تمرین سری چهارم | علی قاسم زاده | ۴۰۱۱۰۶۳۳۹

مسئله‌ی ۱. تبدیل متغیر

حل.

الف) از آنجایی که Z_1 و Z_2 از توزیع نرمال هستند و مستقل اند(چون ماتریس cov برابر است با I) و در درایه ی اول و دوم Y ترکیبی با مجموع ضرایب یک و ضرایب مثبت از Z_1 و Z_2 داریم، پس توزیع درایه ی اول و دوم Y نیز گاوسی است، حالا باید پارامتر های توزیع گاوسی که Y از آن امده است را بدست آوریم :

$$Z \sim \mathcal{N}\left(\begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix}, I\right) \qquad Y = \begin{bmatrix} \frac{3}{5}Z_1(t) + \frac{2}{5}Z_2(t) \\ \frac{1}{5}Z_1(t) + \frac{4}{5}Z_2(t) \end{bmatrix}$$

حالا اگر تبدیل خطی بنویسیم برای بدست آوردن Y از روی Z خواهیم داشت که :

$$Y = \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} Z_1(t) \\ Z_2(t) \end{bmatrix}$$

ماتریس سمت چپی را A می نامیم و از طرفی داریم که:

$$\mathbb{E}(Y) = \mathbb{E}(AZ) = A\mathbb{E}(Z) = A \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix} = \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix}$$

$$\Sigma_Y = \mathbb{E}\left((Y - \mathbb{E}(Y))(Y - \mathbb{E}(Y))^T\right) = \mathbb{E}\left((Y - \mathbb{E}(Y))(Y - \mathbb{E}(Y))^T\right) =$$

$$\mathbb{E}\left((AZ - \mathbb{E}(AZ))(AZ - \mathbb{E}(AZ))^T\right) = A\mathbb{E}\left((Z - \mathbb{E}(Z))(Z - \mathbb{E}(Z))^T\right) A^T =$$

$$A\Sigma_Z A^T = AA^T = \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix} \begin{bmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{4}{5} \end{bmatrix}^T = \begin{bmatrix} \frac{13}{25} & \frac{11}{25} \\ \frac{11}{25} & \frac{17}{25} \end{bmatrix}$$

پس خواهیم داشت که :

$$\longrightarrow Y \sim \mathcal{N}\left(\begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix}, \begin{bmatrix} \frac{13}{25} & \frac{11}{25} \\ \frac{11}{25} & \frac{17}{25} \end{bmatrix}\right)$$

حالا به بدست آوردن PDF توزیع این بردار می پردازیم :

$$f_Y(y) = \frac{1}{\sqrt{2\pi}^2 |\Sigma_Y|^{1/2}} exp\left(-\frac{1}{2}(y-\mu)^T \Sigma_Y^{-1}(y-\mu)\right) =$$

$$\mu = \begin{bmatrix} 0.0 \\ 0.0 \end{bmatrix}, \quad |\Sigma_Y| = \left|\begin{bmatrix} \frac{13}{25} & \frac{11}{25} \\ \frac{11}{25} & \frac{17}{25} \end{bmatrix}\right| = \frac{1}{625}\left|\begin{bmatrix} 13 & 11 \\ 11 & 17 \end{bmatrix}\right| = \frac{100}{625} = \frac{4}{25}$$

$$\longrightarrow \Sigma_Y^{-1} = \frac{1}{|\Sigma_Y|}\begin{bmatrix} 17 & -11 \\ -11 & 13 \end{bmatrix}$$

$$\longrightarrow f_Y(y) = \frac{1}{2\pi\frac{2}{5}} exp\left(-\frac{1}{2}y^T\left(\frac{25}{4}\begin{bmatrix}17 & -11\\ -11 & 13\end{bmatrix}\right)y\right)$$

$$\longrightarrow f_Y(y) = \frac{5}{4\pi} exp\left(-\frac{25}{8}y^T\left(\begin{bmatrix}17 & -11\\ -11 & 13\end{bmatrix}\right)y\right)$$

