CS283: Assignment 5

February 24, 2022

Due: Sunday March 15th, 11:59pm
Submit by the **blakcboard system**

By turning in this assignment, I agree with the KAUST student protocol and declare that all of this is my own work.

Requirements:

- Put all files in a single pdf file containing everything. Attach your code with a link from google drive.
- It is an individual assignment, independent write up, and submission in your own hand is required for credit.
- **Only** LaTeX pdf is accepted for this assignment if you are a CEMSE student. For other students, please write your submission in a **clear** way.

Homework Description

1 Zero-Shot Learning and Vision-Language

For the homework, you are required to get familiar with basic zero-shot learning and vision and language. You need to finish the following:

1.1 Basic Understanding (25 points)

In the first section, you're required to fully understand CIZSL [1] and clearly state the main idea and key designs:

- (5pt) What are the differences between GAZSL [2] and CIZSL [1]?
- (5pt) How the creativity loss is connected with the classification head over classes? Why it can be helpful?

- (15pt) Run the code of CIZSL on one text-based dataset (e.g., CUB-wiki). Please report your performance using the provided hyperparameters (your performance may slightly different from the reported due to instability and different hyper-parameters). You can find the code here.

1.2 Explorations (25 points)

- (25pt) Reproduce the results on attribute-based datasets including AWA2 and SUN. You can mainly build your network on top of GAZSL [2].
- (*Bonus Point.*) (20pt) Replace the original text features with the one extracted from CLIP [3]. (Hint. You only need to get the text features from CLIP instead of re-training it.)

2 Score-Based Generative Modeling

In this section, you're required to fully understand [4], answer some questions, and design some experiments to show your findings. You may be asked to have an interview for your feedback.

2.1 Basic Understanding (30 points)

In the first subsection, you are required to answer the following questions:

- (3pt) Describe the pipeline logic of [4] (i.e., forward and backward steps).
- (2pt) What is Energy-Based Models (EBMs) and Score-Based Generative Models (SBGMs)?
- (5pt) What is the difference among Euler-Maruyama sampling, Langevin MCMC sampling, and Predictor-Corrector (PC) sampling?
- (5pt) What is the difference among VE SDE, VP SDE, and sub-VP SDE?
- (3pt) How and why SDE is connected with SBGMs?
- (5pt) How the likelihood is computed in probabilistic flow ODE? Why this can not be done for normal SDE?
- (7pt) Clarify potential disadvantages of discrete noisy perturbation (Hint, suggest reading [5].)

2.2 Explorations (20 points)

- (10pt) You are required to generate images by using Euler Maruyama sampling strategy. You need to fill in all the TODOs in the given .ipynb. You may check the tutorial here.
- (10pt) Compute the likelihood on CIFAR10 (similar to the example provided on MNIST) for probabilistic flow ODE.

References

[1] Mohamed Elhoseiny and Mohamed Elfeki. Creativity inspired zero-shot learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5784–5793, 2019.

[2] Yizhe Zhu, Mohamed Elhoseiny, Bingchen Liu, Xi Peng, and Ahmed Elgammal. A generative adversarial approach for zero-shot learning from noisy texts. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1004–1013, 2018.

[3] 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.

[4] 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*, 2020.

[5] Yang Song and Stefano Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in Neural Information Processing Systems*, 32, 2019.