

Информативни признания в дълбоките невронни мрежи за класификация на изображения. Генеративно моделиране.

Не просто класификация

Не само асоциация
входове към изходи

Разбиране и представяне
на това по какъв начин
процесите се развиват

Разпознаване на обекти,
техните фактори и техните
вариации

Откриване и детекция на
неочаквани събития в
развитието на процесите

Изграждане на концепции
полезни за обосновки и за
вземането на решения

Подготвеност и генериране
на 'по-богати' планове за
решения в бъдещето

Какво представляват генеративните модели?

Модел позволяващ да се изучи и настрои симулатор на данните

Модели позволяващи условни плътностни оценки

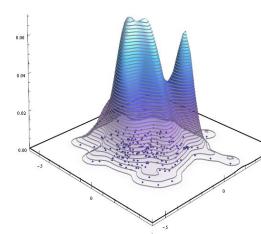
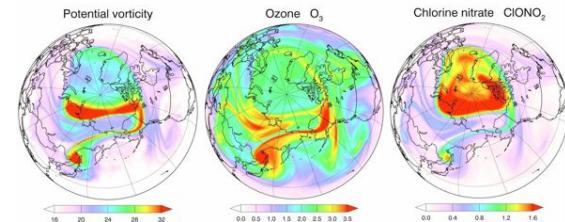
Подход за обучение на модел за данни без учител

Техните характеристики са:

- **Вероятностни** модели на данните, които позволяват описание на неопределености.

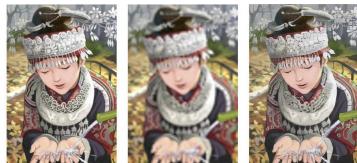
Разпределението на данните $p(x)$ е насочено.

- **Многомерни** изходи.

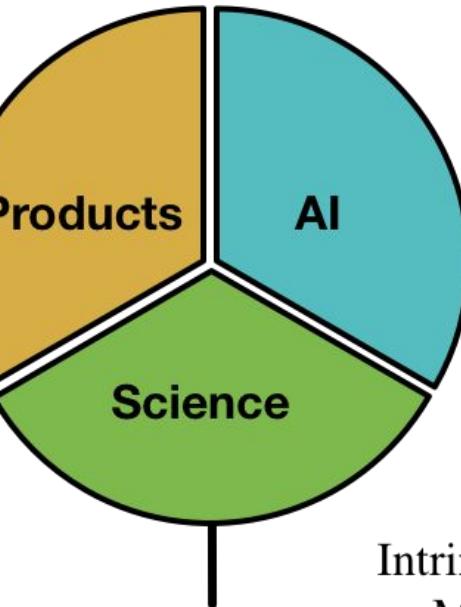
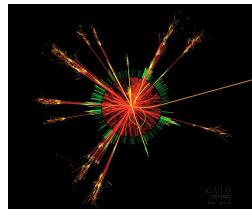


Защо генеративни модели?

Генеративните модели
играят роля в решението на
доста проблеми.



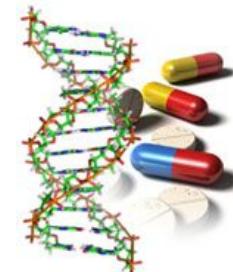
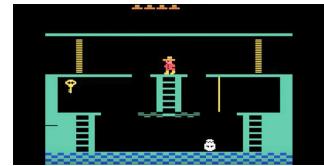
Super-resolution,
Compression,
Text-to-speech



Proteomics,
Drug Discovery,
Astronomy,
High-energy physics

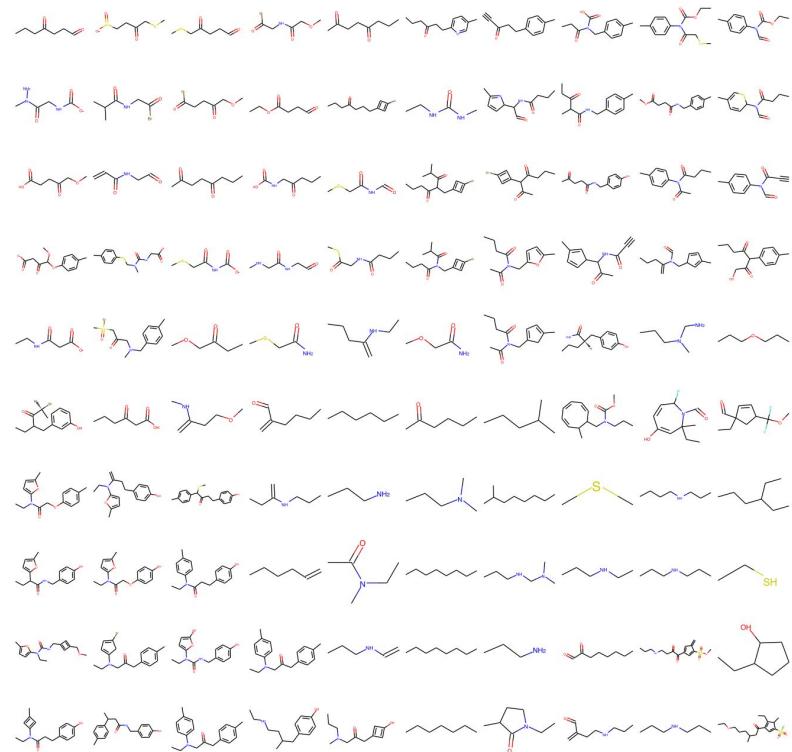
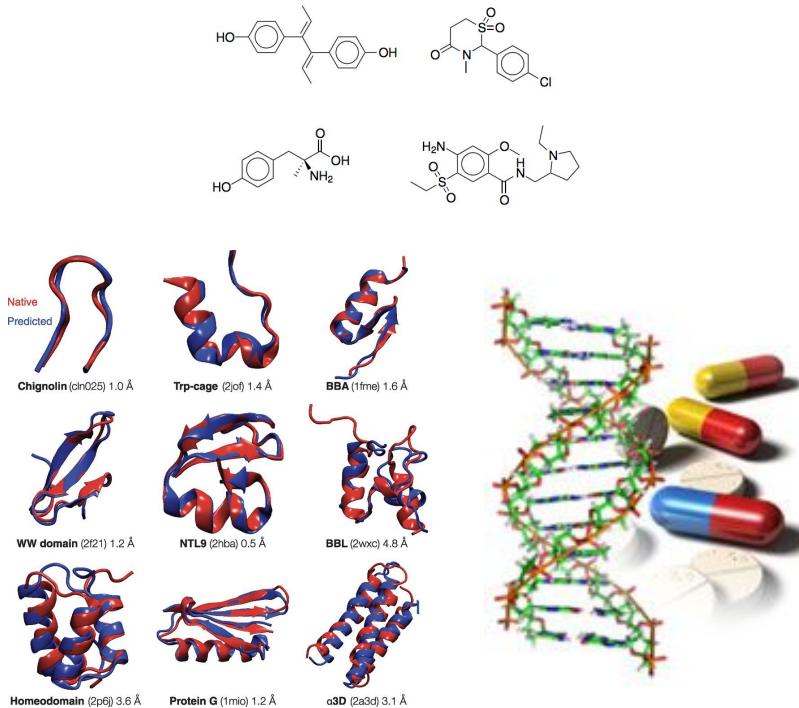


Planning,
Exploration
Intrinsic motivation
Model-based RL



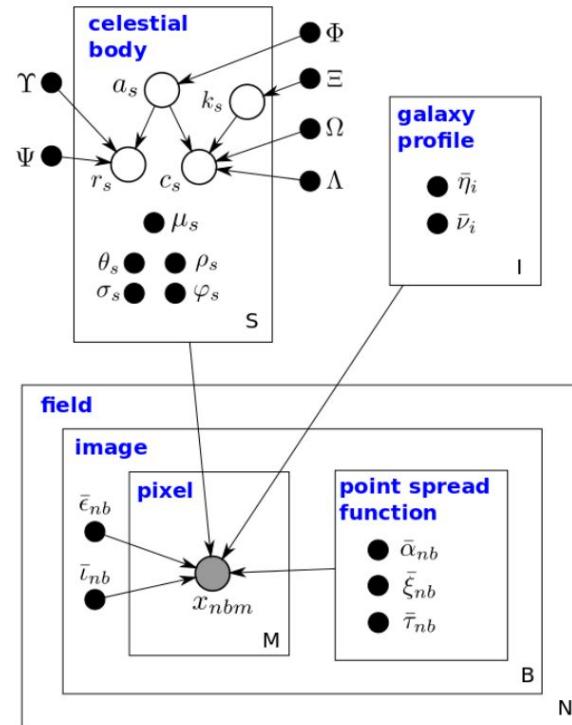
При дизайн на лекарства и очаквани въздействия

Предложения за кандидатни молекули и подобряване на прогнозите с полу-обучение с учител



При откриване и детекция на небесни тела

Генеративни модели за приложения в астрономията и физиката.



Изображения със супервисока разделителна способност

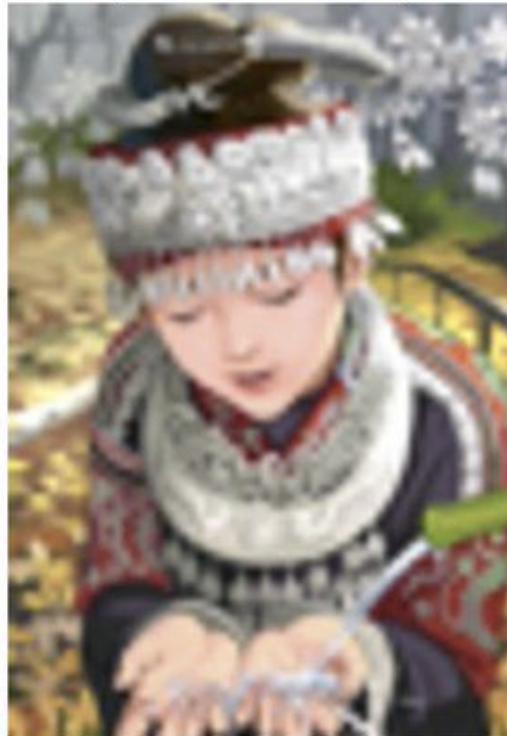
Фотореалистични изображения със super-resolution

original



bicubic

(21.59dB/0.6423)



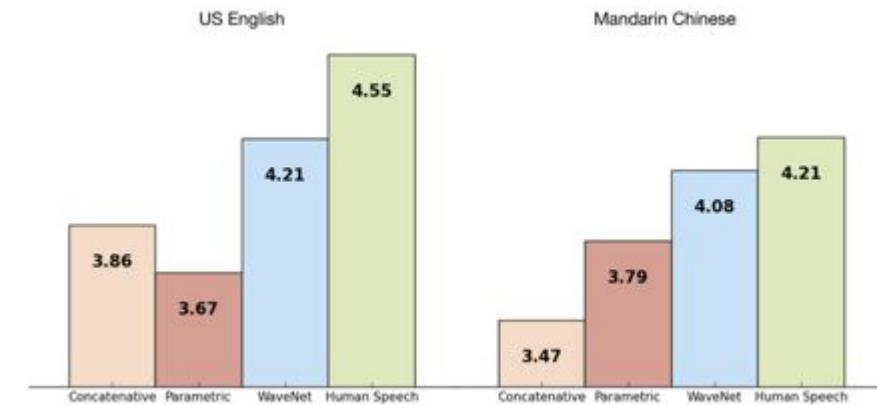
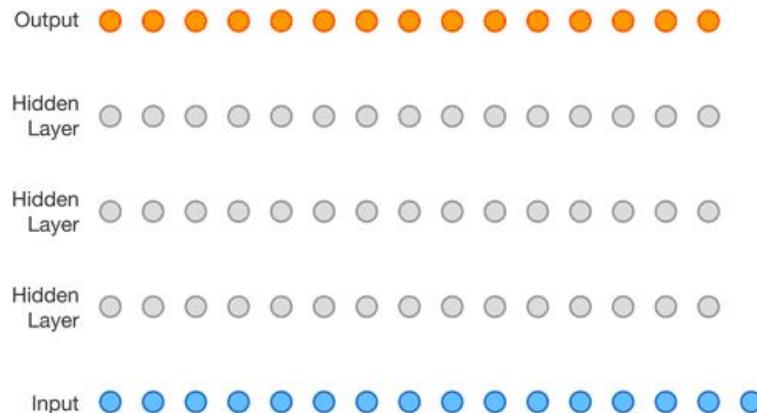
SRGAN

(20.34dB/0.6562)



Синтез на текст към реч

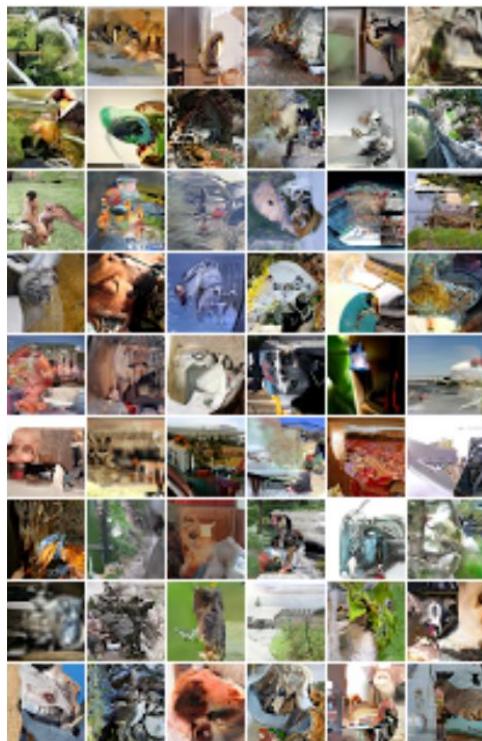
Генериране на звук, обусловен от текста



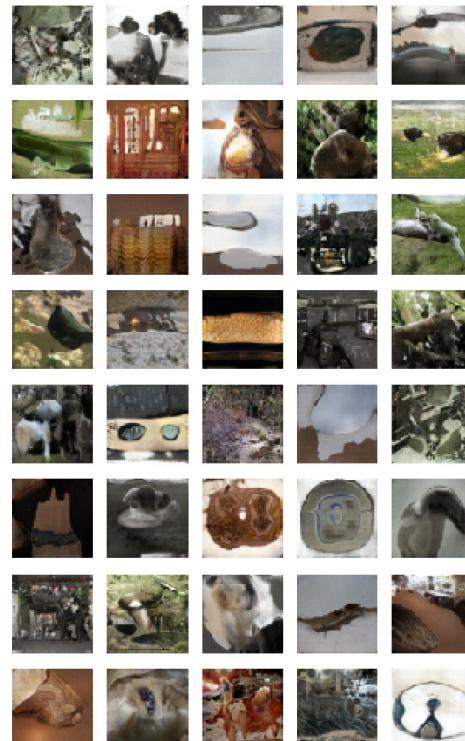
Генериране на изображения и видео съдържание



DRAW



Pixel RNN



ALI

Комуникация и компресия

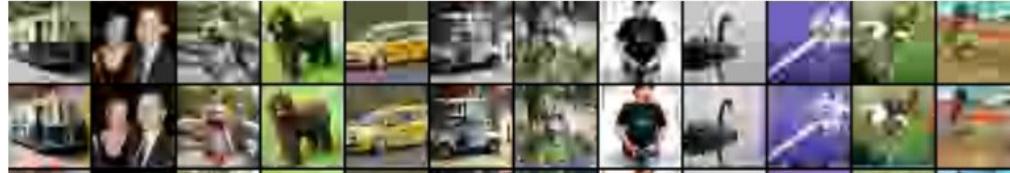
Йерархична компресия на изображения и други данни.

