ML 2026 Spring HW9
Flow Matching

TA: 林育正、吳岳霖、林禹融、蘇炳揚、陳品睿、江履方

ntu-ml-2026-spring-ta@googlegroups.com

Deadline: 2026/**06/11** 23:59:59 (UTC+8)

Outline

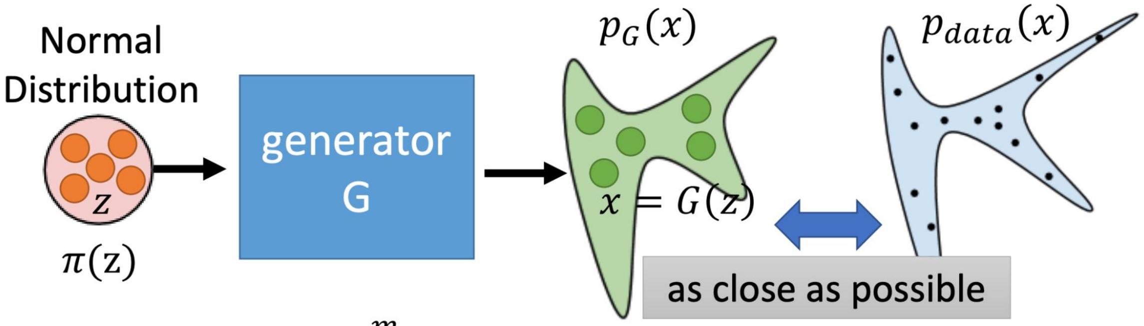
- Task Description
- Assignment Format
- Dataset
- Metric
- Submission & Grading

Useful Links

- Sample code (colab)
- Pytorch Tutorial
- NTU Cool HW9 作業討論區

Prerequisite

- Please watch Prof. Lee's following lecture video before working on this HW.



$$G^* = arg \max_G \sum_{i=1}^{m} log P_G(x^i)$$

$\{x^1, x^2, \dots, x^m\}$ from $P_{data}(x)$

[Flow-based Generative Model]

Prerequisite

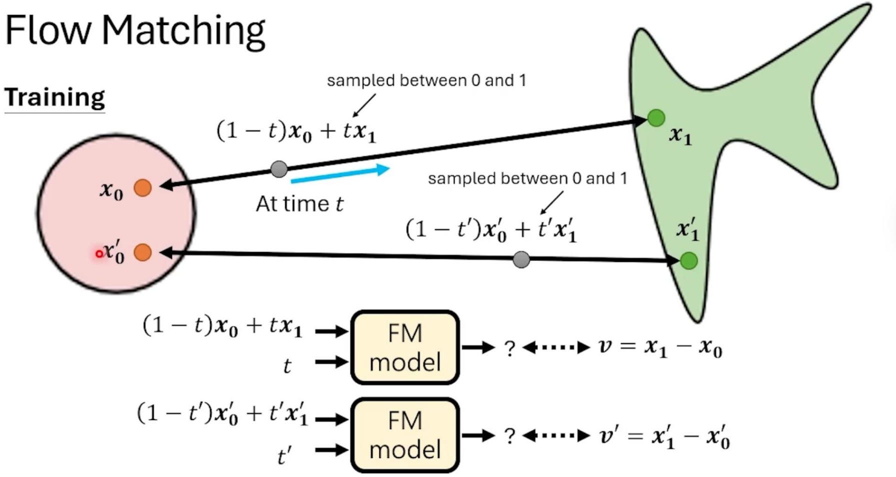
- Please watch Prof. Lee's following lecture video before working on this HW.

【生成式AI】Diffusion Model 原理剖析

Prerequisite

- Please watch Prof. Lee’s following lecture video before working on this HW.

(Ignore the condition in the slides.)



[【生成式人工智慧與機器學習導論2025】第 9 講：影像和聲音上的生成策略]

Task Description

- Goal: Learn about popular current generative models based on diffusion and flow algorithms, then compare their inference-time efficiency.

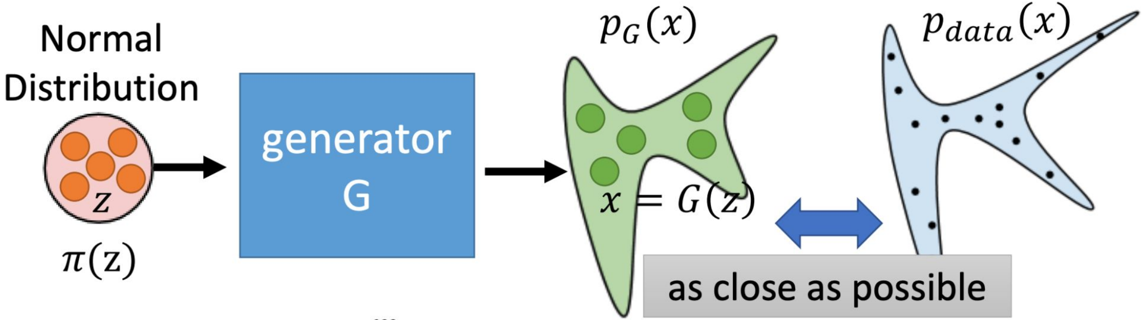
- We recommend students read these papers for a high-level overview:
  - Diffusion Model
  - Flow Matching
  - Rectified Flow
  - MeanFlow

- Here, we provide a tutorial on generative modeling, **skipping** much of the math.

Generative Modeling

- Goal: Learn how to transport from one distribution to another.

