Comp000311

بسم الله الرحمن الرحیم   جلسه ۱۱/۳/۱۴۰۰

Markov Chain Monte Carlo method
(MCMC method)

یک الگوریتم جستجو برای بهینه سازی (بسته به اینکه چه مسئله فیزیکی را دنبال می کنیم، تعیین می شود)

به منظور آموزش این الگوریتم از یک مسئله شروع می کنیم.

جستجو در فضای پارامترهای آزاد مدل به منظور تعیین مقادیر آن پارامترها که نهایتاً به
model free parameters

تکمیل مدل منجر می شود.

(یادتون هست که من یک سری داده به شما داده بودم و گفته بودم برازش دهید
Data

fitting procedure

observed Data

x	y
x_1	y_1
x_2	y_2
$\vdots$	
x_N	y_N

$y(x) = A x^{B}$
theory

$A = ?$
$B = ?$
مجهول هستند

به مقایسه با داده آنها را تعیین کنید.)

(یادتون هست که گفتم فرض کنید می خواهید قانون اهم را بررسی کنید

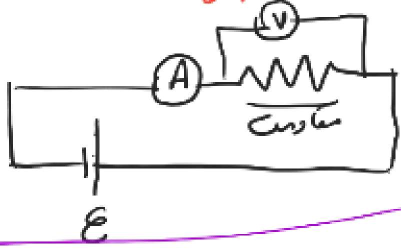
$$\underline{V = RI}$$

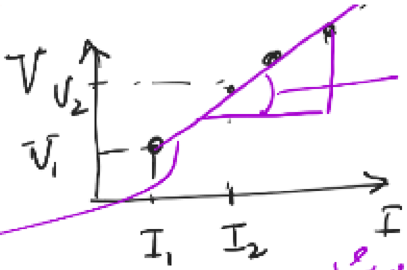
$$\boxed{R = cts}$$

قانون اهم ← رابطه بین ولتاژ و جریان که از یک سیم R می گذرد که دارای به صورت $V = RI$ — نظریه

آزمایشگاه

مستقل از ولتاژ ...





slope ≡ R

↑ پارامتر آزاد مدل که با آزمایش تعیین می‌شود.

پیدا کردن R به‌طوری که خط برازش شده بهترین تطابق با داده‌ها را داشته باشد / از بیشتر نقاط بگذرد

$R \equiv$ Model Free Parameter

اگر قرار است نقش مقاومت را بازی کند

$R \geq 0 \rightarrow p(\theta) = \begin{cases} 1 & R \geq 0 \\ 0 & R < 0 \end{cases}$

مسئله‌ای که در سری آن می‌خواهم MCMC را آموزش دهم یافتن بهترین مقادیر برای پارامترهای آزاد مدل است. که خود مدل نمی‌تواند مقادیر آنها را تعیین کند و صرفاً با مقایسه با اندازه‌گیری‌ها این مقادیر تعیین می‌شوند.

$\{D\}$: $\{Data\}$: Observed Data ste

$\{\Theta\}$: $\{Parameters\}$: Free Parameters

$\rightarrow \boxed{\{\theta\}_{Best} = ?}$

$\boxed{p(\{\theta\} | \{D\})}$ ← به لحاظ احتمالی

تابع احتمال شرطی: احتمال داشتن دسته $\{\theta\}$ به شرطی که $\{D\}$ مشاهده و اندازه‌گیری شده باشند.

با بیشینه کردن $p(\{\theta\}|\{D\}) \rightarrow \{\theta\}_{Best}$ یعنی ادعا می‌کنیم که

$$p\left(\{\theta\}_s, \{\theta\}_{Best} \,/\, \{D\}\right) \equiv \underline{\underline{\text{Maximum}}}$$

$$\boxed{p(\{\theta\}|\{D\}) = \frac{\mathcal{L}(\{D\}|\{\theta\})\, p(\{\theta\})}{\int d\{\theta\}\, \mathcal{L}\, p(\{\theta\})}}$$

Bayse theorem

Posterior PDF

تابع چگالی احتمال پسینی

Likelihood PDF

تابع چگالی احتمال درست نمایی

Prior PDF

تابع چگالی احتمال پیشینی

(شامل همه اطلاعات اولیه در خصوص مقادیر $\{\theta\}$ است)

اگر اطلاعات اولیه را نداشته باشیم ← $p(\{\theta\})_s = cts = 1$

اگر در یک مسئله خاص اطلاعات اولیه در خصوص مقدار پارامتر آن نداشتیم تابع پیشینی را ثابت در نظر می‌گیریم

$$\underbrace{p(\{\theta\}|\{D\})}_{\text{to be Maximized}} \sim \mathcal{L}(\{D\}|\{\theta\}) \rightarrow \text{to be maximized}$$

پس از این به بعد هم تمرکز را روی بیشینه کردن <u>تابع درست نمایی</u> قرار می‌دهیم.