Оригинални изображения

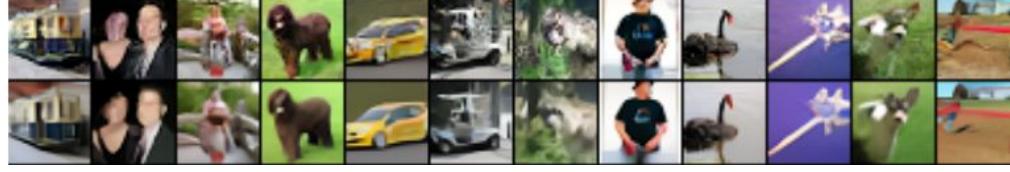


Компресия с ниво: 0.2bits/размерност

JPEG



JPEG-2000



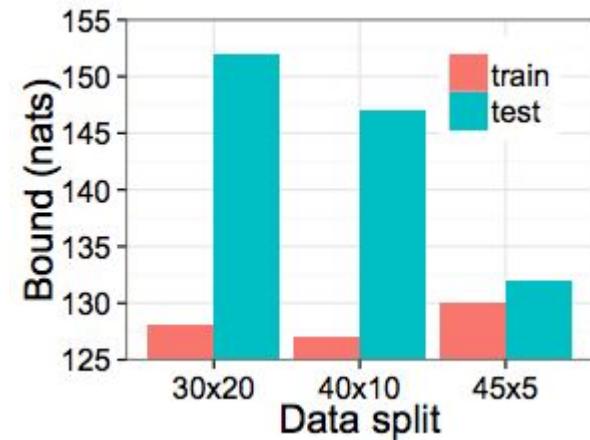
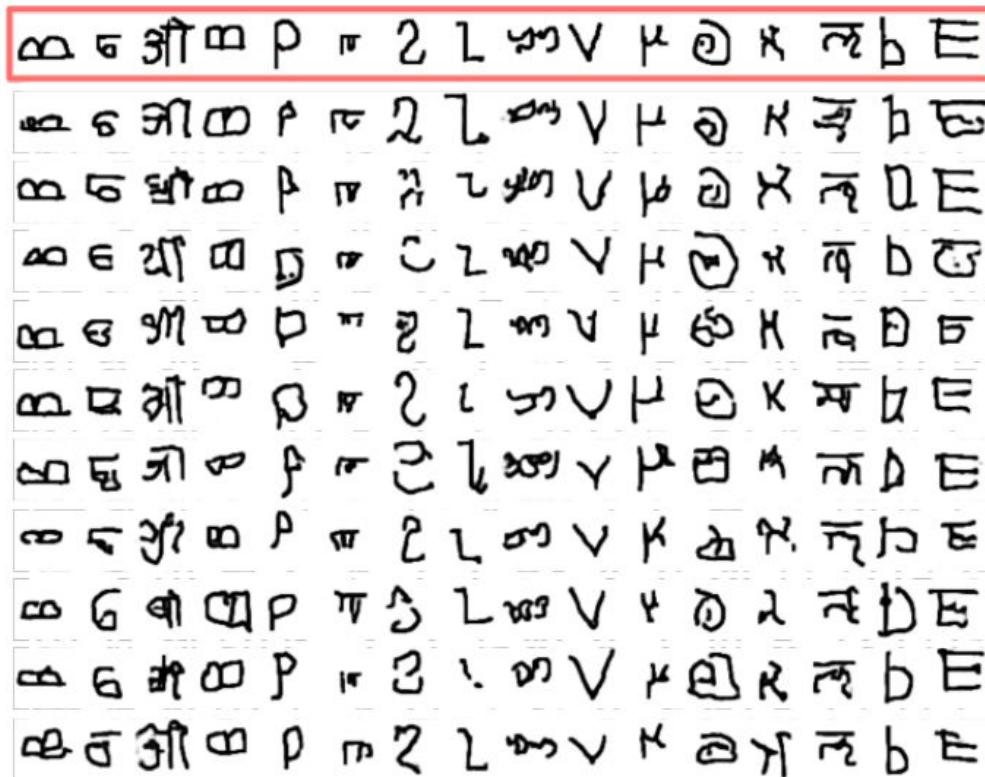
RVAE v1



RVAE v2

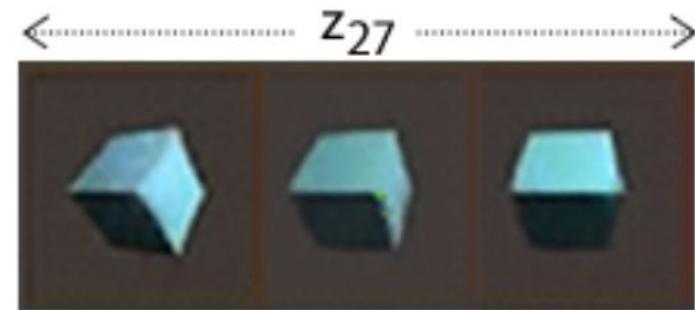
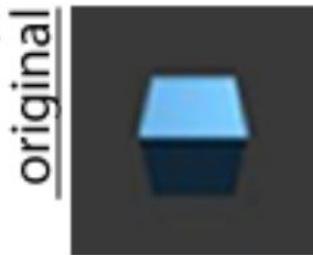
Еднократно обобщение

Бързо обобщение на нови концепции

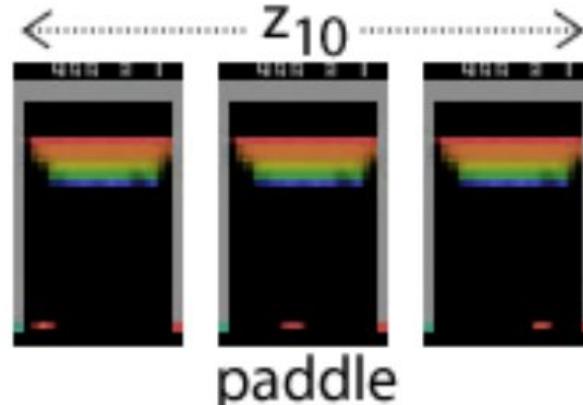
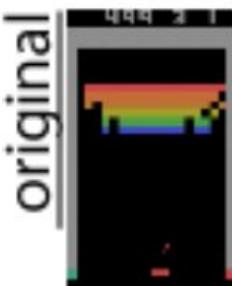


Изучаване на визуални концепции

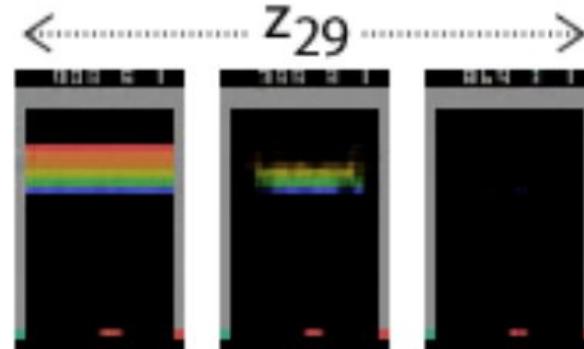
Разбиране на факторите на вариация и инвариантност.



scale



paddle



bricks/score/lives

Симулиране на бъдещи действия

Симулиране на бъдещи траектории на среди въз основа на действия за планиране

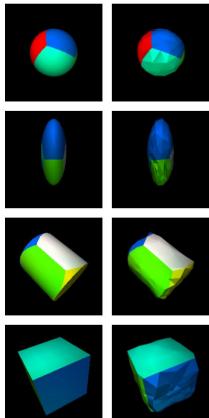
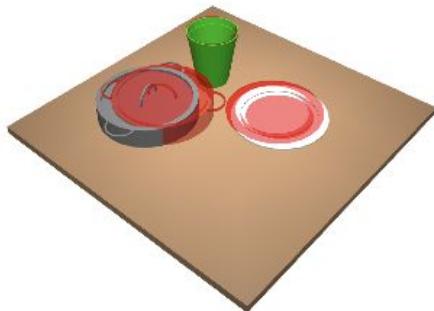


Симулация на ръка на робот

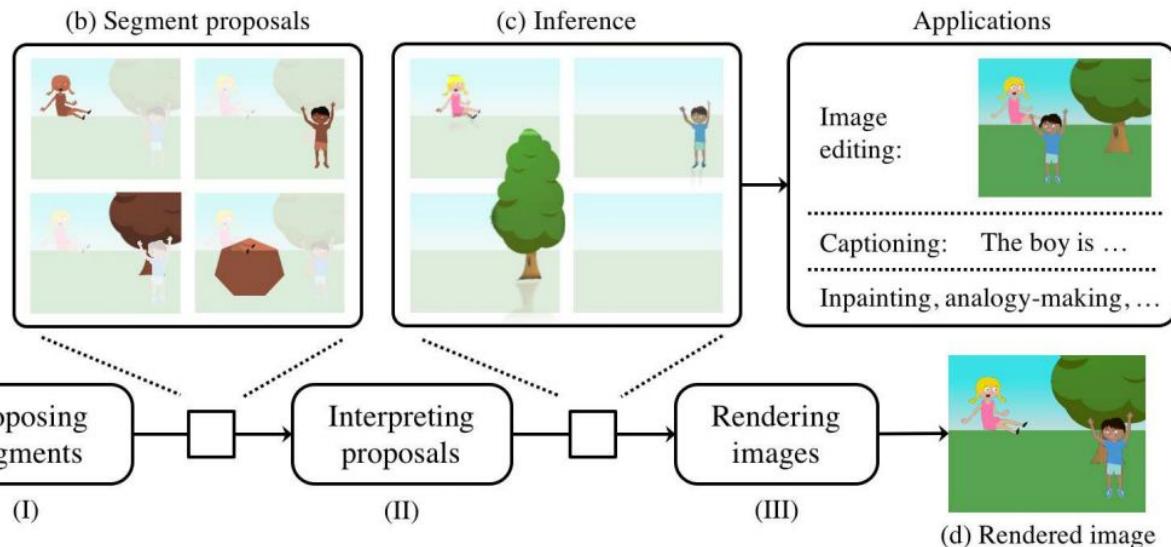
Разбиране на сцени



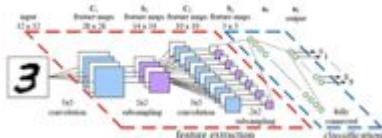
Разбиране на компонентите на сцените и техните взаимодействия



(a) Input image

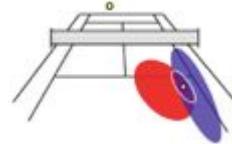


Две направления на машинното обучение



Дълбоко обучение

- + Богати нелинейни модели за класификация и прогнозиране на последователността.
- + Мащабирамо обучение, използващо стохастично приближение и концептуално е просто реализуемо.
- + Лесно композиране с други градиентни методи.
- Само точкови оценки.
- Трудно оценяват модели и извършват селекция и са твърде сложни.



Вероятностни разсъждения

- Основно конюгирали и линейни модели.
- Потенциално недостижими изводи, с прекомерно голям изчислителен ресурс необходим или дълго време за симулация.
- + Единна рамка за изграждане на модели, изводи, прогнозиране и вземане на решения.
- + Изрично отчитане на несигурността и променливостта на резултатите.
- + Издръжлив към пренапасване; с набор за избор и композиция на модела.

Допълващи се силни страни, което прави комбинирането им естествено

Как е приет подходът за машинно обучение



3. Алгоритми



1. Модели



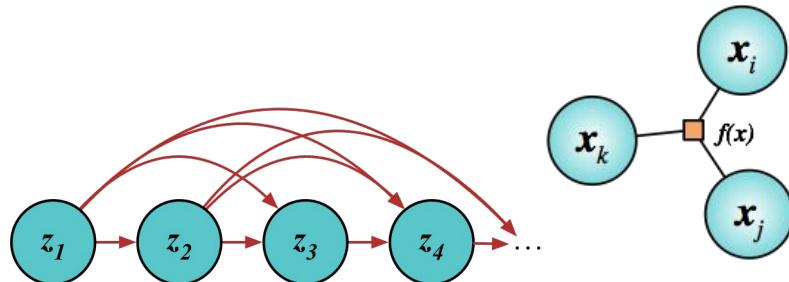
2. Принципи на обучение

Видове генеративни модели



Пълно-наблюдаеми модели

Данни наблюдавани директно с модела, без нужда от въвеждане на нови ненаблюдавани локални променливи.

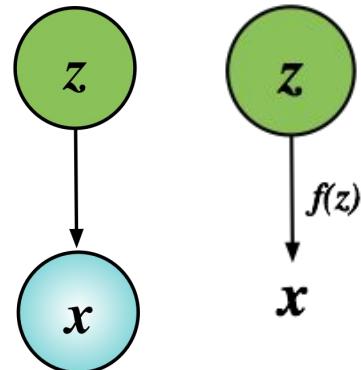


Модели с латентни променливи

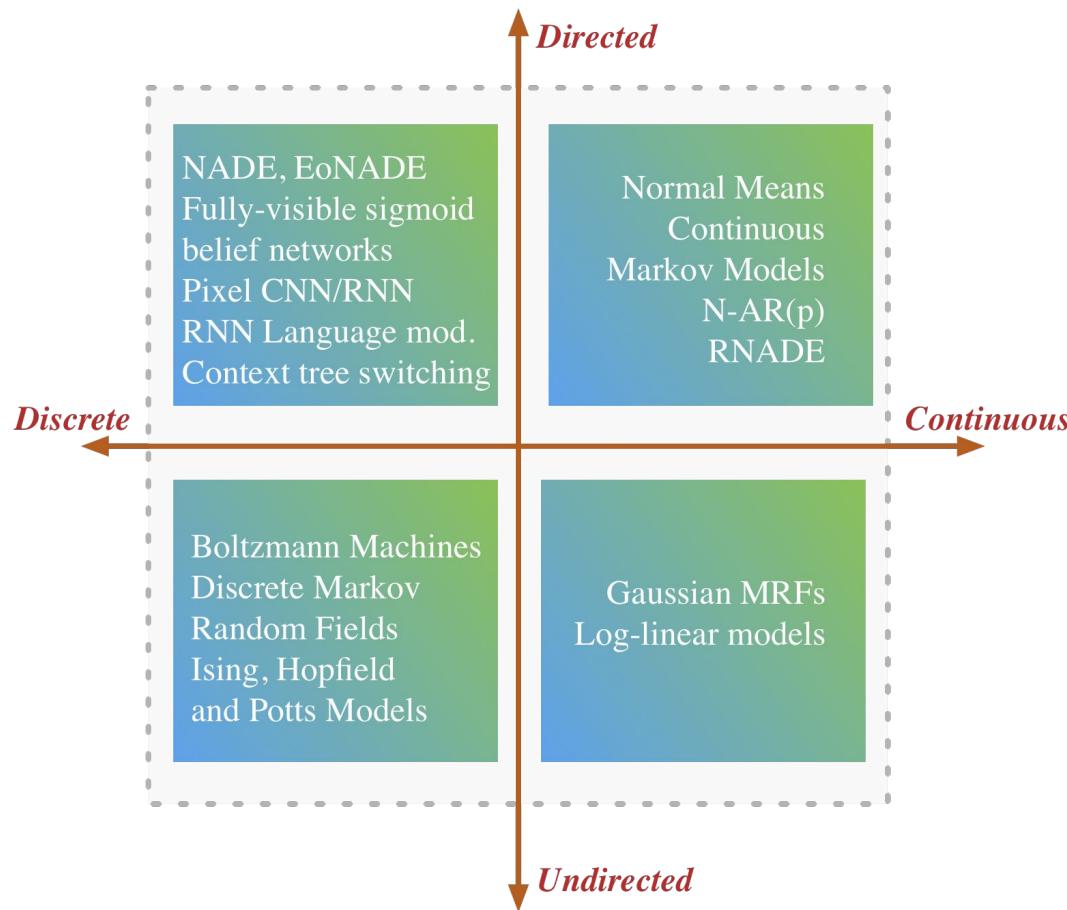
Въвеждане на ненаблюдавана произволна променлива за всяка наблюдавана точка от данни, за да се обяснят скритите причини.

- **Предписани модели:** използване на вероятностите на наблюдателя и прием на шум от наблюдението.
- **Имплицитни модели:** модели без близки подобия.