- Usually, we map an **easy-to-sample** distribution to the data distribution:



$$G^* = arg\max_G \sum_{i=1}^{m} log P_G(x^i)$$

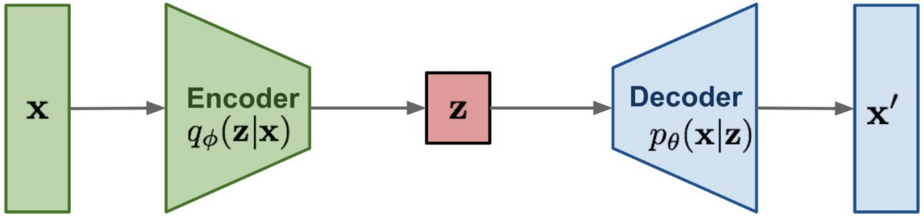
$\{x^1, x^2, \ldots, x^m\}$ from $P_{data}(x)$

Variational Autoencoder (VAE)

- First, we aim to learn how to **encode** an informative latent variable that is similar to a Gaussian distribution and can then be **decoded** to reconstruct the data.

- Here, we omit some of the derivation. If you're interested in variational inference and the **evidence lower bound (ELBO)**, this is closely related to both.

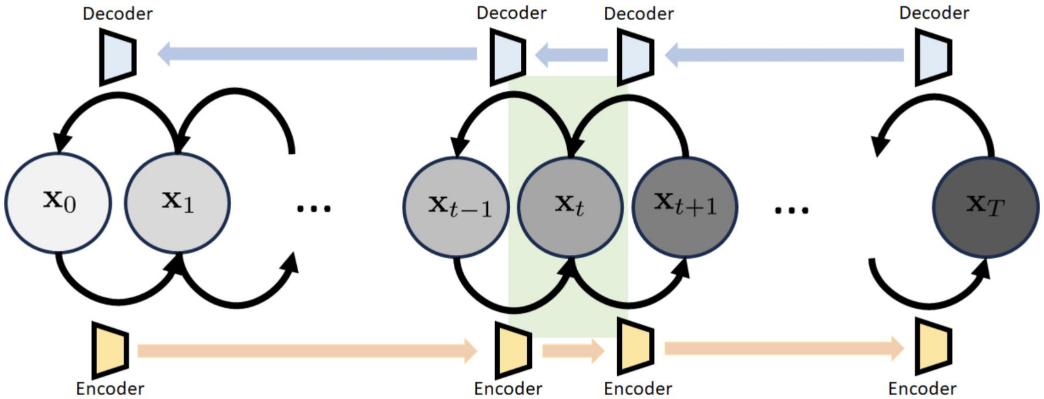
VAE: maximize ELBO.



From VAE to Diffusion

- Is one-step latent encoding too hard for the model? Why not **gradually add noise** to the input? Can the **scheduler** do this instead of the encoder?

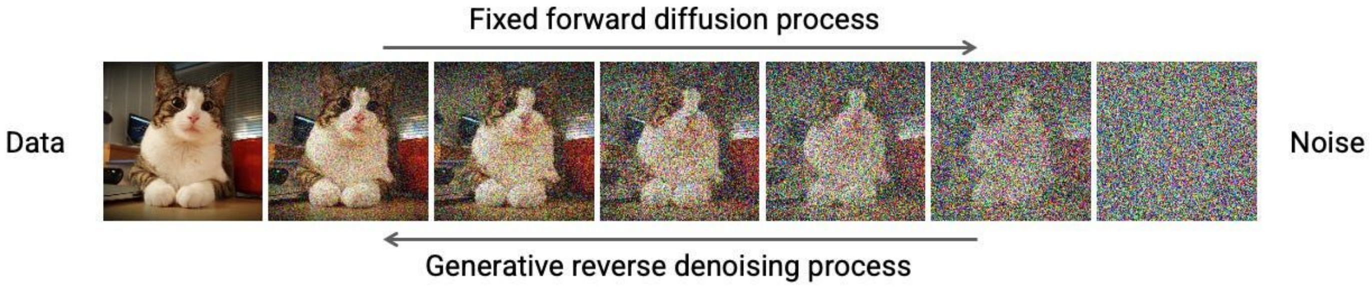
- What about using a Markov chain instead, and then encoding and decoding it?



Diffusion Model

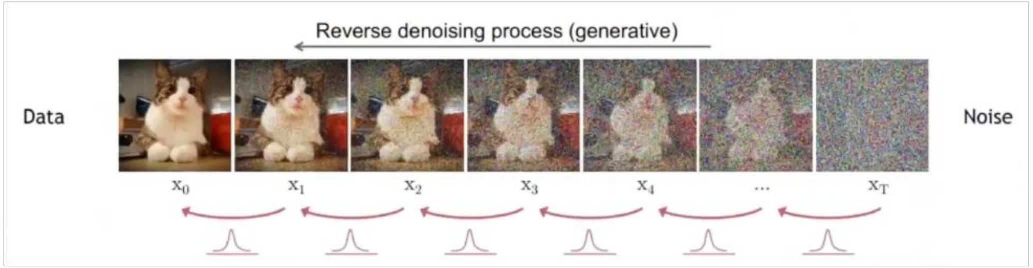
- We can think of the **encoder as a calculation** called the forward process, where noise is manually added, and the decoder as the reverse process.

- So, the goal here is to learn how to **gradually decode (denoise)** the latent variable.



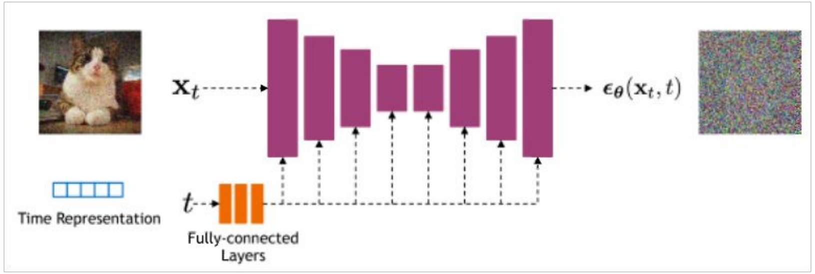
Diffusion Model (Cont'd)

- It can be viewed as first predicting the **clean image**, then estimating the next noisy state. If you're curious why this works, check out these two papers (DDPM, DDIM).



