CS 598: Principles of Generative AI

Spring 2024, TR 2:00–3:15pm (1035 - Campus Instructional Facility)

- Course description: Recent advancements in generative AI have equipped machine learning algorithms with the ability to learn from and accurately replicate observed data, creating new, similar data instances. This course provides an in-depth exploration of the key algorithmic developments in generative models, together with their underlying mathematical principles. We will cover a range of topics such as normalizing flows, variational autoencoders, Langevin algorithms, generative adversarial networks, diffusion models, and sequence generation models, etc.
- Instructor: Prof. Tong Zhang (tozhang@illinois.edu)
  - Office: Siebel Center 2122
  - Office hour: TR 3:15-4:30pm
- TA: Jiaqi Guan (jiaqi@illinois.edu)
  - Office hour: TBD
- Prerequisites: knowledge of machine learning, linear algebra, calculus, and probability, python programming
- Course mechanism:
  - First 16 lectures (before spring break): I will give the lectures, and distribute lecture slides.
  - The first 11 lectures after spring break will be student paper presentations and scribes.
  - The last 2 lectures of student project presentations.
- Grading:
  - individual assignemtns: two theoretical and programming assignments (40%)
  - group assignments (form groups of 3-4)
    * paper presentations (15%) + scribes (10%) + jupiner note experiments explaining the presented papers (10%)
    * final project (20%) + presentations (5%)
- Course Material:
  - Lecture slides
  - Papers
- Lectures
  1. Introduction
  2. Basic neural network models and optimization
  3. Energy-based models and Boltzmann machine
     - **EBM:** Yann LeCun, Sumit Chopra, Raia Hadsell, M Ranzato, and Fujie Huang. A tutorial on energy-based learning. *Predicting structured data*, 1(0), 2006. URL http://yann.lecun.com/exdb/publis/orig/lecun-06.pdf

- **RBM:** Ruslan Salakhutdinov, Andriy Mnih, and Geoffrey Hinton. Restricted boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning*, pages 791–798, 2007. URL https://paperswithcode.com/paper/restricted-boltzmann-machines-for
- **DBM:** Ruslan Salakhutdinov and Geoffrey Hinton. Deep boltzmann machines. In *Artificial intelligence and statistics*, pages 448–455. PMLR, 2009. URL http://proceedings.mlr.press/v5/salakhutdinov09a/salakhutdinov09a.pdf

4. Variational inference

- **VI-1:** Martin J Wainwright, Michael I Jordan, et al. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1–2):1–305, 2008. URL https://www.nowpublishers.com/article/DownloadSummary/MAL-001
- **VI-2:** David M Blei, Alp Kucukelbir, and Jon D McAuliffe. Variational inference: A review for statisticians. *Journal of the American statistical Association*, 112(518):859–877, 2017. URL https://arxiv.org/abs/1601.00670
- **Bayes Inference:** Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International conference on machine learning*, pages 1278–1286. PMLR, 2014. URL http://proceedings.mlr.press/v32/rezende14.pdf

5. Variational auto encoder

- **AE:** Pierre Baldi. Autoencoders, unsupervised learning, and deep architectures. In *Proceedings of ICML workshop on unsupervised and transfer learning*, pages 37–49. JMLR Workshop and Conference Proceedings, 2012. URL http://proceedings.mlr.press/v27/baldi12a/baldi12a.pdf
- **VAE-paper:** Diederik P. Kingma and Max Welling. Auto-encoding variational bayes. In Yoshua Bengio and Yann LeCun, editors, *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*, 2014. URL http://arxiv.org/abs/1312.6114
- **VAE-tutorial:** Diederik P Kingma, Max Welling, et al. An introduction to variational autoencoders. *Foundations and Trends® in Machine Learning*, 12(4):307–392, 2019. URL https://www.nowpublishers.com/article/DownloadSummary/MAL-056
- **Importance-weighting:** Yuri Burda, Roger Grosse, and Ruslan Salakhutdinov. Importance weighted autoencoders. *arXiv preprint arXiv:1509.00519*, 2015. URL https://arxiv.org/pdf/1509.00519