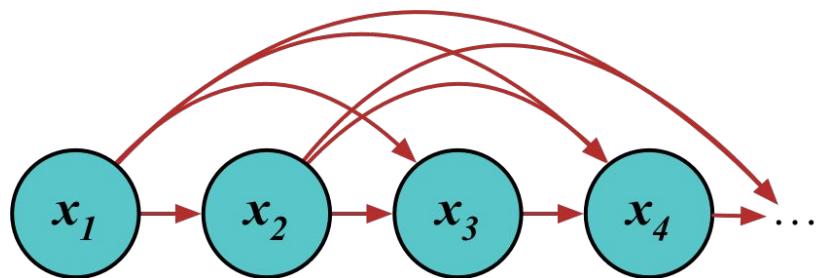
1. Модели



Спектъра на Напълно наблюдаваните модели

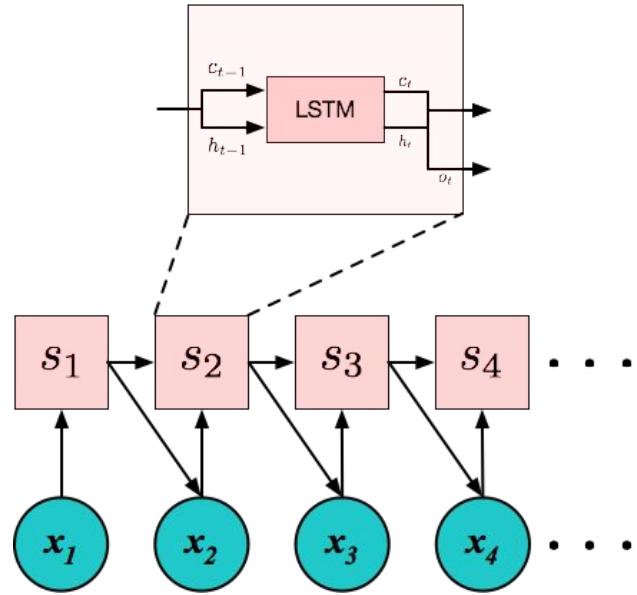


Изграждане на генеративни модели



$$p(x_{1,\dots,N}) = \prod_{i=1}^N p(x_i | x_{1,\dots,(i-1)})$$

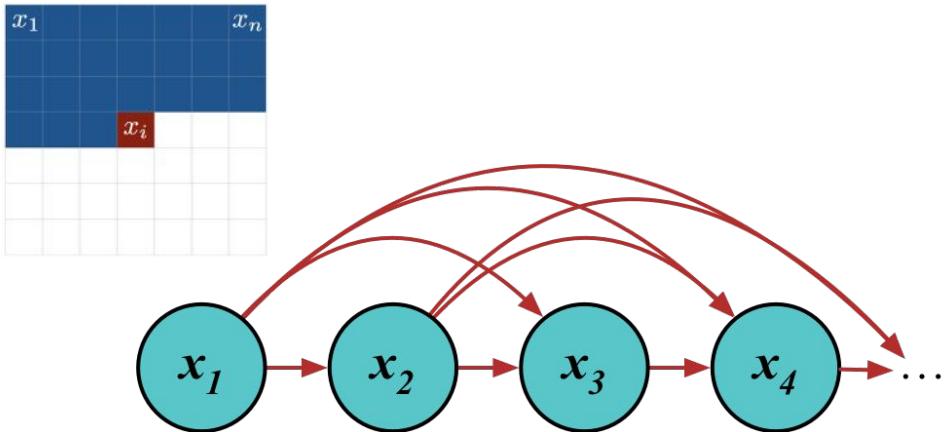
$$p(x_{1,\dots,N}) = \prod_{i=1}^N p(x_i | s_i(s_{i-1}, x_{i-1}))$$



Еквивалентни начини за представяне на едни и същи насочени ациклични графи

Напълно наблюдавани модели

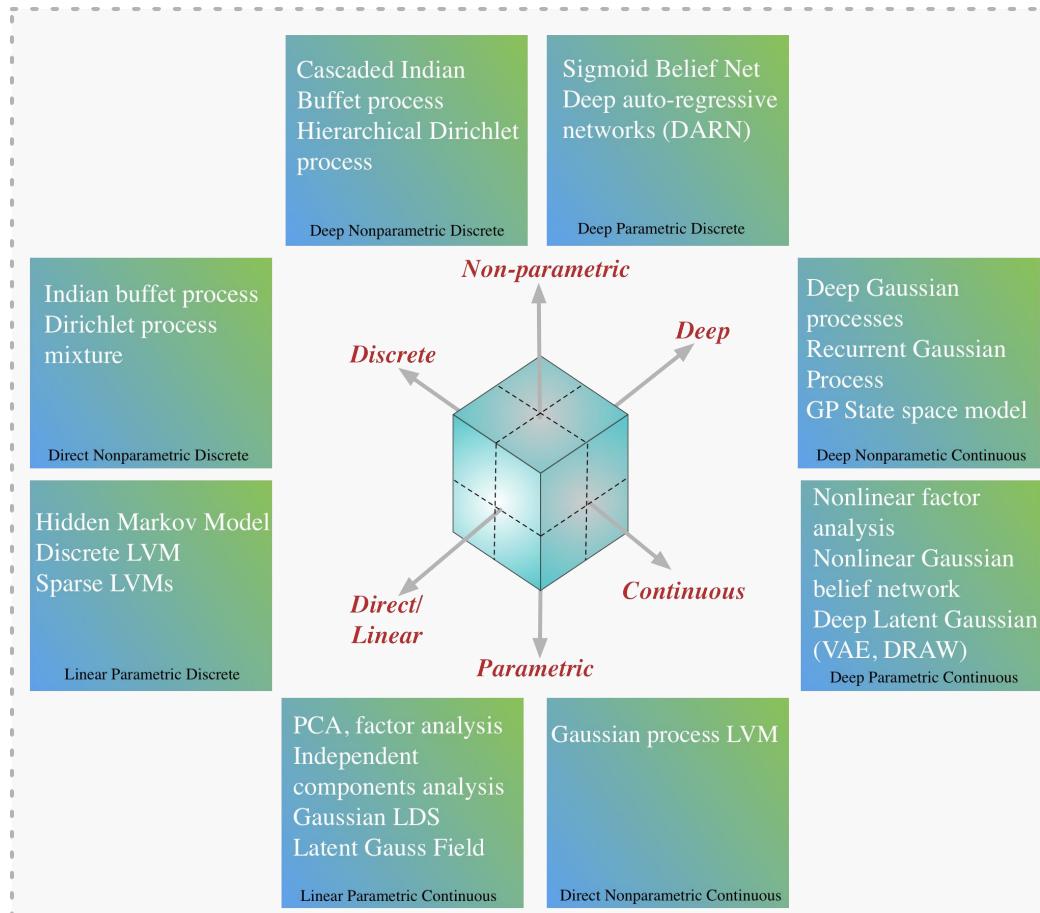
$$p(x_1, \dots, N) = \prod_{i=1}^N p(x_i | x_1, \dots, (i-1))$$



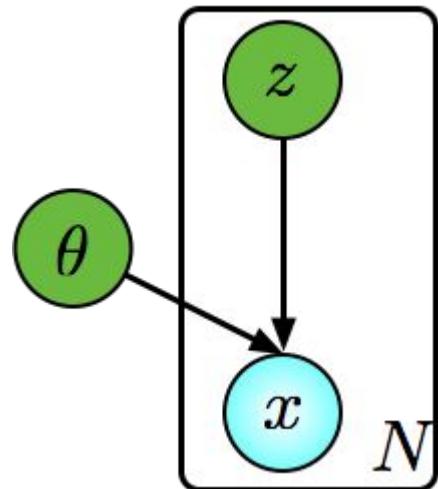
Всички условни вероятности, описани от дълбоки мрежи.

- + Может директно да кодира как са свързани наблюдаваните точки.
- + Может да се използва всеки тип данни
- + За насочени графични модели: запомнянето на параметри е просто
- + Log-сходство се изчислява директно, не е необходимо приближение.
- + Лесен за скалиране към по-големи модели, налични са много инструменти за оптимизация.
- Чувствителност към наредбата.
- За неориентирани модели, ученето на параметри е трудно: Необходимо е да се изчислят нормализиращи константи.
- Генерирането може да бъде бавно: последователно итериране на елементи или с помощта на Марковски процеси.

Спектър на моделите с латентни променливи



Изграждане на генеративни модели



$$p(x, z, \theta) = \rho(\theta) \prod_{i=1}^N p(x_i | z_i, \theta) \pi(z_i)$$

$$\pi(z) = \mathcal{N}(0, \mathbb{I}_{d_z})$$

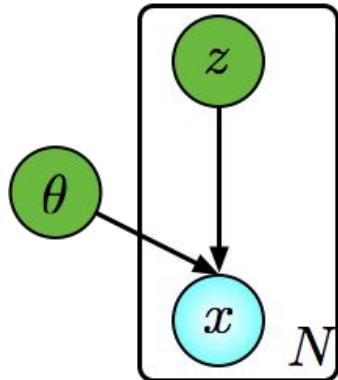
$$\rho(\theta) = \mathcal{N}(0, \kappa^2 \mathbb{I}_{d_\theta})$$

$$p(x | z, \theta) = \mathcal{N}(\theta_0 + \theta_1 z, \exp(\theta_2))$$

$$\theta = \{\theta_0 \in \mathbb{R}^{d_x}, \theta_1 \in \mathbb{R}^{d_x \times d_z}, \theta_2 \in \mathbb{R}^{d_x}\}$$

Изграждане на генеративни модели

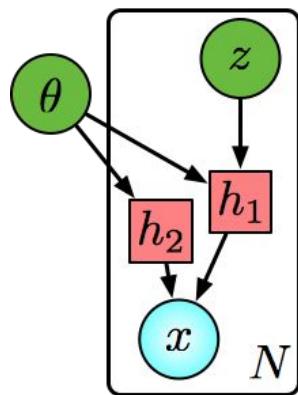
Графични модели + изчислителни графи (Невронни Мрежи)



$$\pi(z) = \mathcal{N}(0, \mathbb{I}_{d_z})$$

$$\rho(\theta) = \mathcal{N}(0, \kappa^2 \mathbb{I}_{d_\theta})$$

$$p(x|z, \theta) = \mathcal{N}(\theta_0 + \theta_1 z, \exp(\theta_2))$$



$$\pi(z) = \mathcal{N}(0, \mathbb{I}_{d_z})$$

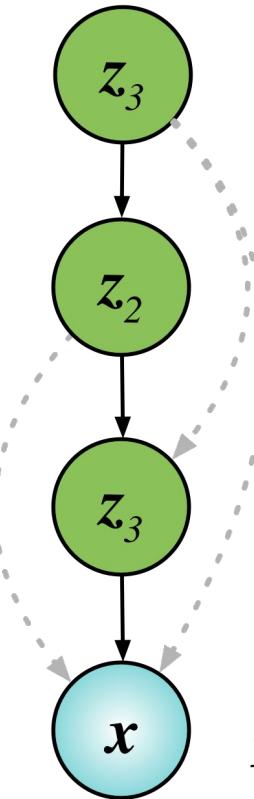
$$\rho(\theta) = \mathcal{N}(0, \kappa^2 \mathbb{I}_{d_\theta})$$

$$h_1 = \theta_0 + \theta_1 z$$

$$h_2 = \exp(\theta_2)$$

$$p(x|z, \theta) = \mathcal{N}(h_1, h_2)$$

Модели с латентни променливи



- Процесът на инверсия за определяне на латентите, съответстващи на даден вход, е труден като цяло
- Трудно е да се изчисли маргинализираната вероятност, изискаща приближения.
- Не е лесно да се посочат богати приближения за латентно задно разпределение.

$$p(x, z, \theta) = \rho(\theta) \prod_{i=1}^N p(x_i | z_i, \theta) \pi(z_i)$$

- + Лесно семплиране.
- + Лесен начин за включване на йерархия и дълбочина.
- + Лесна за кодиране структура
- + Избягва предположенията за зависимост от наредба: маргинализацията предизвиква зависимости.
- + Осигуряват компресия и представителност.
- + Възможност за оценяване, сравнение на модели и избор с помощта на маргинализирани сходства.

Въведат се ненаблюдавани локални произволни променливи, които представляват скрити зависимости.

Избор на принципи на обучение

За даден модел има много конкуриращи се методи за изводи.

- Точни методи (конюгация, изброяване)
- Числова интеграция (квадратура)
- Обобщен метод на моментите
- Максимално сходство (ML)
- Максимална апостериори (MAP)
- Приближение на Лаплас
- Интегрирани вложени приближения на Лаплас
- Максимизиране на очакванията (EM)
- Методи на Монте Карло (MCMC, SMC, ABC)
- Контрастна оценка (NCE)
- Методи за кухина (EP)
- Вариационни методи



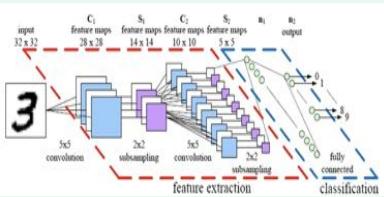
2. Принципи на обучение

Комбиниране на модели и изводи

3. Алгоритми

Даден модел и принцип на обучение могат да бъдат приложени по много начини.

Конволовионна невронна мрежа + наказана максимална вероятност



- Оптимизационни методи (SGD, Adagrad)
- Регуляризация (L1, L2, пакетна нормализация, отпадания)

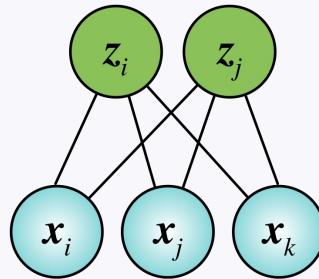
Неявен генеративен модел + Тестване с два образца



- Метод на моменти
- Приближително Бейсово изчисление (ABC)
- Генеративна съревнователна представителна мрежа (GAN)

Модел на латентна променлива + вариационен извод

-
- A diagram showing three circles: a green circle θ , a green circle z , and a blue circle x . Arrows point from θ to x and from z to x .
- VEM алгоритъм
 - Очаквано разпространение
 - Приближително предаване
 - Вариационни автоЕнкодери (VAE)



- Контрастна дивергенция
- Постоянна контр. дивергенция
- Паралелно темпериране
- Естествени градиенти

Въпроси към изводите?

Цел	Количество на интерес
Прогноза	$p(x_{(t+1), \dots, \infty} x_{-\infty, \dots, t})$
Планиране	$J = \mathbb{E}_p \left[\int_0^{\infty} dt C(x_t) \middle x_0, u \right]$
Оценка на параметри	$p(\theta x_{0, \dots, N})$
Експериментален дизайн	$EIG = D[p(f(x_{t, \dots, \infty}) u); p(f(x_{-\infty, \dots, t}))]$
Тестване на хипотези	$\frac{p(f(x_{-\infty, \dots, t}) H_0)}{p(f(x_{-\infty, \dots, t}) H_1)}$

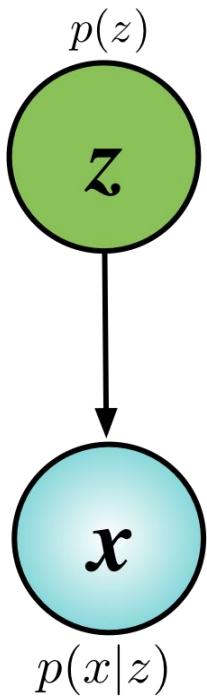
Модели с латентни променливи

$$x \in \mathbb{R}^{d_x} \quad z \in \mathbb{R}^{d_z} \quad \theta \in \mathbb{R}^{d_\theta}$$

$$\mathcal{D} = \{x_i\} \quad i \in \{1, \dots, N\}$$

$$\log p_\theta(x) = \log \int p_\theta(x|z)p(z)dz = \log \mathbb{E}_{p(z)}[p_\theta(x|z)]$$

$$\log p_\theta(\mathcal{D}) = \sum_{i=1}^N \log \mathbb{E}_{p(z)}[p_\theta(x_i|z)]$$



Методи за приблизителни изводи и заключения

- **Апроксимации на Лаплас**
- **Извадка по важност**
- **Вариационни приближения**
- **Корекции на външните параметрични смущения**
- Други методи: MCMC, Langevin, HMC, Adaptive MCMC

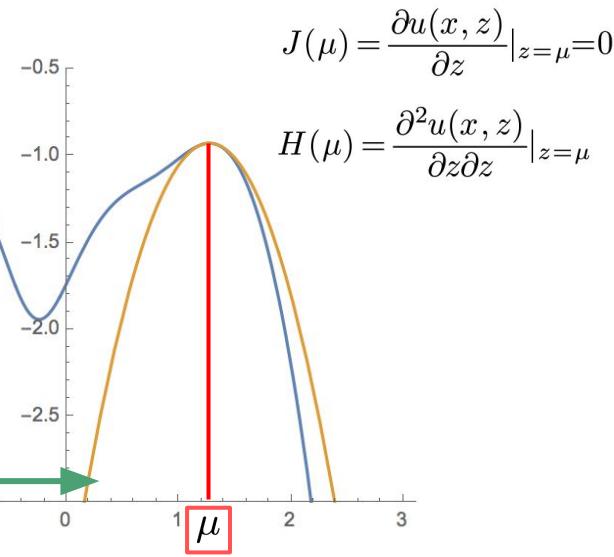
Апроксимация на Лаплас

$$\begin{aligned}\log \mathbb{E}_{p(z)}[p_\theta(x|z)] &= \log \int p_\theta(x|z)p(z)dz \\ &= \log \int e^{-u(x,z)}dz\end{aligned}$$

$$u(x, z) = -\log p_\theta(x|z)p(z)$$

$$u(x, z) \approx u(x, \mu) + \frac{1}{2}(z - \mu)^T H(\mu)(z - \mu)$$

$$\begin{aligned}\log \mathbb{E}_{p(z)}[p_\theta(x|z)] &\approx \log \int e^{-u(x, \mu) - \frac{1}{2}(z - \mu)^T H(\mu)(z - \mu)} dz \\ &= -u(x, \mu) - \frac{1}{2} \ln \det(2\pi H^{-1}(\mu))\end{aligned}$$