Algorithm 1 Training

1: **repeat**
2: $\mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $t \sim \mathrm{Uniform}(\{1, \ldots, T\})$
4: $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: Take gradient descent step on
$$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$$
6: **until** converged

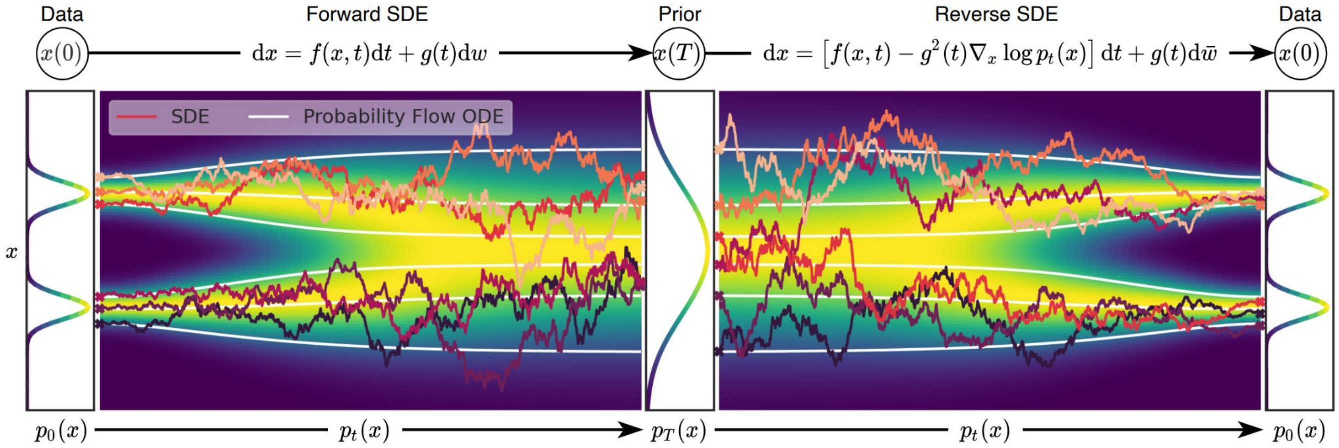


Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

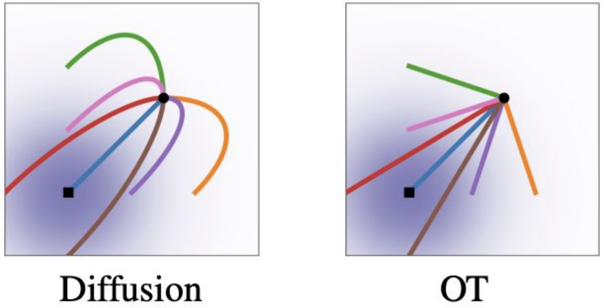
Score-Based Generative Model

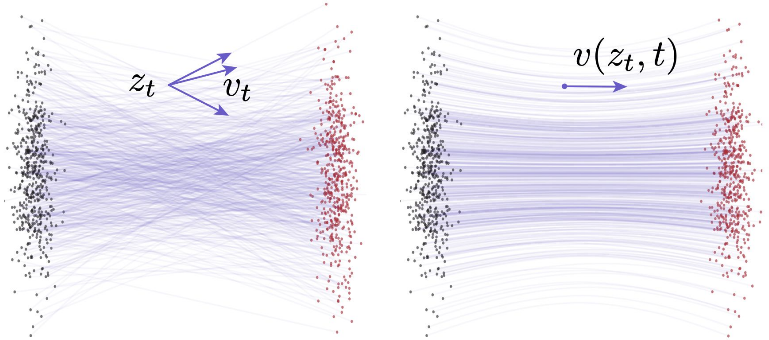
- Let's consider another idea (you can view this paper; not covered in this hw):
    - Make the transitions infinite, indexed by a Markov process, so the model becomes continuous and connects to a **stochastic differential equation**.



Flow Matching

- From this perspective, denoising process may deviate from the conditional **optimal transport (OT)** path, causing diffusion sampling to overshoot.



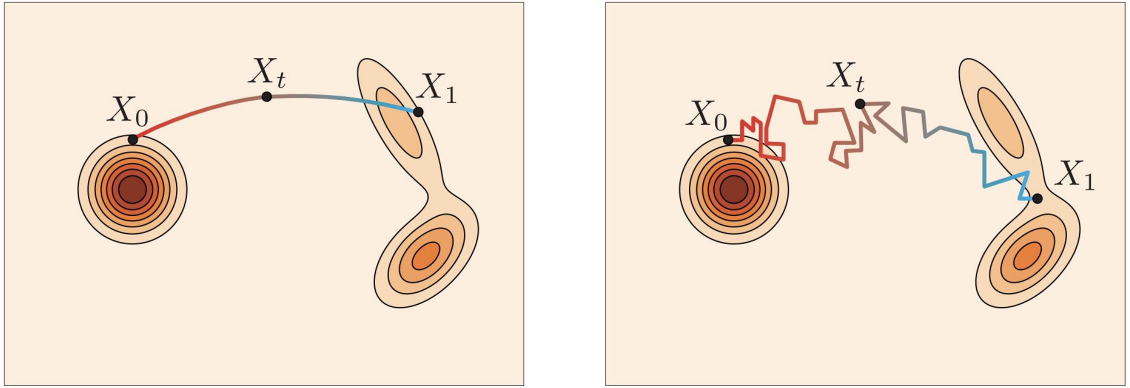


- Why don't we construct a **linear transport** (flow map ψ) and regress its velocity?

$$\mathcal{L}_{\text{CFM}}(\theta) = \mathbb{E}_{t,q(x_1),p(x_0)} \left\| v_t(\psi_t(x_0)) - \frac{d}{dt}\psi_t(x_0) \right\|^2$$

Flow Matching (Cont'd)

- Once our model is trained, we can simply integrate the learned **velocity field** using **ODE solvers** (e.g., Euler) to recover the data, completing the sampling process.