شکل تابع احتمال درست‌نمایی چگونه است؟

قبلاً به این سئوال پاسخ داده‌ام.

$$\{D\}: \{D_1, D_2 \ldots D_N\}$$

$$= \{(x_1,y_1),(x_2,y_2) \ldots (x_N,y_N)\} \qquad y = A x^B \quad m=2$$

$$= \{(I_1,V_1),(I_2,V_2),\ldots (I_N,V_N)\} \longrightarrow m=1 \quad (R)$$

$$\underbrace{\mathcal{L}(\{D\}|\{\theta\}) = \mathcal{L}(\{D_1 \ldots D_N\} \,|\, \{\theta_1 \ldots \underbrace{\theta_m}_{\text{تعداد پارامترهای آزاد}}\})}_{\text{N-Joint Conditional PDF}}$$

① اگر این احتمال‌ها از هم مستقل باشند (اندازه‌گیری‌های مختلف نسبت به هم کاملاً مستقل باشند)

$$\mathcal{L} = \prod_{i=1}^{N} \mathcal{L}_i$$

② قضیه حد مرکزی ←

$$\mathcal{L}_i \equiv \text{Gaussian}$$

$$\mathcal{L}_i = \mathcal{L}(D_i|\{\theta\}) \propto e^{-\frac{[\tilde{y}_{obs}(x_i) - \tilde{y}_{theor}(x_i,\{\theta\})]^2}{2\sigma_i^2}}$$

خطای y_{obs}

$$\mathcal{L} \propto e^{-\sum_{i=1}^{N} \frac{[y_{obs} - y_{th}(x_i,\{\theta\})]^2}{2\sigma_i^2}}$$

$$\mathcal{L} \propto e^{-\frac{1}{2}\chi^2} \qquad \chi^2 \equiv \sum_{i=1}^{N} \frac{[y_{obs} - y_{th}(x_i,\{\theta\})]^2}{\sigma_i^2}$$

بیشینه کنیم

کمینه شود.

$$\mathcal{L} \to \mathcal{L}_{Max} \longrightarrow \boxed{\chi^2 \to \chi^2_{min}}$$

$$\boxed{\chi^2_{min} = \chi^2(\{\theta\} = \{\theta\}_{Best})}$$

استدلال تبدیل می‌شود به جستجو در یافتن $\{\theta\}$ هایی که به ازای آنها χ^2 کمینه ($\mathcal{L}$ بیشینه) می‌شود.

الگوریتم MCMC استفاده می‌کنیم.

~~Hamiltonian Monte Carlo (HMC) وجود دارد که برخی از نواقص MCMC را برطرف می‌کند~~

ایده اصلی MCMC جستجو در فضای پارامترها با توجه به اعداد تصادفی به منظور بهینه‌سازی ($\mathcal{L}_{max}$ و χ^2_{min})

$$\chi^2(\{\theta\}) = \begin{cases} \chi^2(A,B) \\ \chi^2(R) \\ E(\alpha) \end{cases}$$

$$E(\alpha) \to \{\theta\}, \quad E(\alpha) \to \chi^2$$

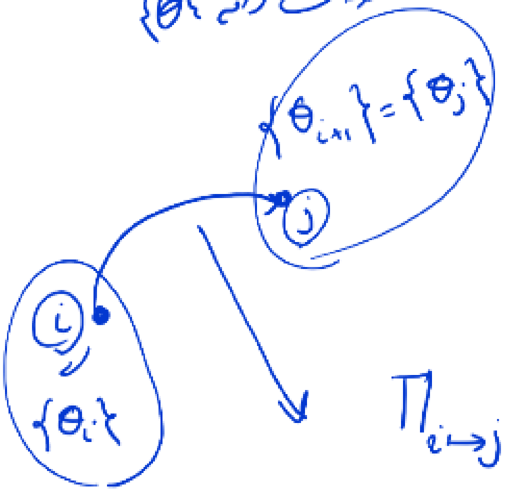
جستجوگری در پیش

$\{\theta_1 \ldots \theta_m\}$ $\quad S\{\theta\}$

$(-\infty, +\infty)$ $\quad (-\infty, +\infty)$

در حالت دائمی $\{\theta\}$

فلسفه نام گذاری روش . $\underline{M}CMC$



$\{\theta_{i+1}\} = \{\theta_j\}$ $\quad \{\theta_i\}$

$\Pi_{i \to j} =$ احتمال رفتن از i به j را نشان می‌دهد

$$\Pi_{i \to j} \equiv \Pi_{ij} = \begin{cases} ① \; \alpha_{ij} & \text{احتمال جهش از i به j} \\ ② \; p_{ij} & \text{احتمال قبول موقعیت جدید} \end{cases}$$

$$\boxed{\Pi_{ij} = \alpha_{ij} \, p_{ij}}$$

دلیل مارکوف بودن ← در زنجیره های مارکوف وضعیت جدید به طور مستقیم به یک وضعیت قبل برمی‌گردد