Известна е и като:
Приближение към седлова точка,
Делта-метод

Извадка по важност

$$\begin{aligned}\log p(x_i) &= \log \mathbb{E}_{p(z)}[p_\theta(x_i|z)] \\ &= \log \mathbb{E}_{q_\phi(z|x_i)}\left[\frac{p_\theta(x_i|z)p(z)}{q_\phi(z|x_i)}\right] \\ &= \log \mathbb{E}_{q_\phi(z|x_i)}[e^{-\mathcal{F}(x_i, z)}] \\ &\approx \log \sum_{k=1}^K e^{-\mathcal{F}(x_i, z_k)} - \log K\end{aligned}$$

$$\mathcal{F}(x, z) = \ln q(z|x) - \ln p(z) - \ln p(x|z)$$

$$\log p(x) \geq \mathbb{E}_{q_\phi(z|x_i)}\left[\log \sum_{k=1}^K e^{-\mathcal{F}(x_i, z_k)}\right] - \log K$$

Важност на теглата

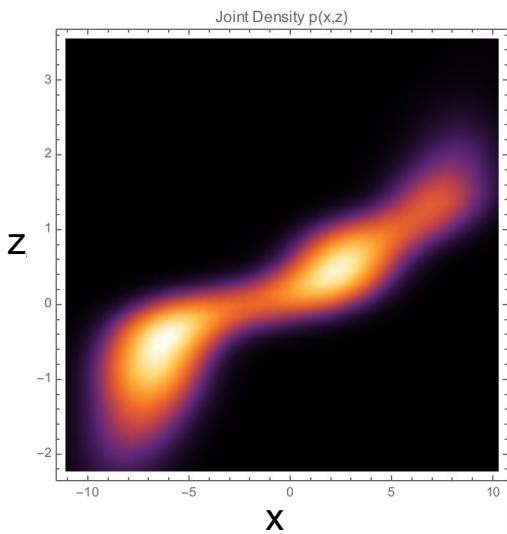
Monte-Carlo

Pointwise Free-energy

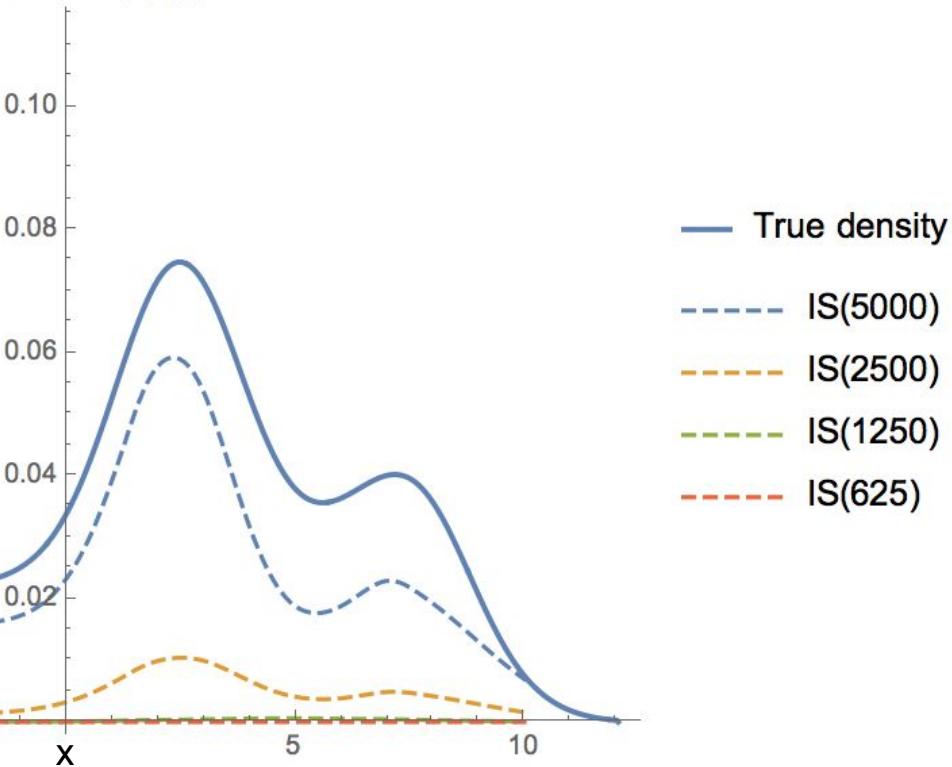
Важни качества

Извадка по важност осигурява граница на очакванията

$$\log p(x) \geq \mathbb{E}_{q_\phi(z|x)} \left[\log \sum_{k=1}^K e^{-\mathcal{F}(x, z_k)} \right] - \log K$$

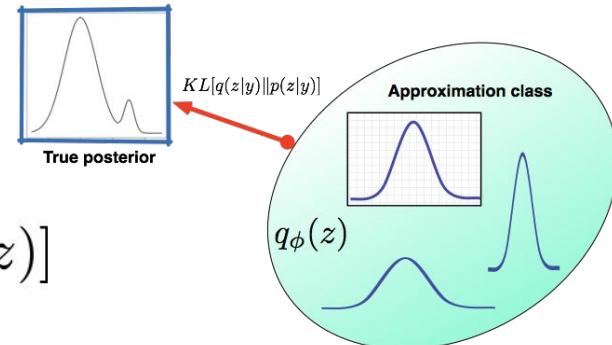


Marginal Density $p(x)$



Вариационни приближения

$$\log p_\theta(\mathcal{D}) = \sum_{i=1}^N \log \mathbb{E}_{p(z)}[p_\theta(x_i|z)]$$



$$\log \mathbb{E}_{p(z)}[p_\theta(x_i|z)] = \log \mathbb{E}_{q_i(z)} \left[\frac{p_\theta(x_i|z)p(z)}{q_i(z)} \right], \quad \forall q_i > 0$$

$$\log \mathbb{E}_{q_i(z)} \left[\frac{p_\theta(x_i|z)p(z)}{q_i(z)} \right] \geq \mathbb{E}_{q_i(z)} \left[\log \frac{p_\theta(x_i|z)p(z)}{q_i(z)} \right]$$

$$\log p_\theta(\mathcal{D}) \geq \sum_{i=1}^N \mathbb{E}_{q_i(z)} \left[\log \frac{p_\theta(x_i|z)p(z)}{q_i(z)} \right]$$

Вариационни приближения

$$\log p_{\theta}(\mathcal{D}) \geq \sum_{i=1}^N \mathbb{E}_{q_i(z)} \left[\log \frac{p_{\theta}(x_i|z)p(z)}{q_i(z)} \right]$$

$$\mathbb{E}_{q_i(z)} \left[\log \frac{p_{\theta}(x_i|z)p(z)}{q_i(z)} \right] = \boxed{\mathbb{E}_{q_i(z)} [\log p_{\theta}(x_i|z)]} - \boxed{\text{KLD}(q_i \| p)}$$

Реконструкция Регуляризация

Корекции на външните параметрични смущения

$$\begin{aligned} \log \mathbb{E}_{p(z)}[p_{\theta}(x|z)] &= \log \int e^{-u(x,z)} dz \\ &= -\mathcal{F}(x) + \log \mathbb{E}_{q(z|x)}[e^{\Delta(x,z)}] \\ &= -\mathcal{F}(x) + \log \mathbb{E}_{q(z|x)}\left[\sum_{k=0}^{\infty} \frac{\Delta(x,z)^k}{k!}\right] \\ &= -\mathcal{F}(x) + \log \sum_{k=0}^{\infty} \frac{1}{k!} \mathbb{E}_{q(z|x)}[\Delta(x,z)^k] \end{aligned}$$
$$\begin{aligned} \mathcal{F}(x, z) &= \ln q(z|x) + u(x, z) \\ \mathcal{F}(x) &= \mathbb{E}_{q(z|x)}[\mathcal{F}(x, z)] \\ \Delta &= -\mathcal{F}(x, z) + \mathcal{F}(x) \\ e^y &= \sum_{k=0}^{\infty} \frac{y^k}{k!} \end{aligned}$$

Избори в процеса на дизайна

Избор на Модел

Изчислителни графики, рендери, симулатори и работни среди

Вариационна оптимизация

- Вариационни EM
- Стохастични VEM
- Monte Carlo градиентни оценки

Приблизителни приближения

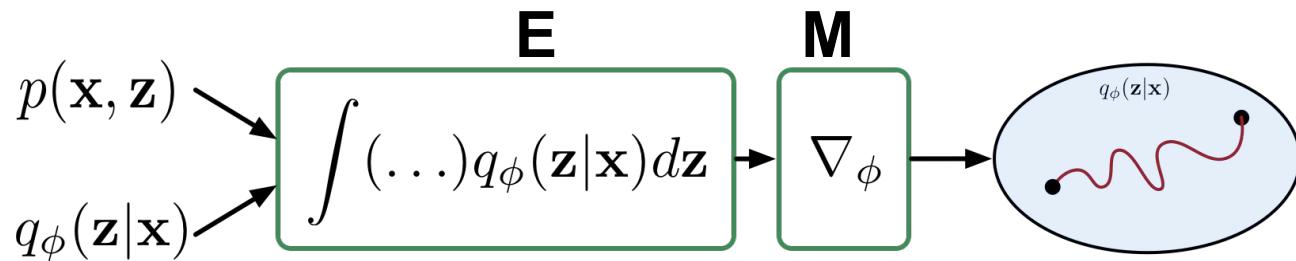
- Усреднени полета
- Структурирани приближения
- Спомагателни променливи методи

Вариационен ЕМ алгоритъм

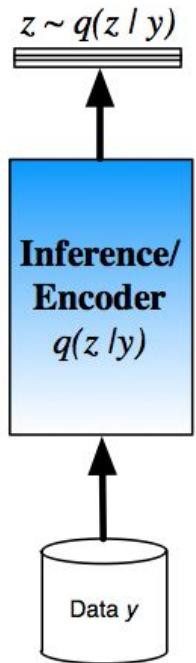
Итерации с фиксирана точка между вариационни и моделни параметри

$$\mathbf{E} \quad q_i^*(z) = \operatorname{argmax}_{q_i} \mathbb{E}_{q_i^*(z)} \left[\log \frac{p_\theta(x_i|z)p(z)}{q_i^*(z)} \right] \Leftrightarrow q_i^*(z) = \frac{p_\theta(x_i|z)p(z)}{p(x_i)}$$

$$\mathbf{M} \quad \theta^* = \operatorname{argmax}_\theta \sum_{i=1}^N \mathbb{E}_{q_i^*(z)} \left[\log \frac{p_\theta(x_i|z)p(z)}{q_i^*(z)} \right]$$



Амортизирани изводи



$$q_i^*(z) = \operatorname{argmax}_{q_i} \mathbb{E}_{q_i^*(z)}[-\mathcal{F}(x_i, z)]$$

Въвеждане на параметрично семейство от условни плътности

$$\operatorname{argmax}_{q_i} \mathbb{E}_{q_i^*(z)}[-\mathcal{F}(x_i, z)] \Rightarrow \operatorname{argmax}_\phi \mathbb{E}_{q_\phi(z|x)}[-\mathcal{F}_\phi(x_i, z)]$$

Вариационни автоЕнкодери

Най-простата инстанция на вариационни автоЕнкодери

Дълбок латентен гаусов модел $p(x, z)$

$$\text{prior sample} \quad z \sim \mathcal{N}(0, \mathbb{I})$$

$$\text{data sufficient statistics} \quad \eta = f_\theta(z)$$

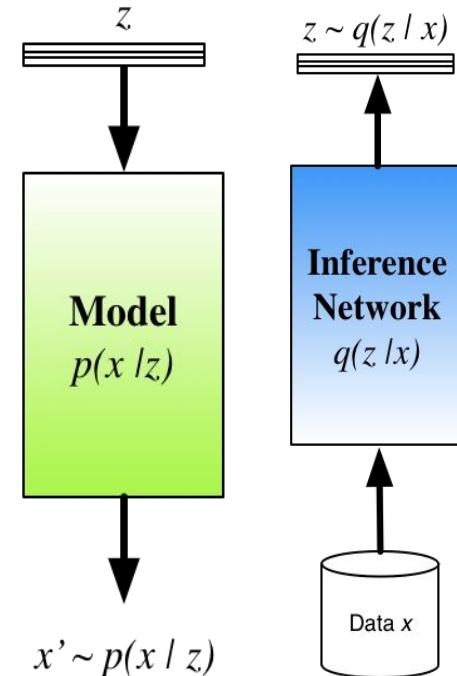
$$\text{data conditional likelihood} \quad x \sim \mathcal{N}(\eta)$$

Модел на Гаусово разпознаване $q(z)$

$$\text{data sample} \quad x \sim \mathcal{D}$$

$$\text{latent sufficient statistics} \quad \eta = f_\phi(x)$$

$$\text{posterior sample} \quad z \sim \mathcal{N}(\eta)$$

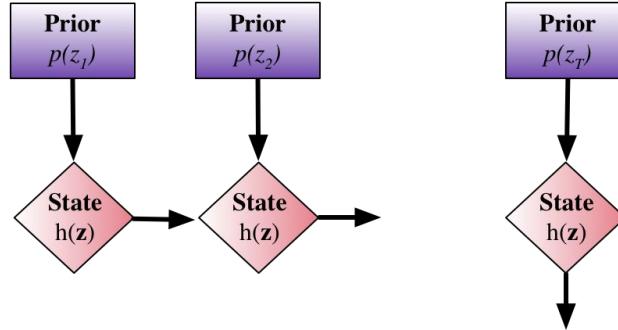


$$\mathbb{E}_{q_i(z)}[\log p_\theta(x_i | z)] - \text{KLD}(q_i \| p)$$

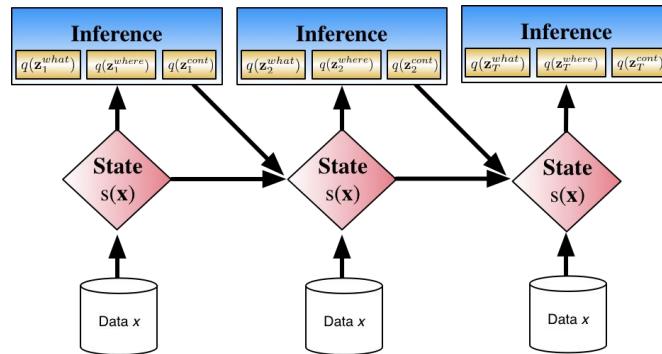
След това оптимизираме модела със свободна енергия и вариационните параметри

По-богати вариационни автоенкодери

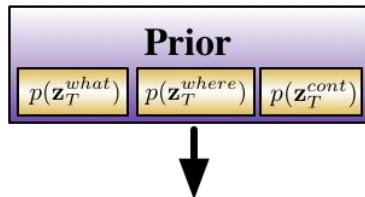
DRAW: Рекурентни/Зависими от предходните