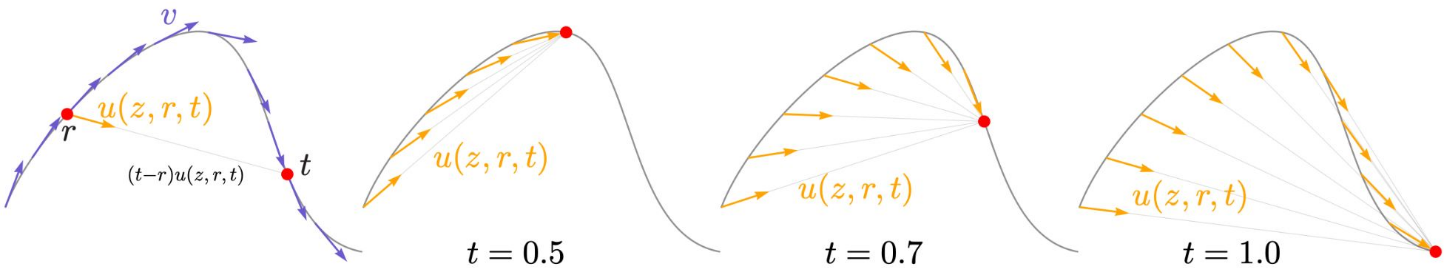


(a) Flow **(b)** Diffusion

MeanFlow

- Up to this point, one challenge is the quality and efficiency trade-off:
  - Diffusion and Flow models require an ODE solver and must query the model multiple times, resulting in a larger **number of function evaluations (NFE)**.

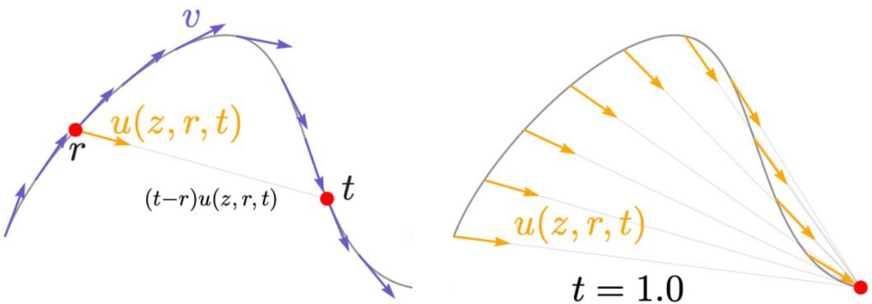
- What if we focus on **average velocity**? Then we can enable few-step generation!



MeanFlow (Cont’d)

- **MeanFlow Identity** provides a training objective for exact few-step inference!

$$\underbrace{u(z_t, r, t)}_{\text{average vel.}} = \underbrace{v(z_t, t)}_{\text{instant. vel.}} - (t - r) \underbrace{\frac{d}{dt} u(z_t, r, t)}_{\text{time derivative}}$$



Algorithm 1 MeanFlow: Training.

Note: in PyTorch and JAX, `jvp` returns the function output and JVP.

```
# fn(z, r, t): function to predict u
# x: training batch

t, r = sample_t_r()
e = randn_like(x)

z = (1 - t) * x + t * e
v = e - x

u, dudt = jvp(fn, (z, r, t), (v, 0, 1))

u_tgt = v - (t - r) * dudt
error = u - stopgrad(u_tgt)

loss = metric(error)
```

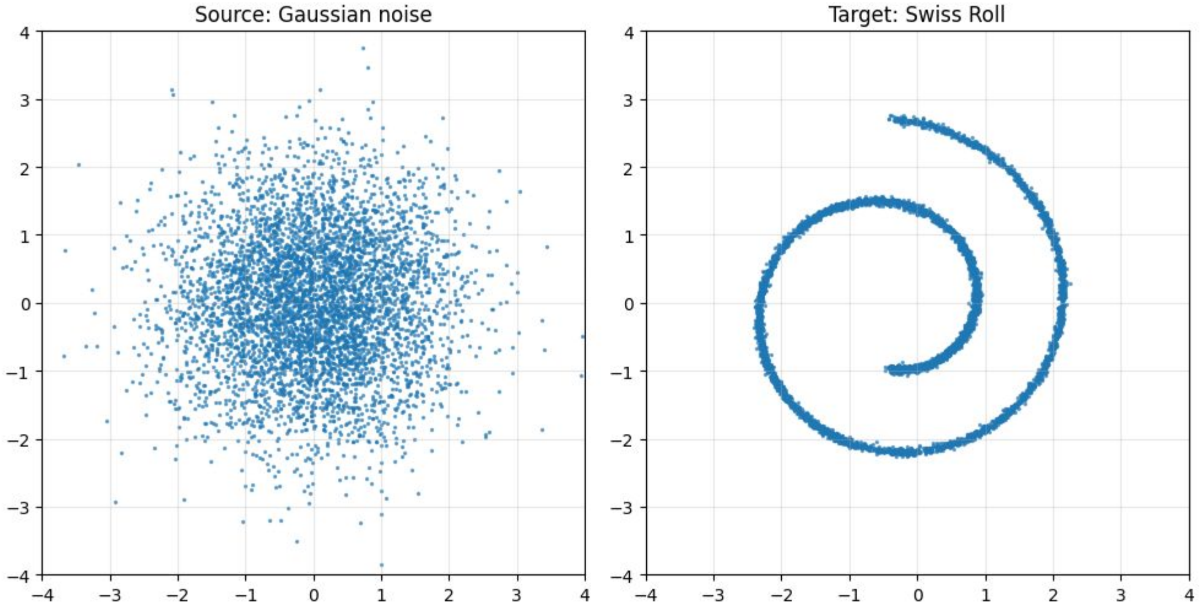
Assignment Format

- This homework consists of 19 questions worth a total of 10 points.
    - Part 1: 16 Paper Reading questions (0.5 pt each).
    - Part 2: 3 Coding questions (graded as 0.5 / 0.5 / 1.0 pt).