6. Normalizing flow

- **NICE:** Laurent Dinh, David Krueger, and Yoshua Bengio. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014. URL https://arxiv.org/pdf/1410.8516.pdf
- **NF-paper:** Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015. URL http://proceedings.mlr.press/v37/rezende15.pdf
- **NF-general:** George Papamakarios, Eric Nalisnick, Danilo Jimenez Rezende, Shakir Mohamed, and Balaji Lakshminarayanan. Normalizing flows for probabilistic modeling and inference. *The Journal of Machine Learning Research*, 22(1):2617–2680, 2021. URL https://www.jmlr.org/papers/volume22/19-1028/19-1028.pdf

7. Sampling basics

- **MCMC Introduction:** Christophe Andrieu, Nando De Freitas, Arnaud Doucet, and Michael I Jordan. An introduction to mcmc for machine learning. *Machine learning*, 50:5–43, 2003. URL https://link.springer.com/content/pdf/10.1023/A:1020281327116.pdf

8. MCMC and Langevin algorithms

- **ULA:** Gareth O Roberts and Osnat Stramer. Langevin diffusions and metropolis-hastings algorithms. *Methodology and computing in applied probability*, 4:337–357, 2002. URL https://link.springer.com/article/10.1023/A:1023562417138

- **SGLD:** Max Welling and Yee W Teh. Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 681–688, 2011. URL http://www.icml-2011.org/papers/398_icmlpaper.pdf

9. Brownian motion, SDE, and Fokker–Planck equation
   - **SDE-book:** Simo Särkkä and Arno Solin. *Applied stochastic differential equations*, volume 10. Cambridge University Press, 2019. URL https://users.aalto.fi/~ssarkka/pub/sde_book.pdf
10. Hamiltonian Monte Cardo and under damped Langevin algorithm
    - **HMC:** Michael Betancourt. A conceptual introduction to hamiltonian monte carlo. *arXiv preprint arXiv:1701.02434*, 2017. URL https://arxiv.org/pdf/1701.02434.pdf
    - **Underdamped SGLD:** Tianqi Chen, Emily Fox, and Carlos Guestrin. Stochastic gradient hamiltonian monte carlo. In *International conference on machine learning*, pages 1683–1691. PMLR, 2014. URL http://proceedings.mlr.press/v32/cheni14.pdf
11. f-divefrgence, Wasserstein distance, and convergence of Langevin algorithms
    - Gareth O Roberts and Richard L Tweedie. Exponential convergence of langevin distributions and their discrete approximations. *Bernoulli*, pages 341–363, 1996. URL https://www.jstor.org/stable/3318418
    - Yi-An Ma, Yuansi Chen, Chi Jin, Nicolas Flammarion, and Michael I Jordan. Sampling can be faster than optimization. *Proceedings of the National Academy of Sciences*, 116(42):20881–20885, 2019. URL https://www.pnas.org/doi/full/10.1073/pnas.1820003116
    - Xiang Cheng, Niladri S Chatterji, Peter L Bartlett, and Michael I Jordan. Underdamped langevin mcmc: A non-asymptotic analysis. In *Conference on learning theory*, pages 300–323. PMLR, 2018. URL http://proceedings.mlr.press/v75/cheng18a/cheng18a.pdf
12. Stein variaiontal method, and MMD-distance based generation
    - **SVGD:** Qiang Liu and Dilin Wang. Stein variational gradient descent: A general purpose bayesian inference algorithm. *Advances in neural information processing systems*, 29, 2016. URL https://arxiv.org/abs/1608.04471
    - **MMD:** Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1):723–773, 2012. URL https://www.jmlr.org/papers/volume13/gretton12a/gretton12a.pdf
    - **MMD-GEN:** Gintare Karolina Dziugaite, Daniel M Roy, and Zoubin Ghahramani. Training generative neural networks via maximum mean discrepancy optimization. *arXiv preprint arXiv:1505.03906*, 2015. URL https://arxiv.org/pdf/1505.03906
    - **PSR:** Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, pages 359–378, 2007. URL https://sites.stat.washington.edu/raftery/Research/PDF/Gneiting2007jasa.pdf
    - **PSR-Gen:** Xinwei Shen and Nicolai Meinshausen. Engression: Extrapolation for nonlinear regression? *arXiv preprint arXiv:2307.00835*, 2023. URL https://arxiv.org/pdf/2307.00835
13. Score matching
    - **SM:** Aapo Hyvärinen and Peter Dayan. Estimation of non-normalized statistical models by score matching. *Journal of Machine Learning Research*, 6(4), 2005. URL https://www.jmlr.org/papers/volume6/hyvarinen05a/hyvarinen05a.pdf
    - **Denoising SM:** Pascal Vincent. A connection between score matching and denoising autoencoders. *Neural computation*, 23(7):1661–1674, 2011. URL https://www.iro.umontreal.ca/~vincentp/Publications/smdae_techreport.pdf
14. Diffusion model (I)
    - **Original:** Jascha Sohl-Dickstein, Eric Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pages 2256–2265. PMLR, 2015. URL https://arxiv.org/pdf/1503.03585.pdf

