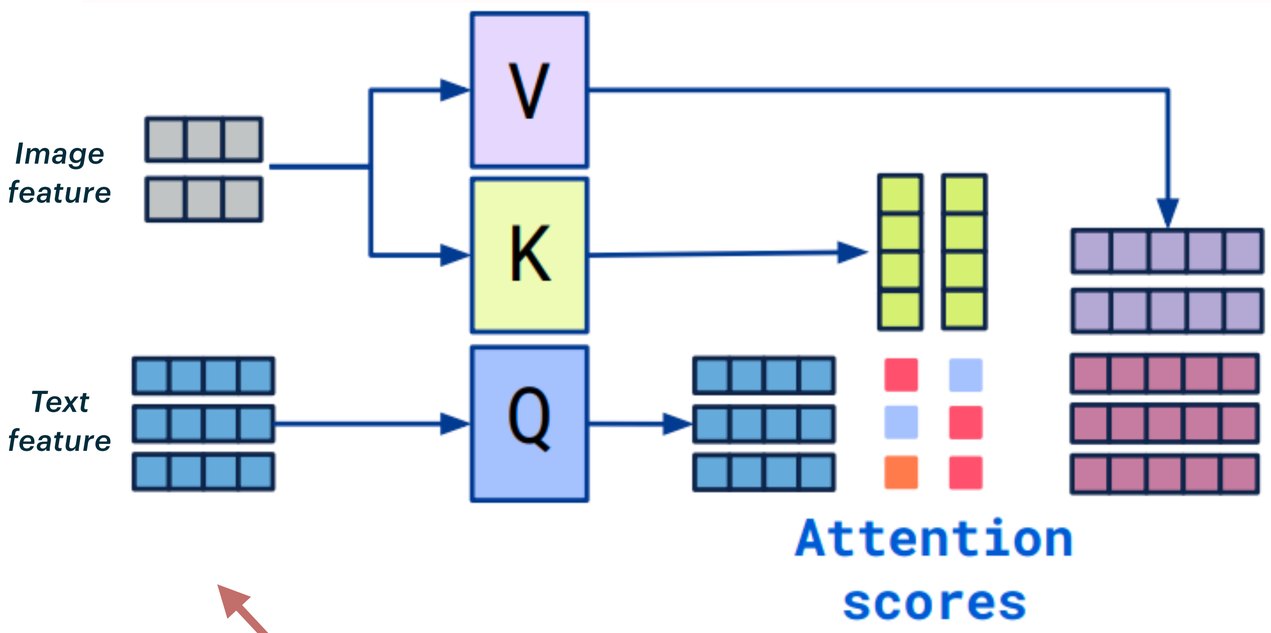
- **Attention enables decoder to "attend" to the whole input sentence.**
- **Self-Attention: A type of attention that relates to different positions of a sentence**



Vaswani, Ashish, et al. "Attention is all you need." *Advances in neural information processing systems* 30 (2017).

3. ATTENTION MECHANISM

- **Cross-Attention: Self-Attention defined on two asymmetric embedding sequences**

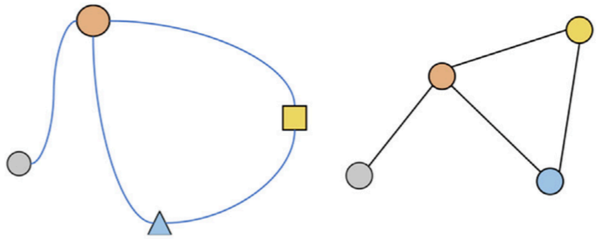


Capture complex relationship between modalities and sequins

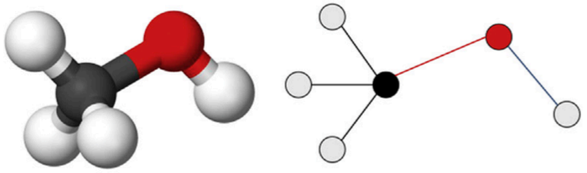
Cross-Attention is All You Need: Adapting Pretrained Transformers for Machine Translation (Gheini et al., EMNLP 2021)

5. GRAPH REPRESENTATION LEARNING

- **Graph neural networks (GNNs): deep learning based methods that operate on graph domain**