- You only need to complete the quiz on NTU Cool and submit it.
- It might be useful to go through a Colab tutorial on these topics first, then answer some paper-reading questions.

Dataset

- Use flow matching and MeanFlow to transform Gaussian noise into a **Swiss roll**.



Metric

- Jensen-Shannon Divergence (JSD)

$$D_{JS}(p||q) = \frac{1}{2} D_{KL}\left(p|| \frac{p+q}{2}\right) + \frac{1}{2} D_{KL}\left(q|| \frac{p+q}{2}\right)$$

  - Measures the "distance" between two distributions
  - Symmetric version of Kullback–Leibler (KL) divergence

Image source: https://arize.com/blog-course/jensen-shannon-divergence/

Submission & Deadline

- Submit your homework to **NTU Cool**, you don't need to submit your code.
- There is no submission limit for the NTU Cool quiz. Your highest score among all attempts will be taken as your final grade.
- **Deadline: 6/11 (Thu.) 23:59**
- **No late submission is allowed**

Grading Release Date

- The grading of the homework will be released by 2026/**06/14** 23:59:59 (UTC+8)

Grading - Regulations

- You should NOT plagiarize, if you use any other resource, you should cite it in the reference.
- Do NOT share codes or prediction files with any living creatures.
- Your final grade x 0.9 and get a score 0 for that homework if you violate any of the above rules first time (within a semester).
- Your will get F for the final grade if you violate any of the above rules multiple (> 1) times (within a semester).
- Prof. Lee & TAs preserve the rights to change the rules & grades.

If You Have Any Questions

- NTU Cool **HW9** 作業討論區
  - 如果同學的問題不涉及作業答案或隱私，請**一律使用** NTU Cool 討論區
  - 助教們會優先回答NTU Cool討論區上的問題

- Email: ntu-ml-2026-spring-ta@googlegroups.com
  - Title should start with [ML 2026 Spring **HW9**]
  - Email with the wrong title will be moved to trash automatically

- TA Hours
  - Each Friday before / after class:
    - (Fri.) 13.20 ~ 14.10 / 17:30~18:00
    - Location: 博理112

Chinese + English

Q1

Q1.

(Choose three)

下列哪些敘述符合 Flow Matching 的核心概念？

A. 它通常學習一個 time-dependent velocity field，用來把 simple prior distribution transport 到 data distribution
B. 它可以被視為一種 continuous normalizing flow / neural ODE 形式的 generative modeling
C. 它的訓練目標通常直接要求模型預測每個時間點的 velocity，而不一定需要模擬完整 reverse diffusion chain
D. 它一定需要像 DDPM 一樣在 inference 時加入 Gaussian noise 才能 sample
E. 它讓 ODE solver 在 inference 的時候一定可以解出微分方程的最佳解

Q2

Q2.

(Choose three)

Flow Matching 論文中，為什麼作者特別討論 **Optimal Transport path**，而不是只使用 diffusion path？

A. OT path 可以讓資料點和噪聲點之間的 interpolation 更接近有效率的 transport
B. OT path 的目標是讓所有樣本在 forward process 中完全變成同一個點
C. OT path 可能使 sampling 所需的路徑更短、更直接，因此改善 sampling efficiency
D. OT path 讓 Flow Matching 完全不需要神經網路，只需要解 closed-form mapping
E. 使用 OT-inspired path 可以展示 Flow Matching 不只是 diffusion model 的另一種寫法

Q3

Q3.

(Choose four)

下列哪些情境下，Flow Matching 的效果或效率可能受限？

A. Learned velocity field 在某些時間區間不 夠 smooth，導致 ODE solver 需要更多 steps 才穩定
B. Target跟prior中間的interpolation設計不好，使得中間 distribution 太難學或 velocity magnitude 太大
C. Model capacity 不足，無法準確近似 conditional velocity field
D. 使用較高階的 ODE solver 或較多 inference steps 來求解 learned velocity field
E. Inference budget 極低，模型只可以用較小的 step去估計

Q4

Q4.

(Choose four)

在 **Mean Flows for One-step Generative Modeling** 中，作者為什麼認為傳統 Flow Matching 的 instantaneous velocity 不一定是 one-step generation 最自然的學習目標？

A. One-step generation 需要的是從起點到終點的整體位移效果，而不只是某一瞬間的速度
B. Flow matching中 Instantaneous velocity 適合用很多小步積分，但直接用一步跨過整個時間區間時可能不 夠對齊目標
C. Average velocity 可以描述一段時間 內的平均 transport behavior，因此可以使用 1-NFE 生成
D. MeanFlow 的重點是完全移除時間變數，讓模型不再需要考慮任何 time conditioning
E. MeanFlow 透過 average velocity 與 instantaneous velocity 之間的 identity，可建立跨時間的 objective

Q5

Q5.

(Choose three)

關於 MeanFlow 和其他 fast sampling 方法的差異，下列哪些 說法較符合論文主張？

A. MeanFlow 不是單純把已訓練好的 flow model 蒸餾成 one-step student
B. MeanFlow 主張可以從 scratch 訓練 one-step generative model
C. MeanFlow 的方法設計需要一個 pretrained 的 flow matching model 並且學習中間的 average velocity
D. MeanFlow 的核心貢獻是把 DDPM 的 Markov chain 加長，讓 one-step model 變成 thousand-step model
E. MeanFlow 企圖縮小 one-step model 與 multi-step diffusion / flow model 的品質差距

Q6

Q6.

(Choose three)

下列哪些情境下，MeanFlow 的優勢可能較明顯？

A. Inference budget 非常低，例如希望 1-NFE 或極少 function evaluations 生成樣本
B. 部署場景重視速度，例如 edge device、real-time generation 或互動式應用
C. 使用者希望避免長時間 multi-step ODE/SDE sampling
D. 任務完全不在乎 inference latency，只追求大量 sampling steps 下的最高品質
E. Model 的 average velocity 學得不夠準，但 Instaneous velocity 學得很好

Q7

Q7.