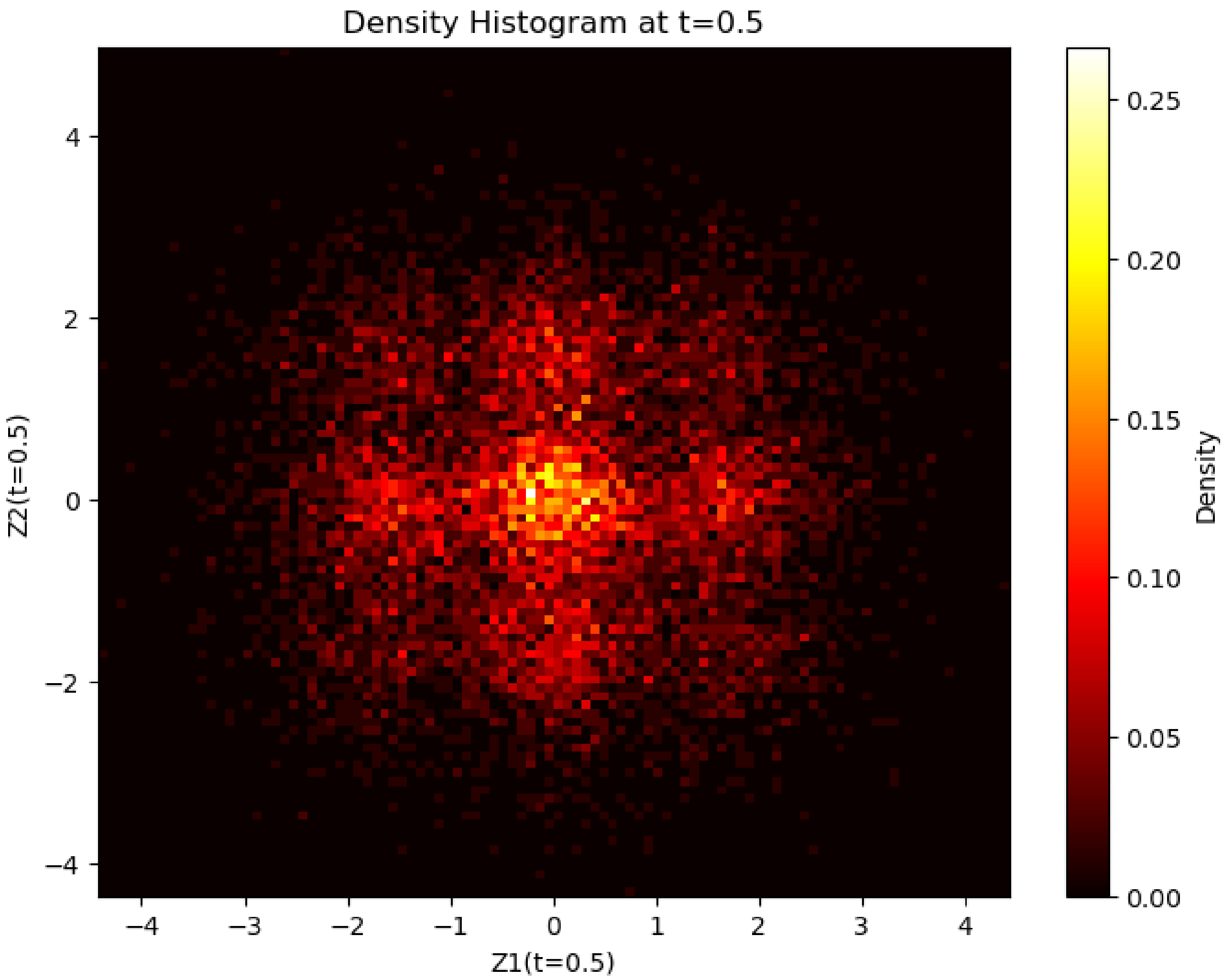
ب ۱) از معادله ها داریم که :