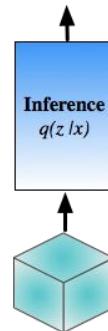
Рекурентни/Зависими мрежи за изводи



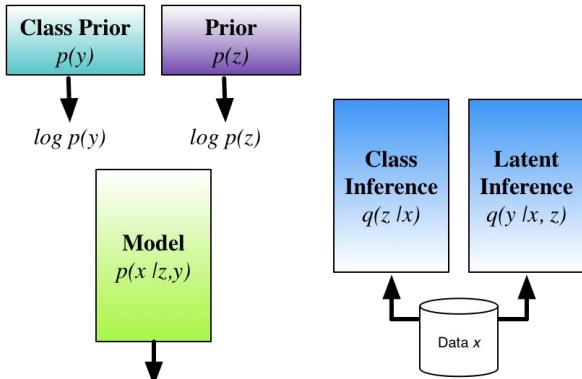
AIR: Структурирани предходни



Обемни и
последователни данни

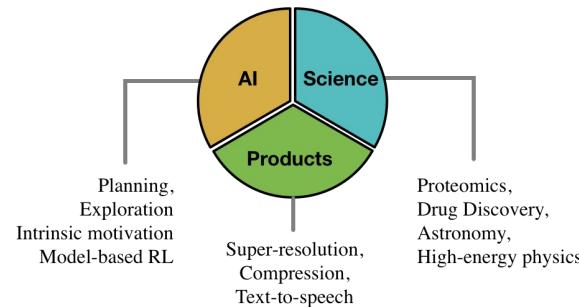


Полуконтролирано обучение

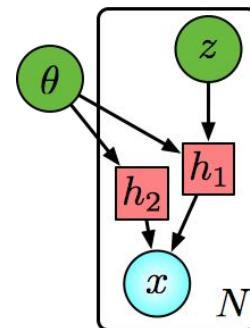


Обобщение до тук

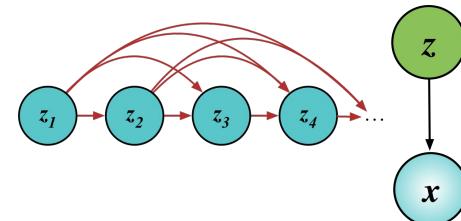
Приложения на генеративни модели



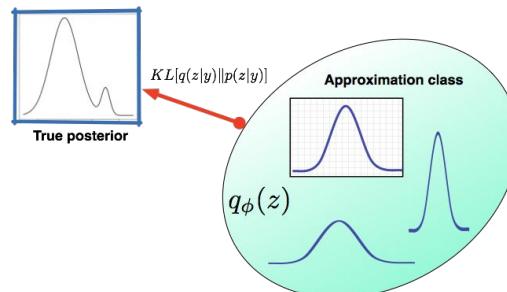
Вероятностно дълбоко обучение



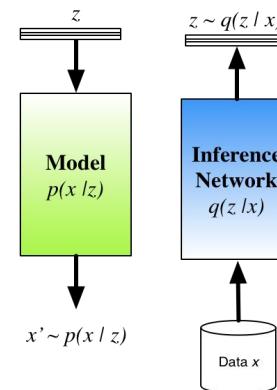
Видове генеративни модели



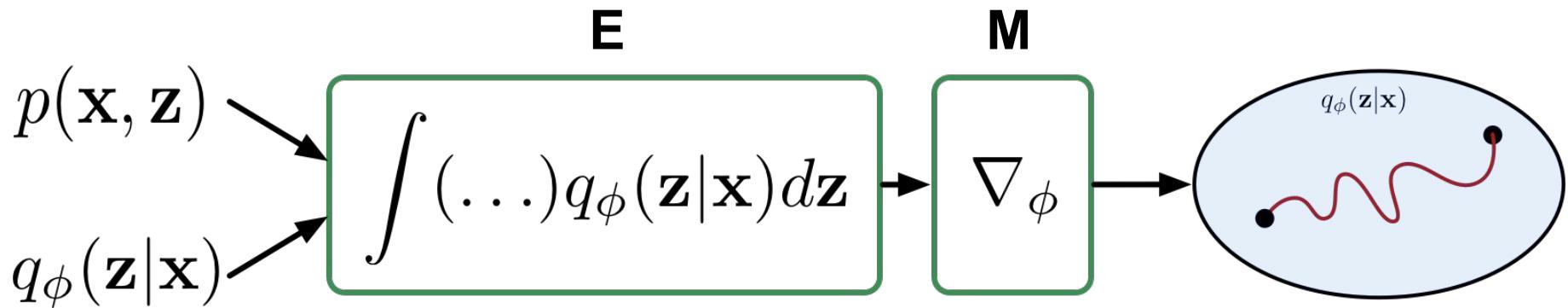
Вариационни принципи



Амортизирани изводи

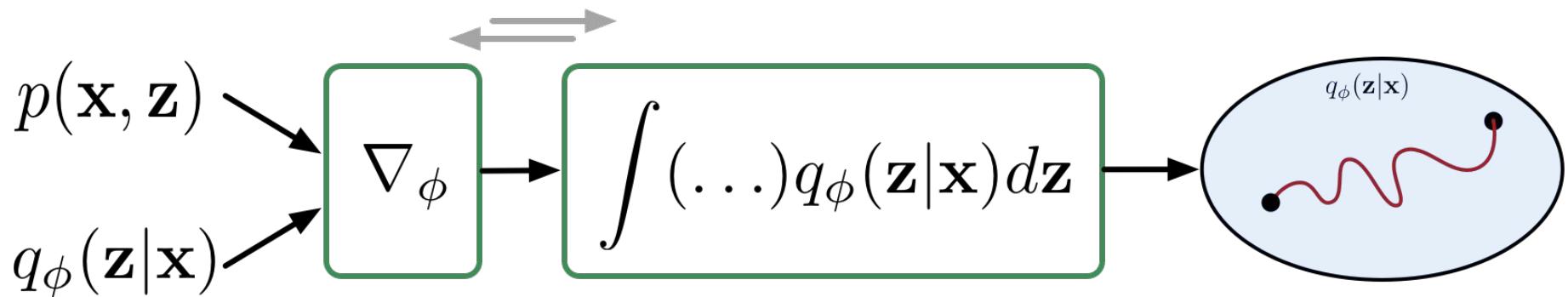


Класически изведен подход



Изчислете очакванията, след това М-стъпковите градиенти

Стохастичен подход за изводи



Като цяло няма как да знаем очакванията.

Градиентът е на параметрите на разпределението, по отношение на което е взето очакването .

Стохастични градиентни оценители

$$\nabla_{\phi} \mathbb{E}_{q_{\phi}(\mathbf{z})}[f_{\theta}(\mathbf{z})] = \nabla \int [q_{\phi}(\mathbf{z}) | f_{\theta}(\mathbf{z})] d\mathbf{z}$$

Оценка на функцията за оценка:

Разграничете плътността $q(\mathbf{z}|\mathbf{x})$

Оценка на градиента по пътя:

Разграничете функцията $f(\mathbf{z})$

Типични зони с проблеми:

- Генеративни модели и изводи
- Подсилено обучение и управление
- Изследване на операции и контрол на инвентар
- Монте Карло симулация
- Финанси и ценообразуване на активи
- Оценка на чувствителността

Оценка на функцията оценители

$$\nabla_{\phi} \mathbb{E}_{q_{\phi}(\mathbf{z})}[f_{\theta}(\mathbf{z})] = \nabla \int q_{\phi}(\mathbf{z}) f_{\theta}(\mathbf{z}) d\mathbf{z}$$

$$= \mathbb{E}_{q(z)}[f_{\theta}(\mathbf{z}) \nabla_{\phi} \log q_{\phi}(\mathbf{z})]$$

Градиентът се претегля от стойността на функцията

Извесно е още като:

- Likelihood-ratio trick
- Radon-Nikodym derivative
- REINFORCE and policy gradients
- Automated inference
- Black-box inference

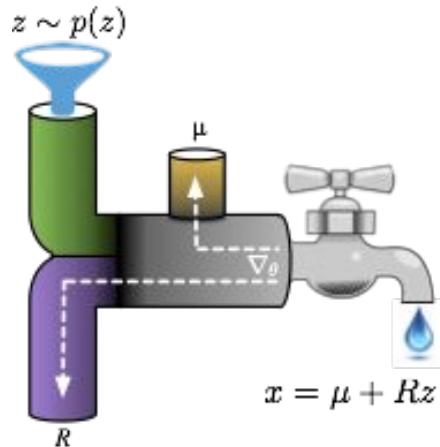
Кога се прилага:

- Функцията не е диференцируема.
- Разпределението q предоставя лесна възможност за извличане на данни от него.
- Плътноста q е известна и диференцируема.

Репараметризация

$$\nabla_{\phi} \mathbb{E}_{q_{\phi}(\mathbf{z})}[f_{\theta}(\mathbf{z})] = \nabla \int q_{\phi}(\mathbf{z}) \boxed{f_{\theta}(\mathbf{z})} d\mathbf{z}$$

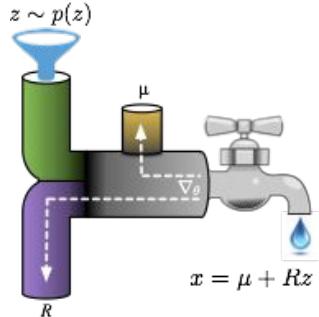
Намерете обратима функция $g(\cdot)$, която изразява z като трансформация на основно разпределение.



$$\mathbf{z} = g_{\phi}(\boldsymbol{\epsilon}) \quad \boldsymbol{\epsilon} \sim p(\boldsymbol{\epsilon})$$

$$\mathbb{E}_{q_{\phi}(z|x)}[f(z)] = \mathbb{E}_{p(\boldsymbol{\epsilon})}[f(g_{\phi}(x, \boldsymbol{\epsilon}))]$$

Оценител на производни по пътя



$$\mathbf{z} = g(\epsilon, \phi) \quad \epsilon \sim p(\epsilon)$$

$$\nabla_{\phi} \mathbb{E}_{q_{\phi}(\mathbf{z})} [f_{\theta}(\mathbf{z})] = \nabla \int q_{\phi}(\mathbf{z}) \boxed{f_{\theta}(\mathbf{z})} dz$$

$$= \mathbb{E}_{p(\epsilon)} [\nabla_{\phi} f_{\theta}(g(\epsilon, \phi))]$$

Известен още и като:

- Трик за препараметризация
- Стохастично обратно разпространение
- Анализ на смущенията
- Афинно-независим извод
- Двойно стохастична оценка
- Йерархични нецентрирани параметризации.

Кога да се прилага

- Функция f е диференцируема
- Плътността q може да бъде описана с помощта на по-просто базово разпределение: обратен CDF, преобразуване в мащаб на местоположение или друго координатно преобразуване.
- Лесен за вземане на проби от основното разпределение.

Гаусови стохастични градиенти

$$\nabla_{\phi} \mathbb{E}_{\mathcal{N}(\mu, CC^\top)}[f_\theta(\mathbf{z})]$$

Градиент от първи ред

$$p(\epsilon) = \mathcal{N}(0, 1) \quad g(\epsilon, \phi) = \mu_\phi(x) + C_\phi(x)\epsilon$$

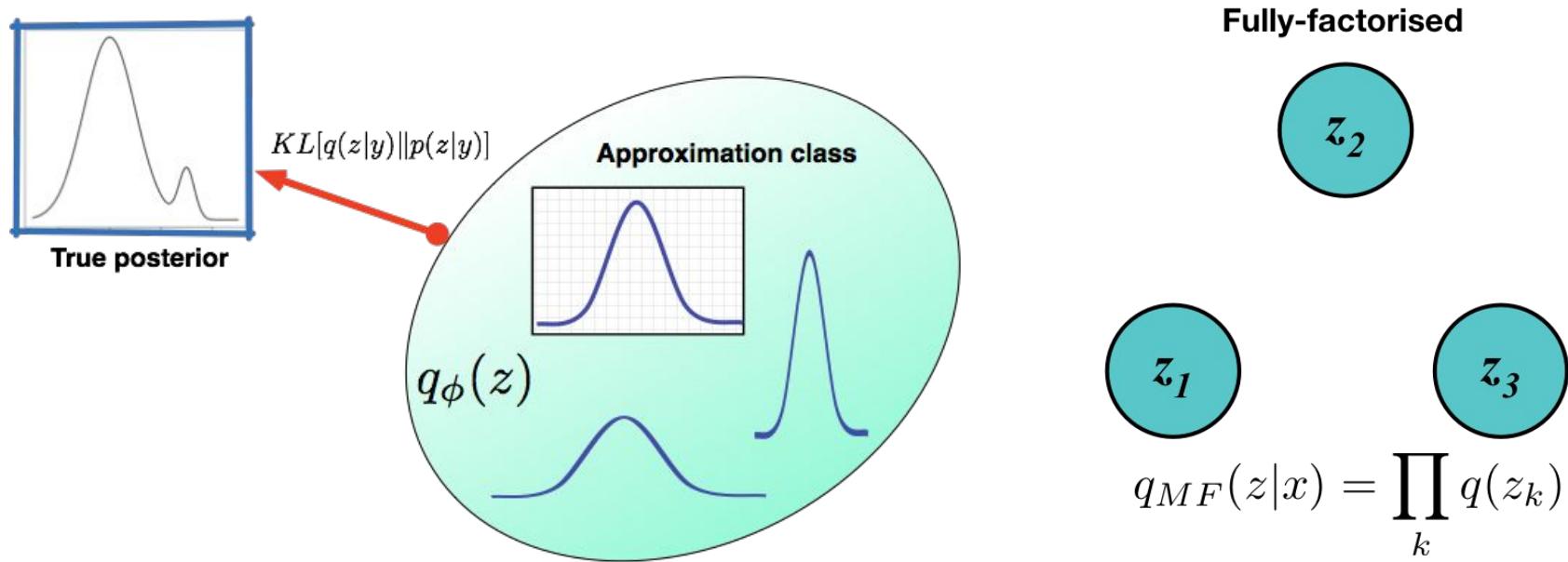
$$\mathbb{E}_{p(\epsilon)}[J^\top (\nabla_{\phi}\mu_\phi + \nabla_{\phi}C_\phi^\top \epsilon)]$$

Градиент от втори ред

$$\mathbb{E}_{q(z)}[J^\top \nabla_{\phi}\mu_\phi + Tr[HC_\phi \nabla_{\phi}C_\phi]]$$

Можем да разработим оценки с ниска вариация, като използваме
знанията за участващите разпределения, когато ги знаем

Апроксимации на усреднени полета

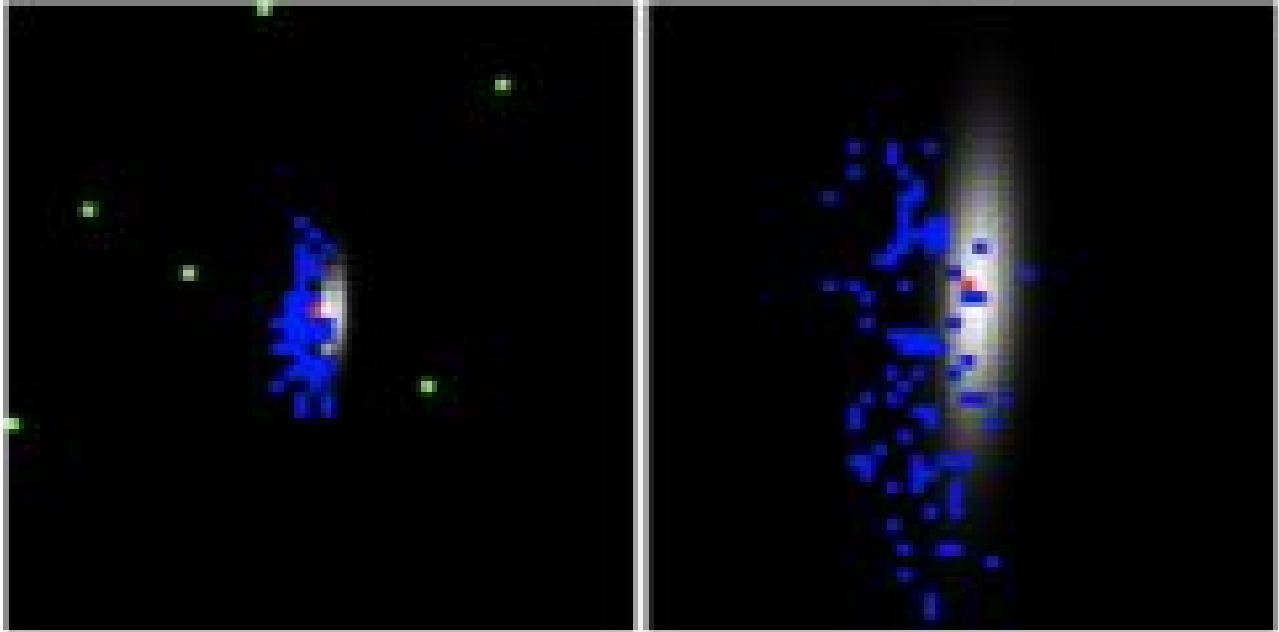
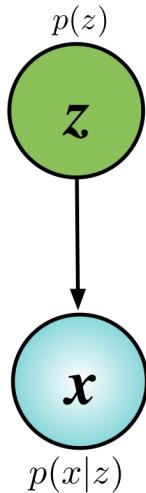


Ключова част от вариационния извод е изборът на приблизително постериорно разпределение q .