(Choose one)

下列哪個比較接近 DDPM 原始論文的核心貢獻？

A. 將 diffusion probabilistic model 和 variational bound 訓練連結起來
B. 提出 transformer self-attention 作為主要貢獻
C. 建立和 denoising score matching / Langevin dynamics 的關係
D. 證明 one-step sampling 一定優於多步 sampling
E. 完全放棄 likelihood 或 latent variable model 的觀點

Q8

Q8.

(Choose two)

關於 DDPM 的敘述，下列哪些理解合理？

A. 它可以看成在不同 noise level 下訓練模型預測加入的 noise
B. 它完全不需要 forward noising process
C. 它讓 reverse process 每一步都學會局部去噪
D. 它的主要功能是訓練模型做 image classification
E. 它表示 DDPM 不需要任何 noise schedule

Q9

Q9.

(Choose three)

DDPM 的 progressive lossy decompression 比較接近下列哪些意思？

A. 生成過程可以被看成從資訊很少的 noise 狀態逐步恢復細節
B. 它表示 DDPM 是純 autoregressive language model
C. 不同 reverse timesteps 可以對應到不同層級資訊的重建
D. 資料會先恢復局部的資訊再恢復細節
E. 它的意思是 DDPM 完全不需要 denoising

Q10

Q10.

(Choose three)

在 Rectified Flow 中，straight path 為什麼和 sampling efficiency 有關？

A. 因為比較直的 trajectory 可能降低 coarse discretization 的誤差
B. 因為 straight path 讓模型完全不需要學任何 mapping
C. 因為 few-step 或 single Euler step 會比較有機會有效
D. 因為它可以減少不必要的彎曲 transport
E. 因為 straightness 的唯一功能是增加 stochastic noise

Q11

Q11.

(Choose three)
 關於 Rectified Flow 的 reflow，下列哪些說法合理？

A. 它可以利用前一輪 flow 產生新的 coupling 或 trajectory
B. 它主要是把方法改成 DDPM-style stochastic reverse chain
C. 因為 convex transport cost 不增加的性質，使得 reflow出來的結果可以被保證是好的
D. 它保證一次 reflow 就能得到完美 optimal transport map
E. 它可能讓後續 flow 的 trajectory 更直

Q12

Q12.

(Choose four)
下列哪些任務或設定符合 Rectified Flow 的方法定位？

A. 從一個 empirical distribution transport 到另一個 empirical distribution
B. 可用於image generation
C.使用reflow 讓生成的path更加的straight
D. 學習連接 source distribution 和 target distribution 的 dynamics
E. 完全不需要 source distribution，只需要 final labels

Q13

Q13.

(Choose three)

從「訓練訊號」角度比較四篇論文，下列哪些合理？

A. DDPM：學不同 noise level 下的 denoising / noise prediction 相關目標
B. Flow Matching：根據指定 probability path 學習prior跟target之間根據連續時間的變化
C. MeanFlow：以 average velocity 作為更貼近 one-step generation 的 target
D. Rectified Flow：完全依賴target distribution的學習

Q14

Q14.

(Choose three)

從few step generation角度比較四篇論文，下列哪些 說法比較準確？

A. MeanFlow 最直接把 one-step generation 放進方法設計核心
B. Flow Matching 的 path design 可能影響 sampling 需要多少步才會達到比較好的效果
C. DDPM 做1步的sample相較於200步的結果不會有顯著差異
D. Rectified Flow 透過 straightening 讓 coarse Euler step 更合理
E. Flow Matching 和 Rectified Flow 都完全沒有步數選擇問題

Q15

Q15.

(Choose four)
若一個模型在 1-NFE sampling 表現很差，但在 50-NFE 表現很好，下列哪些解釋比較符合這四篇論文的觀點？

A. 它可能學到的是 instantaneous local dynamics，而不是適合大步跨越時間區間的 average behavior
B. 它的 path 或 trajectory 可能太彎，使 single Euler step 的近似誤差很大
C. 它可能依賴逐步修正誤差，因此少步 sampling 會暴露訓練與 inference 的 mismatch
D. 這代表模型完全沒有學到 data distribution，因為只要 50-NFE 好就不可能 1-NFE 差
E. 若是 DDPM，這種現象可能和 denoising時資訊被壓縮太嚴重有關

Q16

Q16.

(Choose four)
如果一個方法宣稱「比 DDPM 快」，下列哪些比較是必要或至少很重要的？

A. 比較相同 sample quality 下需要多少 NFE
B. 比較相同 NFE 下 sample quality 是否下降
C. 說明是否使用 distillation、teacher model 或額外訓練成本
D. 完全不需要報告 wall-clock time，因為 NFE 一定完全等於真實速度
E. 若方法使用 ODE solver，應說明 solver type、step size 或 discretization 設定

Q17

請截圖 Colab 裡面的 **Summary Histogram JS Table**。提交結果只有在 **Histogram JS values** 符合收斂標準時才算有效：**Flow Matching ≤ 0.10**，**MeanFlow ≤ 0.40**。

	Dataset	Method	Inference steps	Histogram JS ↓
0	Swiss Roll	Flow Matching	50	[redacted]
1	Swiss Roll	MeanFlow	1	[redacted]

Q18

Q18.