$$Z(t+1) = Z(t) + \frac{dZ}{dt}dt \longrightarrow \begin{bmatrix}Z_1(t+1)\\ Z_2(t+1)\end{bmatrix} = \begin{bmatrix}Z_1(t)\\ Z_2(t)\end{bmatrix} + \frac{d}{dt}\begin{bmatrix}Z_1(t)\\ Z_2(t)\end{bmatrix}dt =$$

$$\begin{bmatrix}Z_1(t)\\ Z_2(t)\end{bmatrix} + \begin{bmatrix}\frac{Z_1(t))}{dt}\\ \frac{Z_2(t)}{dt}\end{bmatrix}dt = \begin{bmatrix}Z_1(t)\\ Z_2(t)\end{bmatrix} + \begin{bmatrix}tanh(Z_1(t)^3)\\ tanh(Z_2(t)^3)\end{bmatrix}dt$$

$$\longrightarrow Z(t+1) = Z(t) + tanh(Z(t))dt$$

حالا با استفاده از python هیستوگرام مربوط را در $t = 0.5$ می کشیم، خواهیم داشت که :



ب ۲) طبق قضیه ی داده شده در مقاله داریم که :

$$\frac{\partial \log p(z(t))}{\partial t} = -tr\left(\frac{df}{dz(t)}\right)$$

همچنین داریم که $f(z(t))$ برابر است با $\frac{dz}{dt}$ ، پس خواهیم داشت که :

$$f(Z(t)) = \begin{bmatrix} tanh(Z_1(t)^3) \\ tanh(Z_2(t)^3) \end{bmatrix}$$

$$\frac{df}{dZ(t)} = \begin{bmatrix} \frac{d\,tanh(Z_1(t)^3)}{d\,Z_1(t)} & 0 \\ 0 & \frac{d\,tanh(Z_2(t)^3)}{d\,Z_2(t)} \end{bmatrix} =$$

$$\begin{bmatrix} 3Z_1(t)^2(1 - tanh^2(Z_1(t)^3)) & 0 \\ 0 & 3Z_2(t)^2(1 - tanh^2(Z_2(t)^3)) \end{bmatrix}$$

اسم این ماتریس را J می گذاریم خواهیم داشت که :

$$tr(J) = 3Z_1(t)^2(1 - tanh^2(Z_1(t)^3) + 3Z_2(t)^2(1 - tanh^2(Z_2(t)^3))$$

$$\frac{\partial \log p(Z(t))}{\partial t} = -tr(J) = -(3Z_1(t)^2(1 - tanh^2(Z_1(t)^3)) + 3Z_2(t)^2(1 - tanh^2(Z_2(t)^3)))$$

حالا از طرفی داریم که :

$$\frac{\partial \log p(Z(t))}{\partial t} = \frac{\frac{\partial p(Z(t))}{\partial t}}{p(Z(t))} \longrightarrow$$