$$\mathcal{F}(q, \theta) = \mathbb{E}_{q_\phi(z|x)} [\log p_\theta(\mathbf{x}|\mathbf{z})] - KL[q_\phi(\mathbf{z}|\mathbf{x}) \| p(\mathbf{z})]$$

Последващи апроксимации на усреднени полета

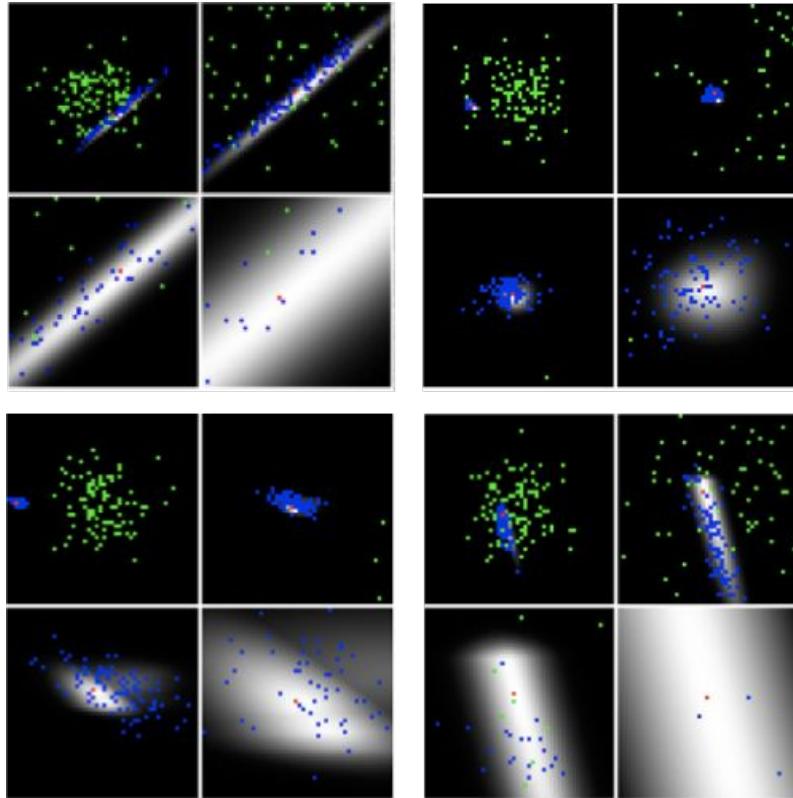
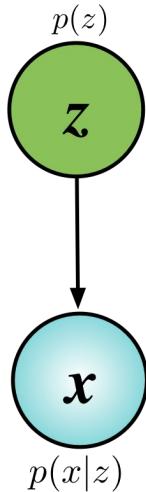
Дълбок
латентен
гаусов модел



Усреднени полета или напълно разложеното на множители последващо обикновено не е достатъчно

Последващи разпределения в реалния свят

Дълбок
латентен
гаусов модел

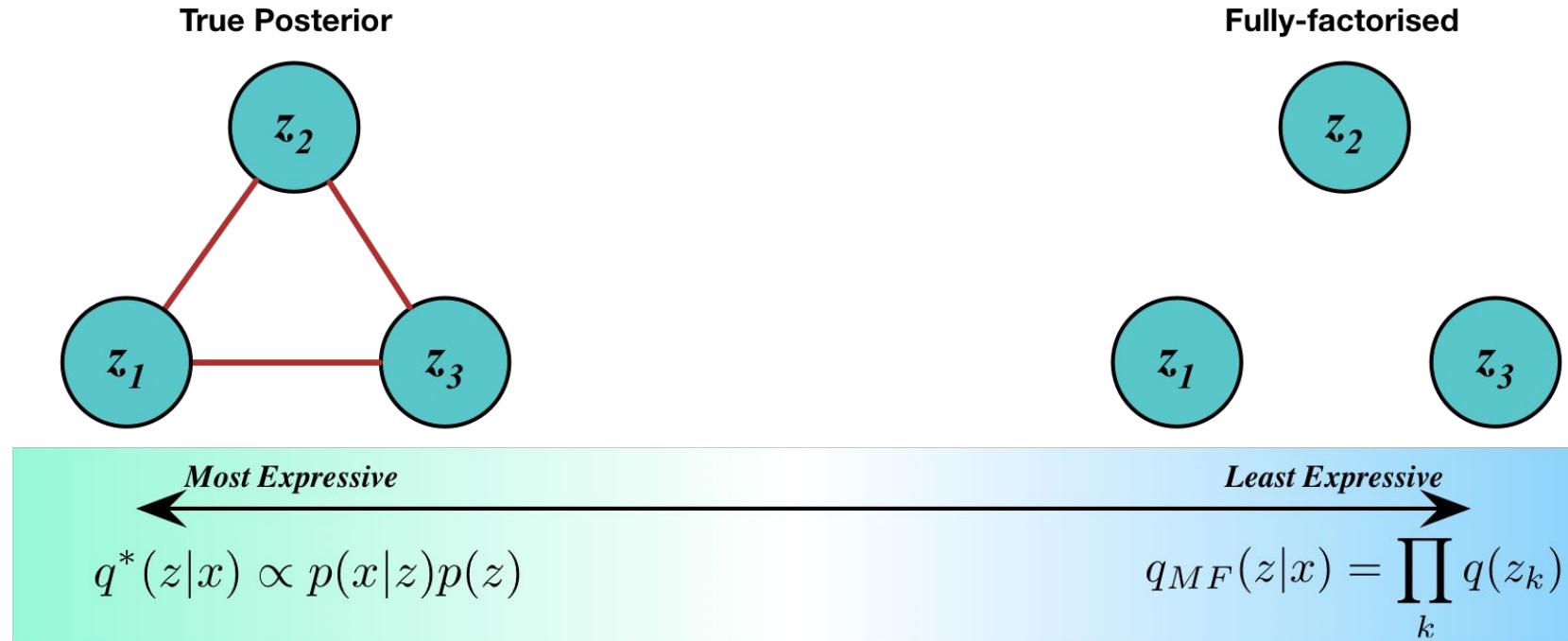


Сложни зависимости · Негаусови разпределения · Множество режими

По-богати класове на последващи разпределения

Две главни цели:

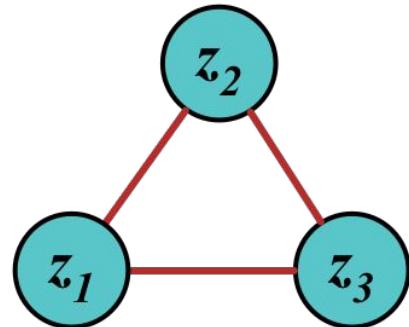
- Изграждане на по-богати приблизителни последващи разпределения.
- Поддържа изчислителна ефективност и мащабируемост.



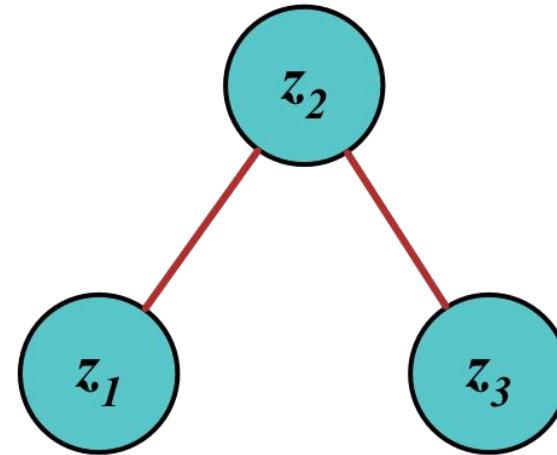
Същото като проблема за определяне на модел на самите данни

Структурирани апроксимации

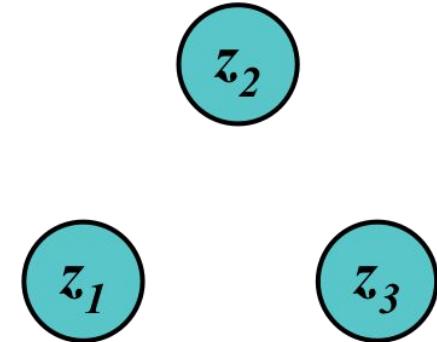
True Posterior



Structured Approx.



Fully-factorised



Most Expressive

$$q^*(z|x) \propto p(x|z)p(z)$$

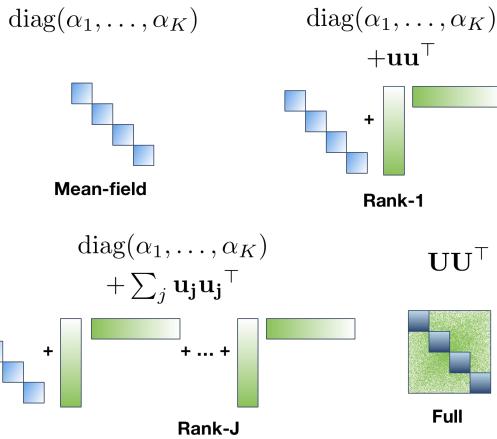
Least Expressive

$$q(z) = \prod_k q_k(z_k | \{z_j\}_{j \neq k})$$

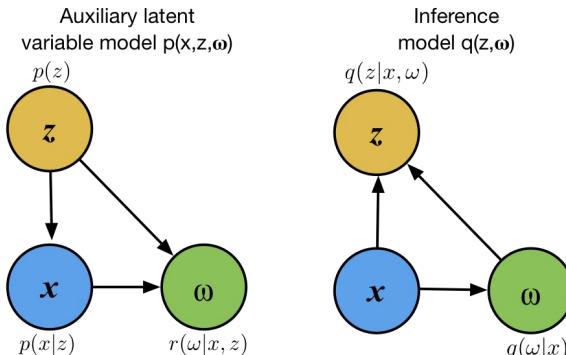
$$q_{MF}(z|x) = \prod_k q(z_k)$$

Групи на последващи разпределения

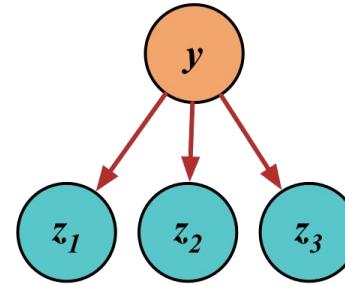
Ковариационни модели



Модели със спомагателни променливи

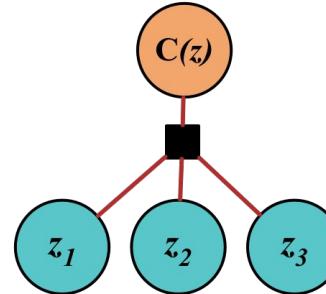


Mixture model



$$q_{mm}(\mathbf{z}; \boldsymbol{\nu}) = \sum_r \rho_r q_r(\mathbf{z}_r | \boldsymbol{\nu}_r)$$

Copula Methods



$$q_{lm}(\mathbf{z}; \boldsymbol{\nu}) = \left(\prod_k q_k(z_k | \boldsymbol{\nu}_k) \right) C(\mathbf{z}; \boldsymbol{\nu}_{k+1})$$

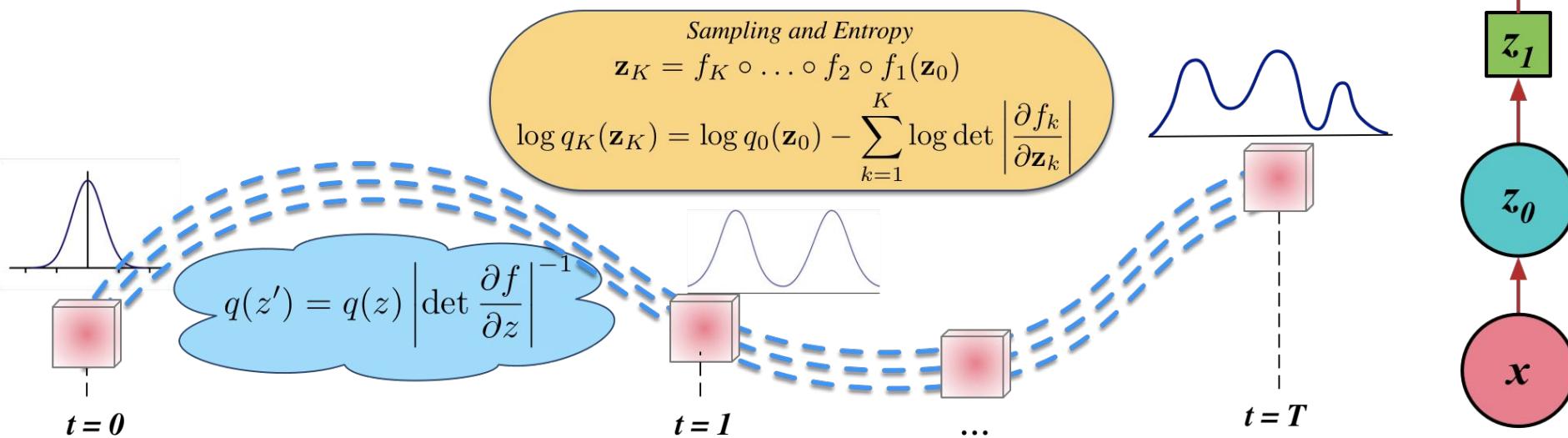
Нормализиране на потоците



Нормализиране на потоците

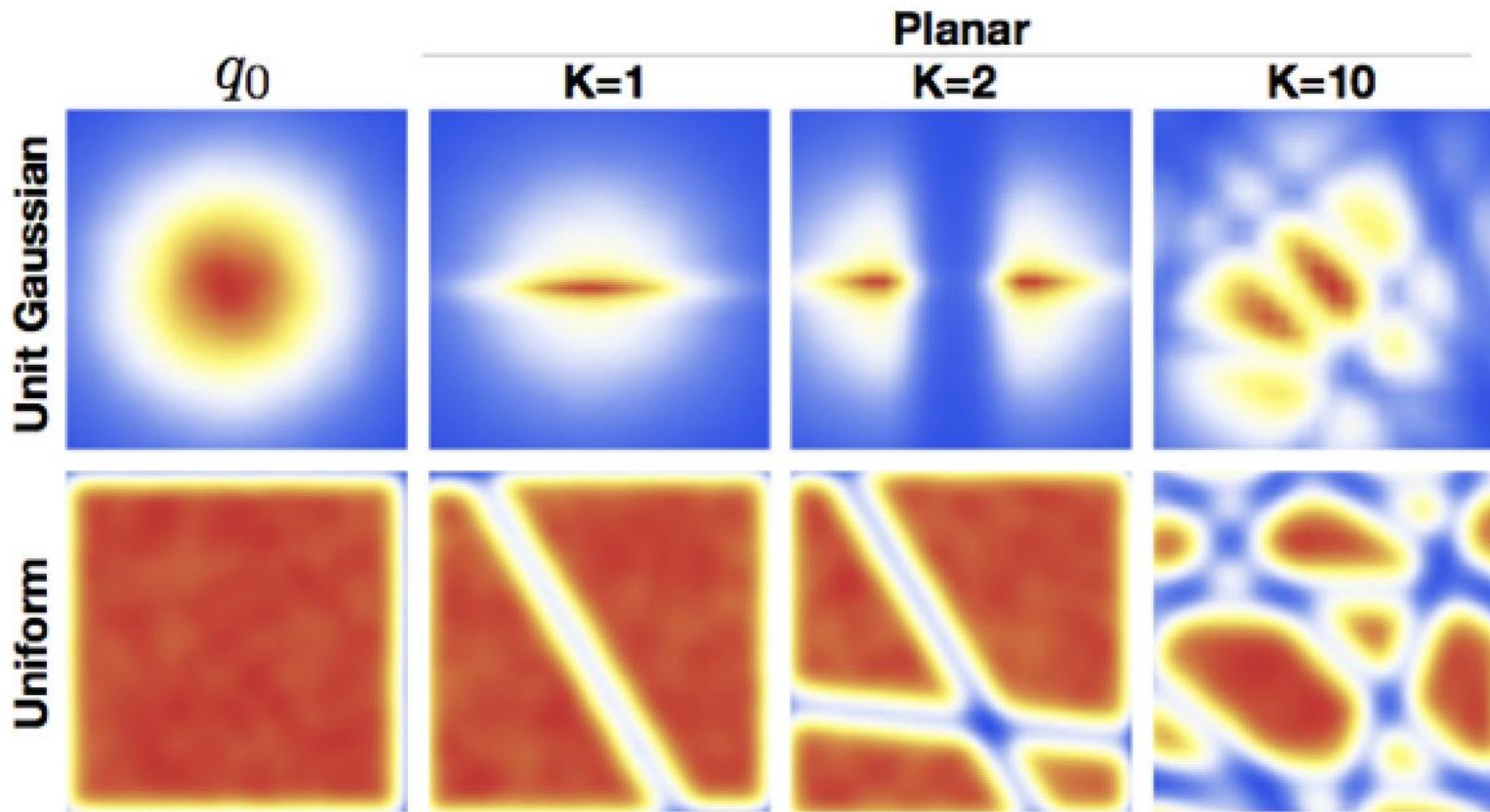
Използвайте правилото за промяна на променливите:

- Започнете с първоначално разпределение
- Приложете последователност от K обратими трансформации

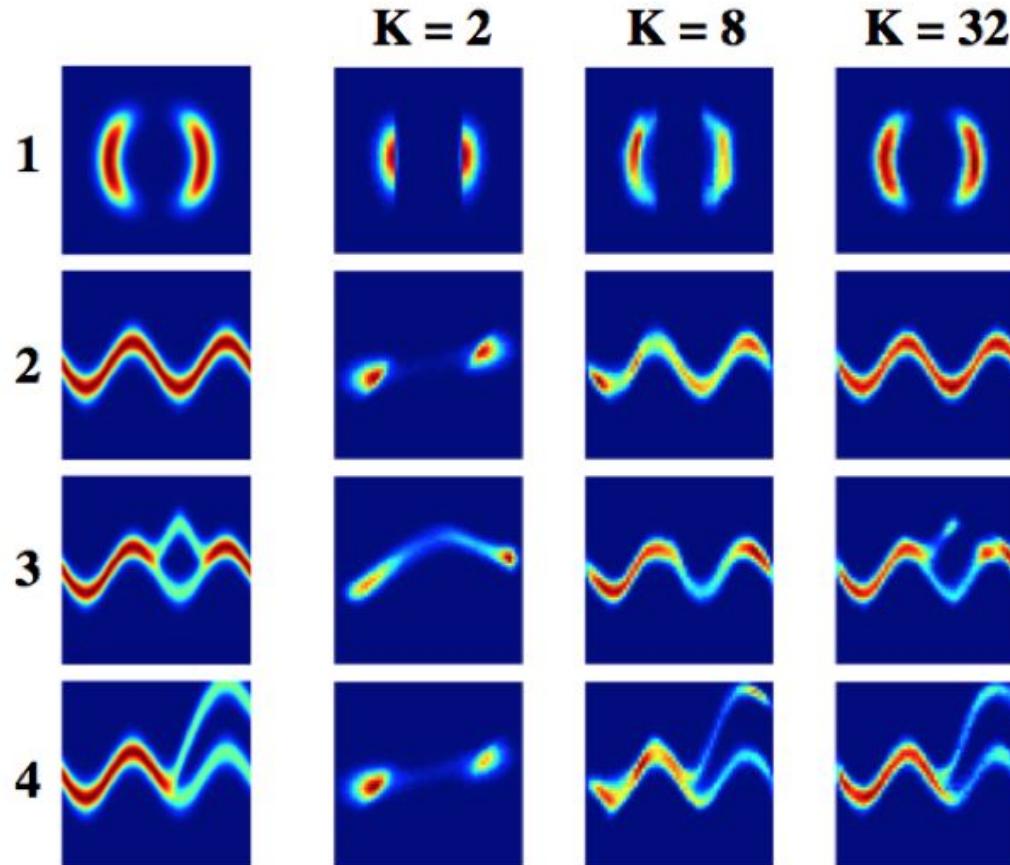


Разпределението преминава чрез последователност от обратими трансформации

Нормализиране на потоците



Нормализиране на потоците



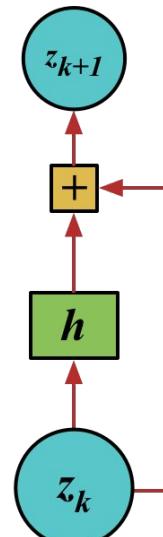
Избор на трансформация

$$\mathcal{L} = \mathbb{E}_{q_0(\mathbf{z}_0)}[\log p(\mathbf{x}, \mathbf{z}_K)] - \mathbb{E}_{q_0(\mathbf{z}_0)}[\log q_0(\mathbf{z}_0)] - \mathbb{E}_{q_0(\mathbf{z}_0)} \left[\sum_{k=1}^K \log \det \left| \frac{\partial f_k}{\partial \mathbf{z}_k} \right| \right]$$

Започнете с напълно Гаусово разложение на множители и подобрете чрез промяна на променливите.

Триъгълните якобиани позволяват изчислителна ефективност.

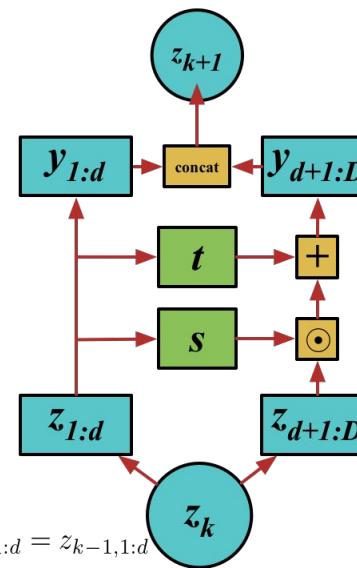
Planar Flow



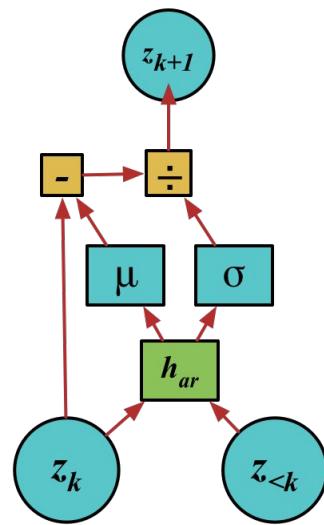
$$z_k = z_{k-1} + u h(w^\top z_{k-1} + b) \quad y_{d+1:D} = t(z_{k-1,1:d}) + z_{d+1:D} \odot \exp(s(z_{k-1,1:d}))$$

Линейно времево изчисляване на детерминантата и нейния градиент.

Real NVP

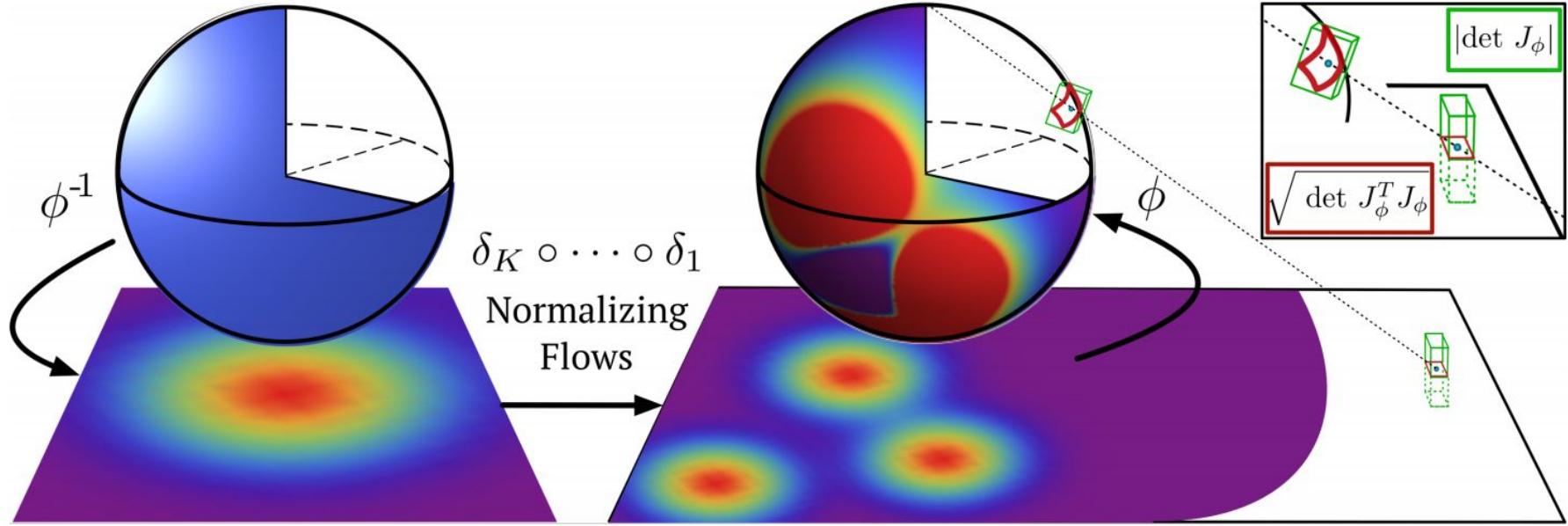


Inverse AR Flow



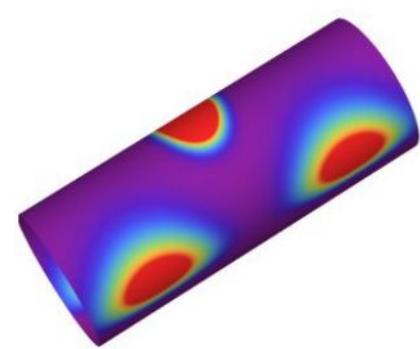
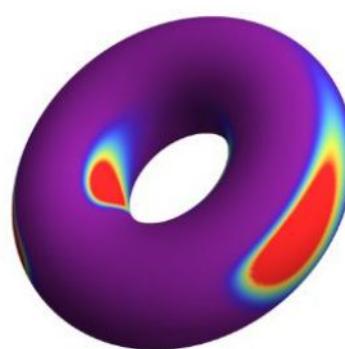
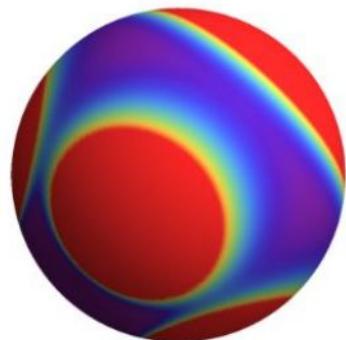
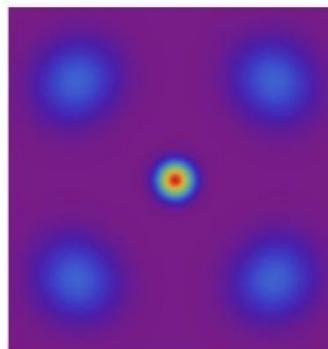
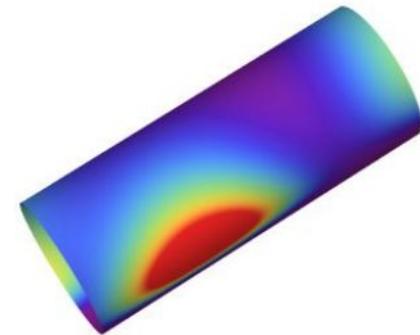
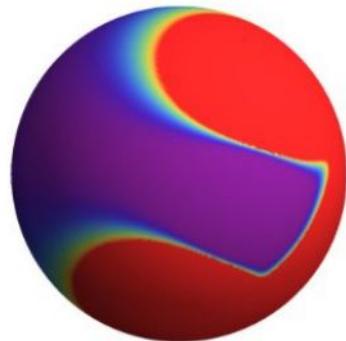
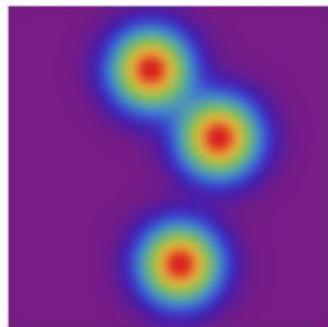
$$z_k = \frac{z_{k-1} - \mu_k(z_{<k}, x)}{\sigma_k(z_{<k}, x)}$$

Нормализиране на потоци върху неевклидови многообразия

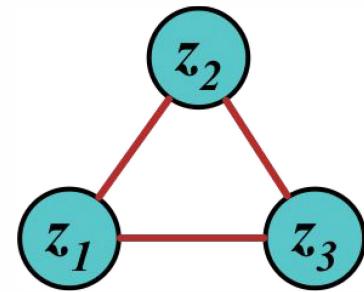


$$\log q_K(\mathbf{z}_K) = \log q_0(\mathbf{z}_0) - \frac{1}{2} \sum_{k=1}^K \log \det |\mathbf{J}_\phi^\top \mathbf{J}_\phi|$$

Нормализиране на потоци върху неевклидови многообразия

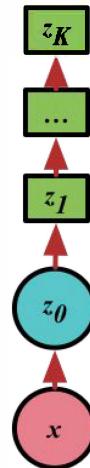


True Posterior

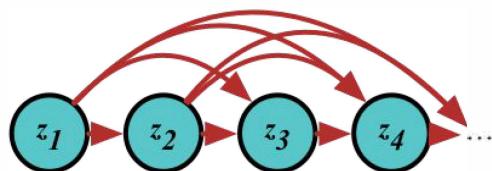


Families of Posterior Approximations

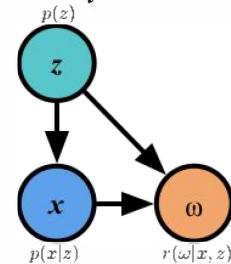
Normalising flows



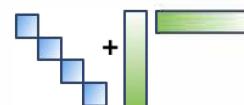
Structured mean-field



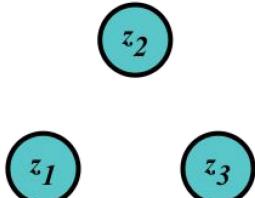
Auxiliary variables



Covariance models



Fully-factorised



Most Expressive

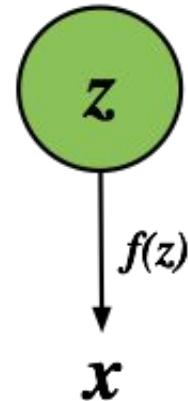
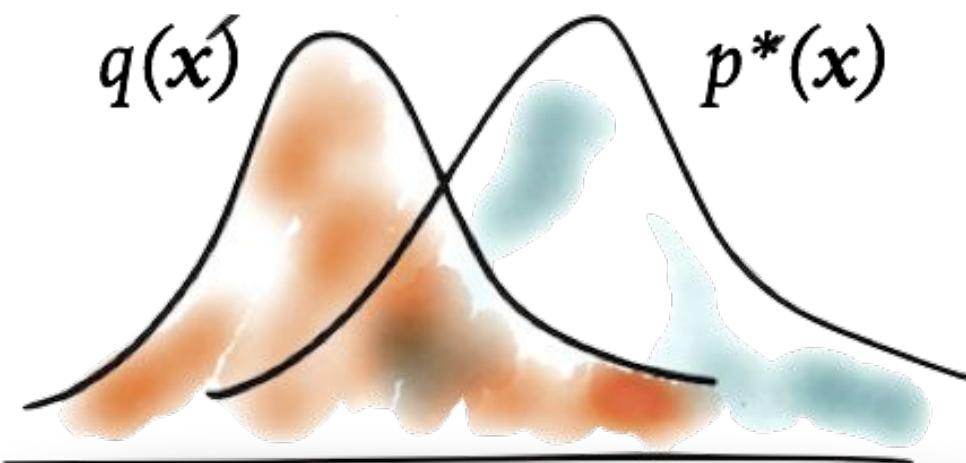
$$q^*(z|x) \propto p(x|z)p(z)$$

Least Expressive

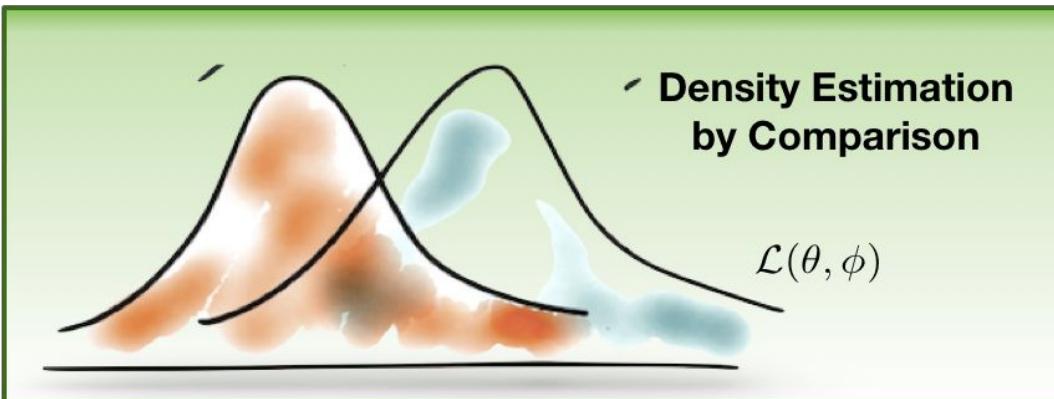
$$q_{MF}(z|x) = \prod_k q(z_k)$$

Учене чрез сравнения

За някои модели имаме достъп само до ненормализирана вероятност, частично познаване на разпределението или симулатор на данни.