(Choose three)
根據訓練過程中觀察到的 **training curves** 和中間生成樣本，下列哪些敘述符合實驗結果？

A. 在訓練過程中，**Flow Matching** 和 **MeanFlow** 的 **losses** 都會隨著訓練進行而平滑下降。
B. 在訓練過程中，**loss** 的波動或暫時上升一定代表生成樣本品質正在變差。
C. **MeanFlow** 比 **Flow Matching** 更難最佳化，而且它的 **training loss** 通常比較不穩定。
D. 即使 **MeanFlow** 的 **training loss curve** 比較不平滑，它仍然可能生成大致符合 **Swiss Roll** 結構的 **one-step samples**。
E. 也應該檢查生成樣本的視覺品質，因為單看 **loss** 不一定能完整反映生成品質。
F. 對兩個模型來說，訓練更多 **epochs** 一定保證會得到更好的生成品質。

Q19(1pt)

在最後的 **inference comparison** 段落中，會比較 **Flow Matching** 和 **MeanFlow** 在不同 **inference settings** 下的生成結果。

請根據你看到的結果，說明兩種方法在生成品質上的主要差異。你的回答至少需要討論下面任兩種比較：

- **Flow Matching** 使用 **1 Euler step** 和 **MeanFlow** 使用 **1 step** 的結果差異。
- **Flow Matching** 在不同 **Euler step counts** 下的結果變化，例如 **1, 5, 10, 20, 50 steps**。
- 在相同 **ODE steps** 下，**Euler** 和 **RK4** 的結果差異。
- 在相近 **model-evaluation budget** 下，**Euler** 和 **RK4** 的結果差異，例如 **Euler 20 steps** 和 **RK4 5 steps**。

不需要回報精確的 **metric values**。請以生成樣本的觀察結果為主，說明 **inference steps**、**solver choice** 和 **computational budget** 會如何影響生成品質。

回答長度約 **100–200 words**。

English problems

Q1

Q1.
(Choose three)
Which of the following statements align with the core concepts of **Flow Matching**?

A. It typically learns a **time-dependent velocity field** that transports a simple prior distribution to the data distribution.
B. It can be viewed as a form of **generative modeling** based on **continuous normalizing flows / neural ODEs**.
C. Its training objective usually directly asks the model to predict the velocity at each time step, without necessarily simulating a full reverse diffusion chain.
D. It necessarily requires adding Gaussian noise during inference, just like DDPM, in order to generate samples.
E. It guarantees that the ODE solver can always obtain the optimal solution to the differential equation during inference.

Q2

Q2.
(Choose three)
In the **Flow Matching** paper, why do the authors specifically discuss the **Optimal Transport path** instead of only using the diffusion path?

A. The OT path can make the interpolation between data points and noise points closer to an efficient transport process.
B. The goal of the OT path is to make all samples collapse into exactly the same point during the forward process.
C. The OT path may make the sampling trajectory shorter and more direct, thereby improving sampling efficiency.
D. The OT path allows Flow Matching to completely avoid using neural networks and only requires solving a closed-form mapping.
E. Using an OT-inspired path demonstrates that Flow Matching is not merely another formulation of diffusion models.

Q3

Q3.
(Choose four)
In which of the following situations might the performance or efficiency of Flow Matching be limited?

A. The learned velocity field is not smooth enough in some time intervals, causing the ODE solver to require more steps for stable sampling.
B. The interpolation design between the prior and target distributions is poor, making the intermediate distributions difficult to learn or causing the velocity magnitude to become too large.
C. The model capacity is insufficient to accurately approximate the conditional velocity field.
D. A higher-order ODE solver or more inference steps are used to solve the learned velocity field.
E. The inference budget is extremely limited, so the model can only use a small number of steps for estimation.

Q4

Q4.

(Choose four)

In **Mean Flows for One-step Generative Modeling**, why do the authors argue that the **instantaneous velocity** used in traditional Flow Matching is not necessarily the most natural learning target for one-step generation?

A. One-step generation requires the overall displacement effect from the starting point to the endpoint, rather than only the velocity at a single instant.
B. In Flow Matching, instantaneous velocity is suitable for integration using many small steps, but it may not align well with the goal when directly crossing the entire time interval in one step.
C. Average velocity can describe the mean transport behavior over a time interval, enabling generation with 1-NFE.
D. The main goal of MeanFlow is to completely remove the time variable, so the model no longer needs any time conditioning.
E. MeanFlow establishes a cross-time objective through an identity between average velocity and instantaneous velocity.

Q5

Q5.
(Choose three)
Regarding the difference between **MeanFlow** and other fast sampling methods, which of the following statements better align with the claims of the paper?

A. MeanFlow is not simply distilling a pre-trained flow model into a one-step student model.
B. MeanFlow argues that a one-step generative model can be trained from scratch.
C. MeanFlow requires a pre-trained Flow Matching model and learns the intermediate average velocity from it.
D. The core contribution of MeanFlow is to lengthen the Markov chain in DDPM, turning a one-step model into a thousand-step model.
E. MeanFlow aims to reduce the quality gap between one-step models and multi-step diffusion / flow models.

Q6

Q6.

(Choose three)

In which of the following scenarios would the advantages of **MeanFlow** likely be more apparent?

A. The inference budget is extremely low, such as when samples need to be generated with 1-NFE or very few function evaluations.
B. The deployment scenario prioritizes speed, such as edge devices, real-time generation, or interactive applications.
C. The user wants to avoid long multi-step ODE/SDE sampling procedures.
D. The task does not care about inference latency at all and only pursues the highest quality under a large number of sampling steps.
E. The model learns the average velocity poorly, but learns the instantaneous velocity very well.

Q7

Q7.
(Choose one)
Which of the following is closest to the core contribution of the original **DDPM** paper?

A. It connects diffusion probabilistic models with training based on a variational bound.
B. It proposes transformer self-attention as its main contribution.
C. It establishes the relationship with denoising score matching and Langevin dynamics.
D. It proves that one-step sampling is always better than multi-step sampling.
E. It completely abandons the perspective of likelihood or latent variable models.

Q8

Q8.
(Choose two)
Regarding **DDPM**, which of the following interpretations are reasonable?

A. It can be viewed as training a model to predict the added noise under different noise levels.
B. It does not require a forward noising process at all.
C. It trains the reverse process to perform local denoising at each step.
D. Its main function is to train a model for image classification.
E. It means that DDPM does not require any noise schedule.