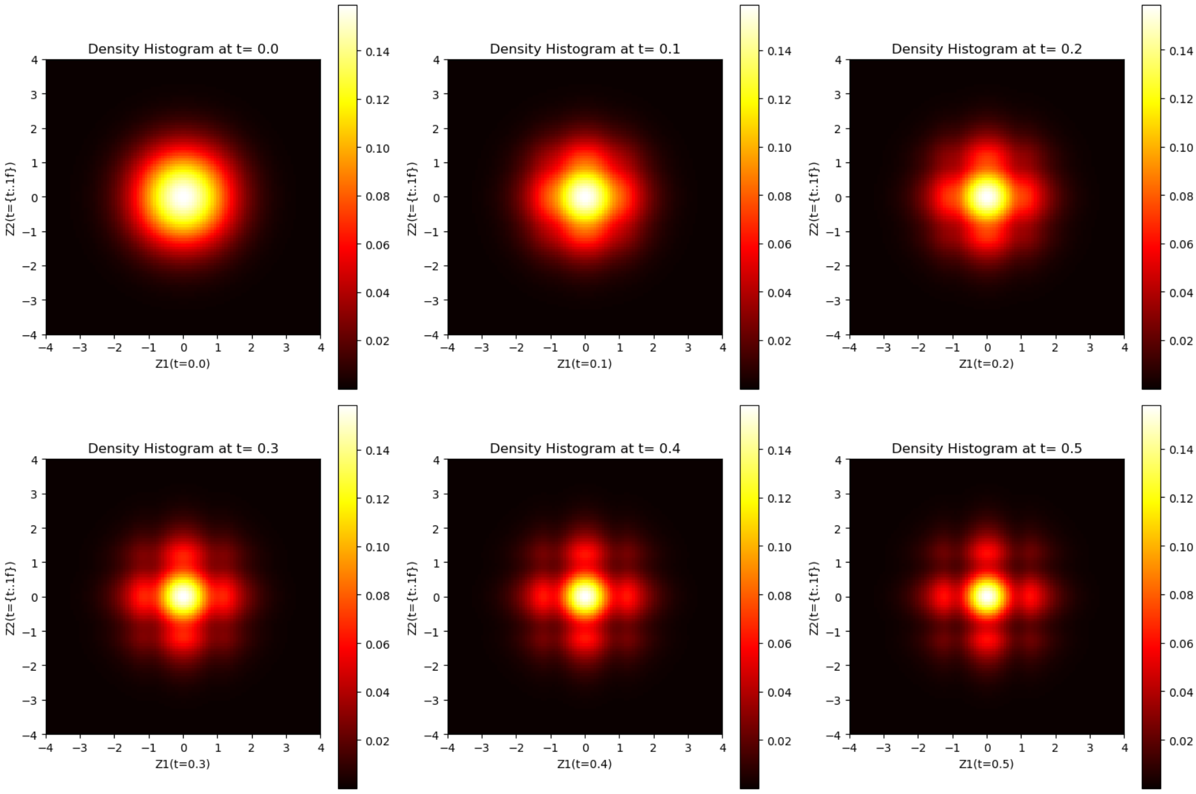
$$\frac{\frac{\partial P(Z(t))}{\partial t}}{p(Z(t))} = -tr(J) \longrightarrow \frac{\partial p(Z(t))}{\partial t} = -tr(J)p(Z(t)) \longrightarrow$$

$$\frac{\partial p(Z(t))}{\partial t} = -p(Z(t))(3Z_1(t)^2(1 - tanh^2(Z_1(t)^3) + 3Z_2(t)^2(1 - tanh^2(Z_2(t)^3)))$$

ب ۳) حالا با حل این معادله ی دیفرانسیل خواهیم داشت که :

$$p(Z(t)) = p(Z(0))exp\left(-\int_0^t (3Z_1(s)^2(1 - tanh^2(Z_1(s)^3) + 3Z_2(s)^2(1 - tanh^2(Z_2(s)^3)))ds\right)$$

حالا نمودار آنرا می کشیم، خواهیم داشت که :



ب ۴) به این روش Sensitivity Adjoint نیز می گویند به این صورت مشتق را حساب می کند که ابتدا معادله ی دیفرانسیل را از t=۰ تا t=T حل می کنیم و Z(T) را بدست می آوریم سپس متغییر Adjoint را به این صورت تعریف می کنیم که :

$$a(t) = \frac{\partial L}{\partial z(t)}$$

این متغییر به ما کمک می کند تا گرادیان را در زمان t محاسبه کنیم.
با استفاده از a(T) (گرادیان در زمان نهایی، T) معادله ی Adjoint زیر را به صورت معکوس از t=T به t=۰ حل می کنیم :

$$-a(t)^T \frac{\partial f}{\partial z}$$

محاسبه ی گرادیان نسبت به پارامتر θ نیاز به محاسبه ی انتگرال زیر دارد :

$$\frac{dL}{d\theta} = -\int_{t_1}^{t_0} \frac{\partial f(z(t), t, \theta)}{\partial \theta} dt$$

در اینجا مسئله هم اینگونه داریم که :

$$L(\theta) = \mathbb{E}_{x \sim Data}[\log p(x;\theta)], \quad f(Z(t), t, \theta) = \frac{dZ(t)}{dt}$$

پس با وجود اینها در روش بالا جایگذاری می کنیم و گرادیان را محاسبه می کنیم.