- **DDPM:** Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020. URL https://proceedings.neurips.cc/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf

15. Diffusion model (II)
    - **DDIM:** Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. In *International Conference on Learning Representations*, 2020a. URL https://openreview.net/pdf?id=St1giarCHLP
    - **Reverse SDE:** Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2020b. URL https://openreview.net/pdf?id=PxTIG12RRHS
16. Efficient generation
    - **DPM:** Cheng Lu, Yuhao Zhou, Fan Bao, Jianfei Chen, Chongxuan Li, and Jun Zhu. Dpm-solver: A fast ode solver for diffusion probabilistic model sampling in around 10 steps. *Advances in Neural Information Processing Systems*, 35:5775–5787, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/260a14acce2a89dad36adc8eefe7c59e-Paper-Conference.pdf
    - **Consistency:** Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency models. *arXiv preprint arXiv:2303.01469*, 2023. URL https://arxiv.org/pdf/2303.01469.pdf
    - **Prompt:** https://stable-diffusion-art.com/prompt-guide/

SpringBreak

17. Diffusion model practice
    - **Diffusion Model Tricks:** Alexander Quinn Nichol and Prafulla Dhariwal. Improved denoising diffusion probabilistic models. In *International Conference on Machine Learning*, pages 8162–8171. PMLR, 2021. URL http://proceedings.mlr.press/v139/nichol21a/nichol21a.pdf
    - **Stable Diffusion:** Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022. URL https://doi.ieeecomputersociety.org/10.1109/CVPR52688.2022.01042
    - **DALL-E 2:** Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2):3, 2022. URL https://arxiv.org/abs/2204.06125
18. GAN basics
    - **GAN:** Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper_files/paper/2014/file/5ca3e9b122f61f8f06494c97b1afccf3-Paper.pdf
    - **f-GAN:** Sebastian Nowozin, Botond Cseke, and Ryota Tomioka. f-gan: Training generative neural samplers using variational divergence minimization. *Advances in neural information processing systems*, 29, 2016. URL https://proceedings.neurips.cc/paper_files/paper/2016/file/cedebb6e872f539bef8c3f919874e9d7-Paper.pdf
    - **W-GAN:** Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017. URL http://proceedings.mlr.press/v70/arjovsky17a/arjovsky17a.pdf
19. GAN practice
    - **DC-GAN:** Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015. URL https://arxiv.org/pdf/1511.06434.pdf

- **Cycle-GAN:** Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017. URL https://openaccess.thecvf.com/content_iccv_2017/html/Zhu_Unpaired_Image-To-Image_Translation_ICCV_2017_paper.html
- **Style-GAN 2:** Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020. URL https://arxiv.org/abs/1912.04958