Q9

Q9.
(Choose three)
Which of the following best describe the idea of **progressive lossy decompression** in DDPM?

A. The generation process can be viewed as gradually recovering details from a noise state that contains very little information.
B. It means that DDPM is a purely autoregressive language model.
C. Different reverse timesteps can correspond to the reconstruction of information at different levels.
D. The data may first recover coarse/local information and then gradually recover finer details.
E. It means that DDPM does not require denoising at all.

Q10

Q10.
(Choose three)
In **Rectified Flow**, why is a straight path related to sampling efficiency?

A. Because a straighter trajectory may reduce the error caused by coarse discretization.
B. Because a straight path means the model does not need to learn any mapping at all.
C. Because few-step or single-step Euler sampling is more likely to be effective.
D. Because it can reduce unnecessary curved transport.
E. Because the only function of straightness is to increase stochastic noise.

Q11

Q11.
(Choose three)
Regarding **reflow** in **Rectified Flow**, which of the following statements are reasonable?

A. It can use the flow from the previous round to produce a new coupling or trajectory.
B. Its main purpose is to turn the method into a DDPM-style stochastic reverse chain.
C. Due to the property that the convex transport cost does not increase, the result after reflow can be theoretically justified as improved or at least not worse under that cost.
D. It guarantees that a perfect optimal transport map can be obtained after only one reflow step.
E. It may make the trajectories of subsequent flows straighter.

Q12

Q12.
(Choose four)
Which of the following tasks or settings align with the methodological positioning of **Rectified Flow**?

A. Transporting from one empirical distribution to another empirical distribution.
B. It can be used for image generation.
C. Using reflow to make the generated paths straighter.
D. Learning the dynamics that connect the source distribution and the target distribution.
E. It does not require any source distribution and only needs final labels.

Q13

Q13.
(Choose three)
From the perspective of **training signals**, which of the following comparisons among the four papers are reasonable?

A. **DDPM:** It learns denoising / noise prediction objectives under different noise levels.
B. **Flow Matching:** It learns the continuous-time transformation between a prior distribution and a target distribution based on a specified probability path.
C. **MeanFlow:** It uses average velocity as a target that is more aligned with one-step generation.
D. **Rectified Flow:** It relies entirely on learning only from the target distribution.

Q14

Q14.
(Choose three)
From the perspective of **few-step generation**, which of the following statements provide a more accurate comparison among the four papers?

A. **MeanFlow** most directly places one-step generation at the core of its method design.
B. In **Flow Matching**, the choice of path design may affect how many sampling steps are needed to achieve good results.
C. In **DDPM**, one-step sampling does not show a significant difference compared with 200-step sampling.
D. **Rectified Flow** uses path straightening to make coarse Euler steps more reasonable.
E. **Flow Matching** and **Rectified Flow** completely avoid the issue of choosing the number of sampling steps.

Q15

Q15.
(Choose four)
If a model performs poorly with **1-NFE sampling** but performs well with **50-NFE sampling**, which of the following explanations better align with the perspectives of these four papers?

A. It may have learned instantaneous local dynamics rather than average behavior suitable for crossing a large time interval in one step.
B. Its path or trajectory may be too curved, causing large approximation error when using a single Euler step.
C. It may rely on gradually correcting errors step by step, so few-step sampling exposes a mismatch between training and inference.
D. This means the model has not learned the data distribution at all, because if 50-NFE works well, then 1-NFE cannot be poor.
E. For DDPM, this phenomenon may be related to the fact that too much information is compressed during the denoising process.

Q16

Q16.
(Choose four)
If a method claims to be “faster than DDPM,” which of the following comparisons are necessary, or at least important?

A. Compare how many NFEs are required to achieve the same sample quality.
B. Compare whether sample quality drops under the same number of NFEs.
C. Clarify whether distillation, a teacher model, or additional training cost is used.
D. There is no need to report wall-clock time, because NFE is always exactly equivalent to real speed.
E. If the method uses an ODE solver, it should specify the solver type, step size, or discretization settings.

Q17

Please screenshot the **Summary Histogram JS Table** from Colab. The submitted result is valid only if the Histogram JS values satisfy the convergence standards: Flow Matching ≤ 0.10 and MeanFlow ≤ 0.40.

	Dataset	Method	Inference steps	Histogram JS ↓
0	Swiss Roll	Flow Matching	50	
1	Swiss Roll	MeanFlow	1	

Q18

Q18.
(Choose three)
Based on the training curves and intermediate generated samples observed during training, which of the following statements are consistent with the experiment?

A. During training, the losses of both Flow Matching and MeanFlow decrease smoothly as training progresses.
B. During training, a fluctuation or temporary increase in loss always indicates that the generated sample quality is getting worse.
C. MeanFlow is harder to optimize than Flow Matching, and its training loss is usually less stable.
D. MeanFlow may still produce one-step samples that roughly follow the Swiss Roll structure, even if its training loss curve is less smooth.
E. The visual quality of generated samples should also be checked, because loss alone may not fully reflect generation quality.
F. Training for more epochs always guarantees better sample quality for both models.

Q19(1pt)

In the final **inference comparison** section of the **Colab assignment**, you will compare the generated samples of **Flow Matching** and **MeanFlow** under different **inference settings**.

Please describe the main differences you observe in the generated samples. Your answer should discuss at least two of the following comparisons:

- **Flow Matching** with **1 Euler step** vs. **MeanFlow** with **1 step**.
- **Flow Matching** with different **Euler step counts**, such as **1, 5, 10, 20, and 50 steps**.
- **Euler** vs. **RK4** under the same number of **ODE steps**.
- **Euler** vs. **RK4** under a similar **model-evaluation budget**, such as **Euler 20 steps** vs. **RK4 5 steps**.

You do not need to report exact **metric values**. Focus on what you observe from the generated samples, and explain how **inference steps**, **solver choice**, and **computational budget** affect generation quality.

Your answer should be around **100–200 words**.