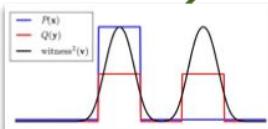


Сравняваме изчисленото разпределение $q(x)$ с истинското разпределение $p^*(x)$, използвайки образци.

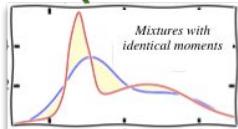


Probability Difference

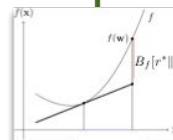
$$r_\phi = p^* - q_\theta$$



*Max Mean
Discrepancy*



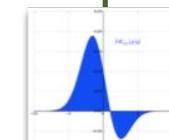
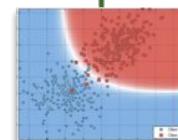
*Moment
Matching*



*Bregman
Divergence*

Probability Ratio

$$r_\phi = \frac{p^*}{q_\theta}$$



$f(u) = u \log u - (u + 1) \log(u + 1)$

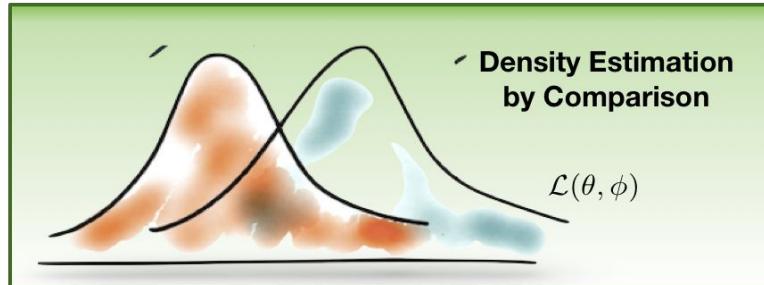
Учене чрез сравнения

Сравнение

Използвайте тест за хипотеза или сравнение, за да изградите спомагателен модел, за да посочите как данните, симулирани от модела, се различават от наблюдаваните данни.

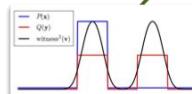
Оценяване

Коригирайте параметрите на модела, за да съответстват по-добре на разпределението на данните, като използвате сравнението.

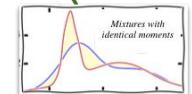


Probability Difference

$$r_\phi = p^* - q_\theta$$



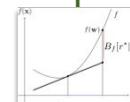
Max Mean
Discrepancy



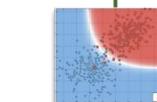
Moment
Matching

Probability Ratio

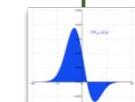
$$r_\phi = \frac{p^*}{q_\theta}$$



Bregman
Divergence



Class Probability
Estimation



$f(u)$
 $f(u) = u \log u - (u + 1) \log(u + 1)$

Коефициенти на плътност и класификация

Съотношение
на плътност $\frac{p^*(\mathbf{x})}{q(\mathbf{x})}$

Правило на
Бейс $p(\mathbf{x}|y) = \frac{p(y|\mathbf{x})p(\mathbf{x})}{p(y)}$

Комбинирайте данни

$$\{\mathbf{x}_1, \dots, \mathbf{x}_N\}$$

Реакни данни Симулирани данни

Присвояване на етикети

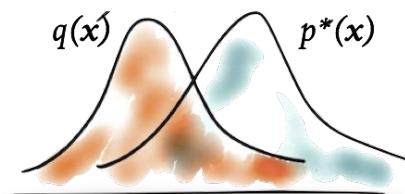
$$\{y_1, \dots, y_N\}$$

$$\{\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_{\hat{n}}\}$$

$$\{\tilde{\mathbf{x}}_1, \dots, \tilde{\mathbf{x}}_{\tilde{n}}\}$$

Еквивалентност

$$p^*(\mathbf{x}) = p(\mathbf{x}|y=1) \quad q(\mathbf{x}) = p(\mathbf{x}|y=-1)$$



Коефициенти на плътност и класификация

Условност

$$\frac{p^*(\mathbf{x})}{q(\mathbf{x})} = \frac{p(\mathbf{x}|y=1)}{p(\mathbf{x}|y=-1)}$$

Заместване на Бейс

$$= \frac{p(y=+1|\mathbf{x})p(\mathbf{x})}{p(y=+1)} \Bigg/ \frac{p(y=-1|\mathbf{x})p(\mathbf{x})}{p(y=-1)}$$

Класова вероятност

$$\frac{p^*(\mathbf{x})}{q(\mathbf{x})} = \frac{p(y=1|\mathbf{x})}{p(y=-1|\mathbf{x})}$$

Изчисляването на съотношение на плътност е еквивалентно на оценката на вероятността за клас.

Unsupervised-като-Supervised обучение

Функция за оценка

$$p(y = +1|\mathbf{x}) = D_\theta(\mathbf{x}) \quad p(y = -1|\mathbf{x}) = 1 - D_\theta(\mathbf{x})$$

Функция на загуби
на Бернули

$$\mathcal{F}(\mathbf{x}, \theta, \phi) = \mathbb{E}_{p^*(x)}[\log D_\theta(\mathbf{x})] + \mathbb{E}_{q_\phi(x)}[\log(1 - D_\theta(\mathbf{x}))]$$

Редуваща се
оптимизация

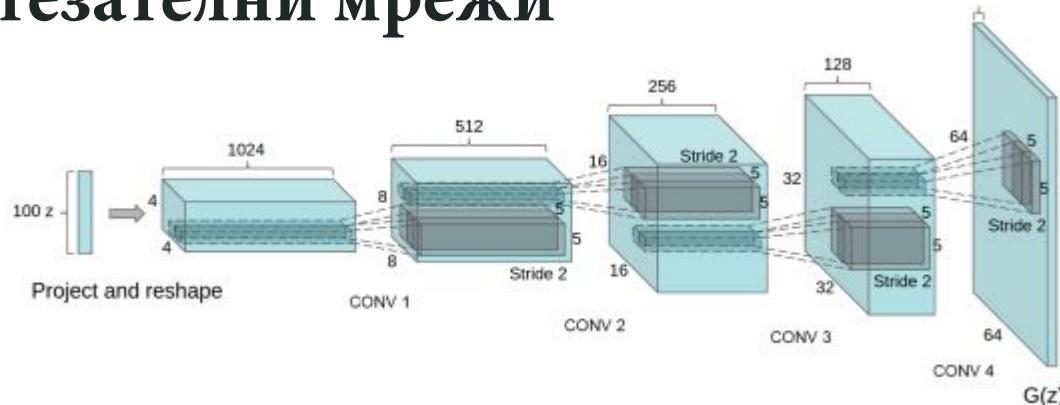
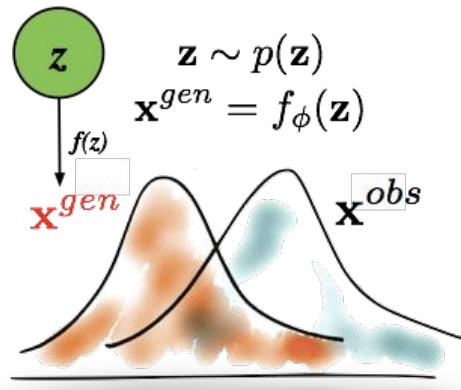
$$\min_{\phi} \max_{\theta} \mathcal{F}(\mathbf{x}, \theta, \phi)$$

- Използвайте, когато имаме различни симулатори и модели
- Може да формира загубата, използвайки всяко правилно правило за оценяване.

Други наименования:

- Учене без надзор и надзор
- Непрекъснато актуализиране на извода
- Класификатор ABC
- Генеративни състезателни мрежи

Генеративни състезателни мрежи



$$\mathcal{F}(\mathbf{x}, \theta, \phi) = \mathbb{E}_{p^*(\mathbf{x})}[\log D_\theta(\mathbf{x})] + \mathbb{E}_{q_\phi(\mathbf{x})}[\log(1 - D_\theta(\mathbf{x}))]$$

Редуваща се оптимизация

$$\min_{\phi} \max_{\theta} \mathcal{F}(\mathbf{x}, \theta, \phi)$$

Загуба при сравнение

$$\theta \propto \nabla_{\theta} \mathbb{E}_{p^*(\mathbf{x})}[\log D_\theta(\mathbf{x})] + \nabla_{\theta} \mathbb{E}_{q_\phi(\mathbf{x})}[\log(1 - D_\theta(\mathbf{x}))]$$

Генеративна загуба

$$\phi \propto -\nabla_{\phi} \mathbb{E}_{q(z)}[\log D_\theta(f_{\phi}(\mathbf{z}))]$$

Интегрални вероятностни метрики

$$\mathcal{M}_f(p, q) = \sup_{f \in \mathcal{F}} |\mathbb{E}_{p(x)}[f] - \mathbb{E}_{q_\theta(x)}[f]|$$

f понякога се нарича
тестова функция, функция свидетел или критик.

Много възможности за избор на f :
класификатори или функции в
определенi пространства.

$$\|f\|_L < 1$$

Васерщайн

$$\|f\|_\infty < 1$$

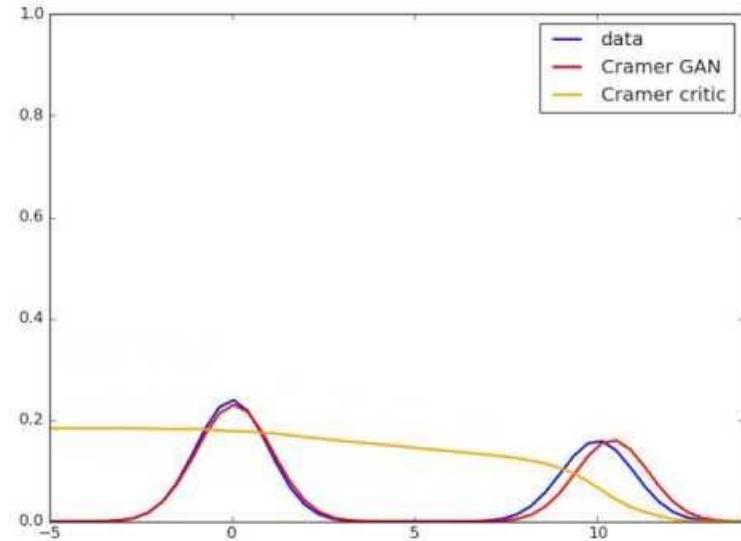
Пълна
Вариация

$$\|f\|_{\mathcal{H}} < 1$$

Максимално усреднo
несъответствие

$$\left\| \frac{df}{dx} \right\|_L < 1$$

Крамер



Вероятностно изучаване на правила

$$u(s, a) \sim \text{Environment}(a) \quad p(R(s)|a) \propto \exp(u(s, a))$$

$$\mathcal{F}(\theta) = \mathbb{E}_{\pi(\mathbf{a}|s)}[R(s, a)] - \text{KL}[\pi_\theta(\mathbf{a}|s) \| p(\mathbf{a})]$$

Актуализация на градиента на правилата

- Уеднаквен преди започването на операнди
- Оценка на градиента на функцията за оценка
(известен още като подсилване)

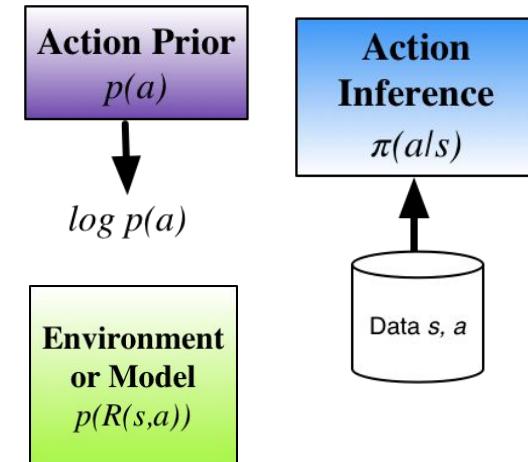
$$\nabla_\theta \mathcal{F}(\theta) = \mathbb{E}_{\pi(\mathbf{a}|s)}[(R(s, a) - c) \nabla_\theta \log \pi_\theta(\mathbf{a}|s)] + \nabla_\theta \mathbb{H}[\pi_\theta(\mathbf{a}|s)]$$

Други алгоритми:

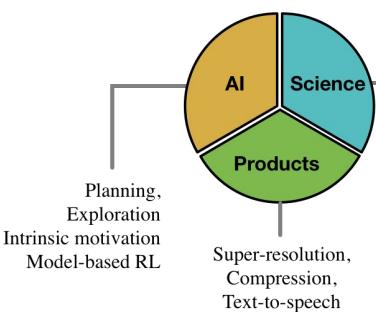
- Търсене на политика за относителна ентропия
- Генеративно състезателно имитационно обучение
- Подсилен вариационен извод

Други наименования:

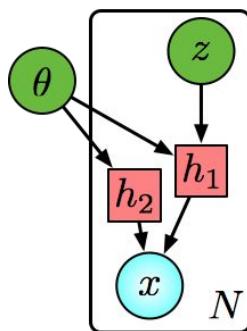
- Planning-as-inference
- Variational MDPs
- Path-integral control



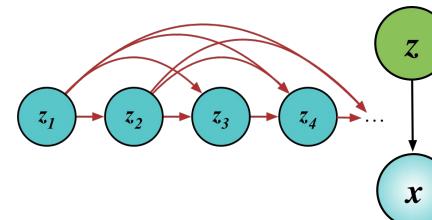
Applications of Generative Models



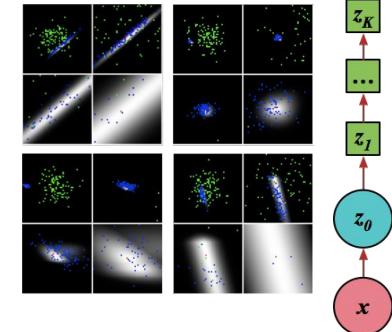
Probabilistic Deep Learning



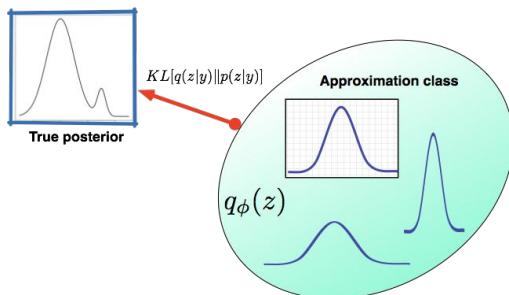
Types of Generative Models



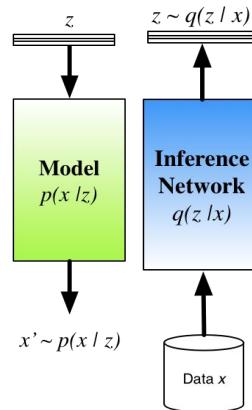
Rich Distributions



Variational Principles



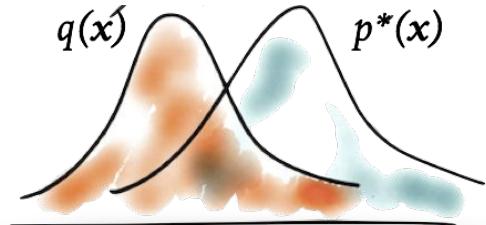
Amortised Inference



Stochastic Optimisation

$$\nabla_\phi \mathbb{E}_{q_\phi(\mathbf{z})} [f_\theta(\mathbf{z})] = \nabla \int [q_\phi(\mathbf{z}) | f_\theta(\mathbf{z})] d\mathbf{z}$$

Learning by Comparison



Предизвикателства

- Възможност за мащабиране до големи изображения, видеоклипове, множество модалности на данни.
- Оценка на генеративни модели.
- Устойчиви и робастни условни модели.
- Дискретни латентни променливи.
- Поддръжка-покритие в модели, режим-колапс.
- Калибриране.
- Несигурност на параметрите.
- Принципи на заключение без вероятност.