20. Neural ODE
    - **Neural ODE:** Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018b. URL https://arxiv.org/abs/1806.07366
    - **Augmented Neural ODE:** Emilien Dupont, Arnaud Doucet, and Yee Whye Teh. Augmented neural odes. *Advances in neural information processing systems*, 32, 2019. URL https://proceedings.neurips.cc/paper/2019/file/21be9a4bd4f81549a9d1d241981cec3c-Paper.pdf
    - **Train Neural ODE:** Chris Finlay, Jörn-Henrik Jacobsen, Levon Nurbekyan, and Adam Oberman. How to train your neural ode: the world of jacobian and kinetic regularization. In *International conference on machine learning*, pages 3154–3164. PMLR, 2020. URL http://proceedings.mlr.press/v119/finlay20a/finlay20a.pdf
21. Representation learning and disentanglement
    - **beta-VAE:** Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *International conference on learning representations*, 2016. URL https://openreview.net/forum?id=Sy2fzU9gl
    - **beta-TCVAE:** Ricky TQ Chen, Xuechen Li, Roger B Grosse, and David K Duvenaud. Isolating sources of disentanglement in variational autoencoders. *Advances in neural information processing systems*, 31, 2018a. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/1ee3dfcd8a0645a25a35977997223d22-Paper.pdf
    - **Impossibility:** Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Raetsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations. In *international conference on machine learning*, pages 4114–4124. PMLR, 2019. URL http://proceedings.mlr.press/v97/locatello19a/locatello19a.pdf
22. Basic Sequence Models (before Transformer)
    - **LSTM:** Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9 (8):1735–1780, 1997. URL https://www.bioinf.jku.at/publications/older/2604.pdf
    - **Seq2Seq:** Ilya Sutskever, Oriol Vinyals, and Quoc V Le. Sequence to sequence learning with neural networks. *Advances in neural information processing systems*, 27, 2014. URL https://proceedings.neurips.cc/paper/2014/file/a14ac55a4f27472c5d894ec1c3c743d2-Paper.pdf
    - **Attention:** Dzmitry Bahdanau, Kyung Hyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015*, 2015. URL https://arxiv.org/pdf/1409.0473.pdf
23. Transformer
    - **Transformer:** Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. URL https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf
    - **VIT:** Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit,

and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy

– **MOE:** William Fedus, Barret Zoph, and Noam Shazeer. Switch transformers: Scaling to trillion parameter models with simple and efficient sparsity. *The Journal of Machine Learning Research*, 23 (1):5232–5270, 2022. URL https://www.jmlr.org/papers/volume23/21-0998/21-0998.pdf

24. Sequence based image generation

– **Pixel-RNN:** Aäron Van Den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks. In *International conference on machine learning*, pages 1747–1756. PMLR, 2016. URL http://proceedings.mlr.press/v48/oord16.pdf

– **VQ-VAE:** Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017. URL https://proceedings.neurips.cc/paper/2017/file/7a98af17e63a0ac09ce2e96d03992fbc-Paper.pdf

– **MUSE:** Huiwen Chang, Han Zhang, Jarred Barber, AJ Maschinot, Jose Lezama, Lu Jiang, Ming-Hsuan Yang, Kevin Murphy, William T Freeman, Michael Rubinstein, et al. Muse: Text-to-image generation via masked generative transformers. *ICML'23*, 2023. URL https://arxiv.org/pdf/2301.00704.pdf

25. Encoder only, encoder decoder, and decoder only sequence models

– **BERT**: Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018. URL https://arxiv.org/abs/1810.04805

– **BART**: Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Ves Stoyanov, and Luke Zettlemoyer. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. *arXiv preprint arXiv:1910.13461*, 2019. URL https://arxiv.org/abs/1910.13461

– **GPT-3**: Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020. URL https://arxiv.org/abs/2005.14165

26. Prompt, Finetuning, and Alignment

– **COT:** Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems*, 35:24824–24837, 2022. URL https://arxiv.org/abs/2201.11903

– **InstructGPT:** Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35: 27730–27744, 2022. URL https://arxiv.org/abs/2203.02155

– **RLHF:** Daniel M Ziegler, Nisan Stiennon, Jeffrey Wu, Tom B Brown, Alec Radford, Dario Amodei, Paul Christiano, and Geoffrey Irving. Fine-tuning language models from human preferences. *arXiv preprint arXiv:1909.08593*, 2019. URL https://arxiv.org/abs/1909.08593

27. Visual foundation model and Visual text alignment

– **CLIP:**: Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021. URL https://arxiv.org/abs/2103.00020

– **Llava:** Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. *arXiv preprint arXiv:2304.08485*, 2023. URL https://arxiv.org/abs/2304.08485

– **Videopoet:** Dan Kondratyuk, Lijun Yu, Xiuye Gu, José Lezama, Jonathan Huang, Rachel Hornung, Hartwig Adam, Hassan Akbari, Yair Alon, Vighnesh Birodkar, et al. Videopoet: A large language model for zero-shot video generation. *arXiv preprint arXiv:2312.14125*, 2023. URL https://arxiv.org/abs/2312.14125

28. Student project presentation
29. Student project presentation