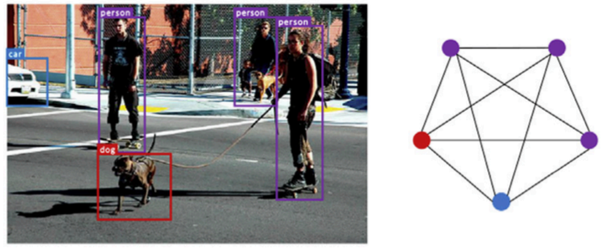


(a) Physics



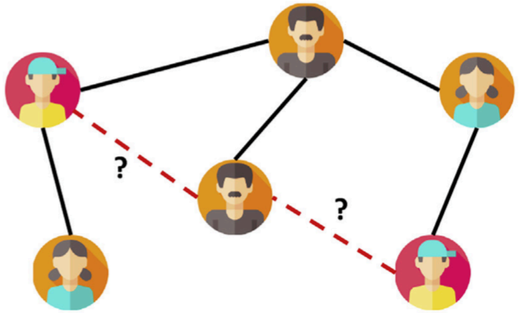
(b) Molecule

Useful for node classification, link prediction, clustering, etc



(c) Image

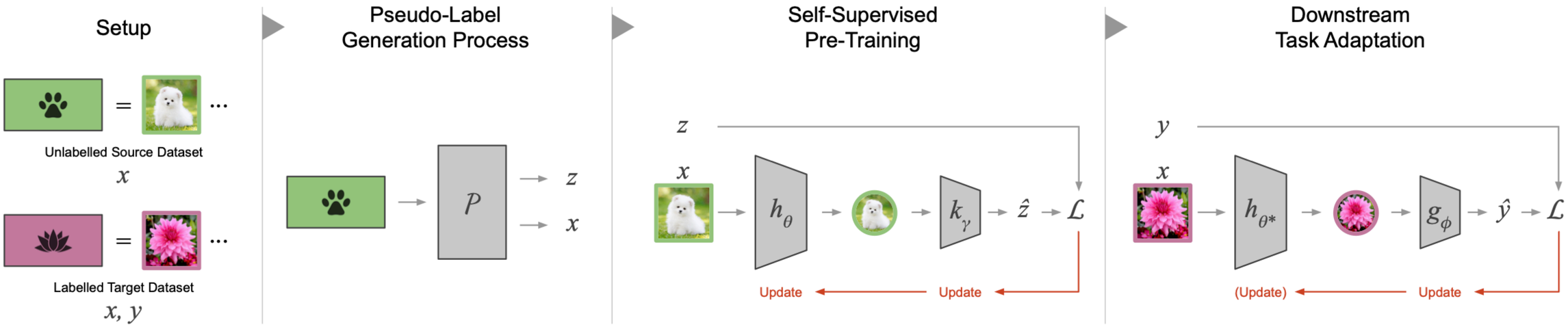
(d) Text



(e) Social Network

5. SELF-SUPERVISED LEARNING

- **Pre-training**: supervised training on a pre-text task (generate labeled training data by editing/destroy-reconstruct/prediction)
- **Fine-tune** learned representation on downstream task



Challenge: how to design a good pre-text task

6. EFFICIENT ADAPTATION

- **Fine-tuning an LLM is very time consuming**
- **How to perform parameter efficient adaptation of LLMs?**

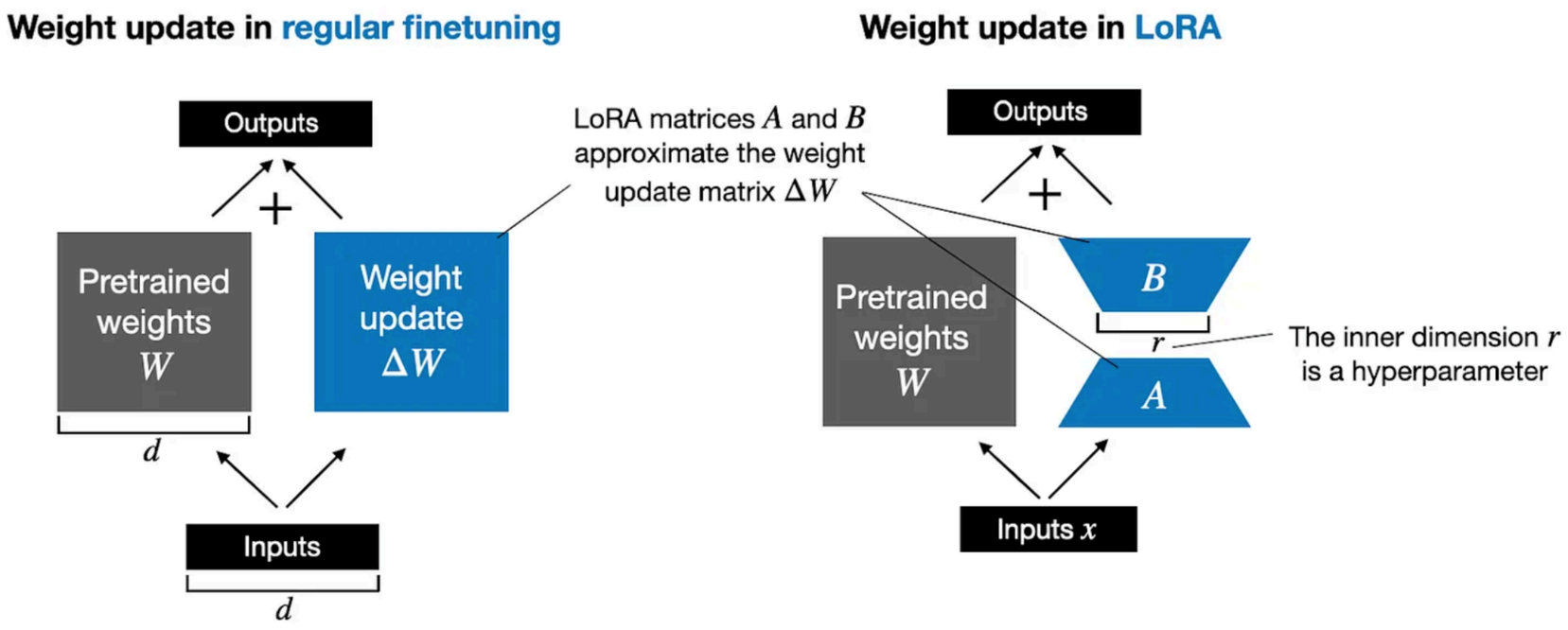


Image: https://magazine.sebastianraschka.com/p/practical-tips-for-finetuning-llms

LECTURE OUTLINE

- About this class
- A preview of the topics we will cover
- Break
- Discussion on *Representation Learning: A New Perspective*

DISCUSSION QUESTIONS

- **Discussion on *Representation Learning: A New Perspective***

NEXT CLASS

- **Reading Assignment： Wright, John & Yu, Lei & Mairal, Julien & Sapiro, Guillermo & Yan, Shuicheng. (2010). Sparse Representation for Computer Vision and Pattern Recognition. Proceedings of the IEEE. 98. 1031 - 1044.**
- **Homework: sign-up and prepare for your role**

Recommended group size： 3(4)

Feel free to use other references