ب ۵) عدم دقت کافی در تقریب‌های عددی: روش‌های عددی مانند اویلر معمولاً برای تقریب راه‌حل معادلات دیفرانسیل استفاده می‌شوند و دقت محدودی دارند. مشتق‌گیری مستقیم از این تقریب‌های عددی خطاهای موجود را تقویت می‌کند و به نتایج ناپایداری منجر می‌شود.

مشکلات ناشی از ناپیوستگی و نویز عددی: روش‌های عددی معمولاً بر پایه گسسته‌سازی کار می‌کنند، و مقادیر به دست آمده از آنها ممکن است شامل ناپیوستگی یا نویز باشند. مشتق‌گیری از این مقادیر باعث تقویت نویز عددی و

ایجاد نتایج غیرواقعی یا ناپایدار می‌شود.
عدم تعمیم‌پذیری به فضاهای پیچیده‌تر: روش‌هایی مثل اویلر برای مدل‌سازی در فضای زمان گسسته طراحی شده‌اند. مشتق‌گیری مستقیم از این روش‌ها ممکن است در فضاهای پارامتری یا شبکه‌های پیچیده‌تری که نیازمند مشتق‌های دقیق‌تر هستند، ناکارآمد باشد.

کارایی کمتر در زمینه‌های بهینه‌سازی: در زمینه‌هایی مانند یادگیری ماشین و بهینه‌سازی، محاسبه گرادیان‌ها باید دقیق و قابل اعتماد باشد. استفاده از روش‌های عددی به‌جای روش‌های تحلیلی (یا روش‌های مناسب تر مثل AutoDiff) ممکن است منجر به کندی و خطاهای غیرقابل قبول در فرآیند بهینه‌سازی شود.

▷

مسئله‌ی ۲. مدل انرژی

حل.

الف) می دانیم که داریم :

$$p_\phi(x) \propto e^{-E_\phi(x)}, \quad Z_\phi = \int e^{-E_\phi(x)} dx$$

یا به عبارتی داریم که :

$$p_\phi(x) = \frac{e^{-E_\phi(x)}}{Z_\phi}$$

پس داریم که :

$$\log p_\phi(x) = \log e^{-E_\phi(x)} - \log Z_\phi = -E_\phi(x) - \log Z_\phi$$

پس خواهیم داشت که :

$$\nabla_\phi \log p_\phi(x) = -\nabla_\phi E_\phi(x) - \nabla_\phi \log Z_\phi$$

حالا داریم که :

$$\nabla_\phi \log Z_\phi = \frac{\nabla_\phi Z_\phi}{Z_\phi} = \frac{\nabla_\phi \int e^{-E_\phi(x)} dx}{Z_\phi} = \frac{\int \nabla_\phi e^{-E_\phi(x)} dx}{Z_\phi} =$$

$$-\frac{\int e^{-E_\phi(x)} \nabla_\phi E_\phi(x) dx}{Z_\phi} = -\int \frac{e^{-E_\phi(x)}}{Z_\phi} \nabla_\phi E_\phi(x) dx =$$

$$-\int p_\phi(x) \nabla_\phi E_\phi(x) dx = \int p_\phi(x)(-\nabla_\phi E_\phi(x)) dx =$$

$$\mathbb{E}_{x \sim p_\phi(x)}[-\nabla_\phi E_\phi(x)]$$

در نتیجه حکم ثابت می شود.

ب) برای سمپل گیری از $p_\phi(x)$ نیاز داریم تا Z_ϕ را محاسبه کنیم که پیچیدگی محاسباتی بالایی دارد زیرا باید روی کل فضای داده انتگرال بگیریم و به همین خاطر سمپل گیری از این توزیع بسیار دشوار است.

▷