Π_{ij}

الگوریتم Metropolic ← $\Pi_{ij} = \checkmark$

احتمال اقدام به حرکت به سوی مکان j ام $\quad$ احتمال پذیرش موقعیت جدید

$$\Pi_{ij} = \alpha_{ij} \times p_{ij}$$

$L_j(x_j^2)$

$$= \alpha_{ij} \; Min\left\{1 \, , \, \frac{\text{وضعیت جدید}}{\ldots}\right\}$$

اگر احتمال درست نمایی حالت جدید بیشتر از حالت قبلی باشد یعنی $\frac{\mathcal{L}_j}{\mathcal{L}_i}>1$ پس یعنی حتماً

موقعیت جدید را بپذیر

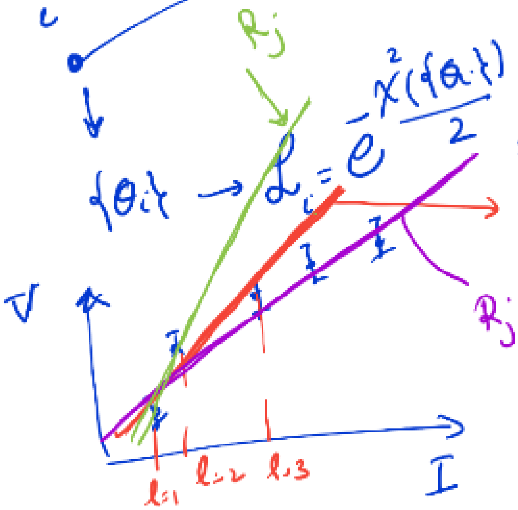
اگر $\mathcal{L}_j<\mathcal{L}_i$ ← $\frac{\mathcal{L}_j}{\mathcal{L}_i}<1$

$$\boxed{\min\left\{1, \frac{\mathcal{L}_j}{\mathcal{L}_i}\right\}}$$

الگوریتم متروپلیس

$i \longrightarrow j$

$\{\theta_i\} \rightarrow \mathcal{L}_i = e^{-\frac{\chi^2(\{\theta_i\})}{2}} = e^{-\chi_i^2}$



obser   theory

$R_i \rightarrow \mathcal{L}_i = e^{-\sum_{l=1}^{N} \frac{[V_l - R_i I_l]^2}{2\sigma_l^2}}$

$R_j \rightarrow \mathcal{L}_j = e^{-\sum_{l=1}^{N} \frac{[V_l - R_j I_l]^2}{2\sigma_l^2}}$

برازشی بهتر است که $\mathcal{L}$ بزرگتری دارد

$R_i = ✓ \longrightarrow R_j = ✓$

$\downarrow \qquad\qquad \downarrow$

$\mathcal{L}_i \qquad\qquad \mathcal{L}_j$

$\mathcal{L}_j > \mathcal{L}_i \rightarrow \frac{\mathcal{L}_j}{\mathcal{L}_i} > 1$

$\mathcal{L}_j < \mathcal{L}_i$

$L_j < L_i$ — $\frac{L_j}{L_i} < 1$

در این حالت یک استراتژی می‌توان این باشد که هیچ‌وقت جدید را نمی‌پذیرم چرا چون وضعیت جدید خوب نیست

پیش‌بینی کردم اگر موقعیت جدید را بپذیرم بعداً موقعیت به شرایط بهتری منتهی می‌شود

$$RA = \min\left\{1, \frac{L_j}{L_i}\right\}$$

$$\min\left\{1, \frac{L_j}{L_i}\right\} = \begin{cases} \text{if } L_j > L_i \rightarrow 1 = \min\{1, \dots\} & \text{پذیرش قطعی} \\ \text{if } L_j < L_i \rightarrow \frac{L_j}{L_i} & \text{با احتمال} \end{cases}$$

با احتمال $\frac{L_j}{L_i}$ وضعیت جدید را می‌پذیرید

0.8 → با احتمال ۸۰ درصد وضعیت جدید را می‌پذیرم (قطعاً وضعیت جدید را رد نمی‌کنید بلکه

با احتمال آن را می‌پذیرید (محافظه‌کارانه است))

$$\text{احتمال پذیرش} = \min\left\{1, \frac{L_j}{L_i}\right\}$$

مثال: $\frac{L_j}{L_i} = 10$ → احتمال پذیرش $= 1$ ✓

مثال: $\frac{L_j}{L_i} = 0.9$ → احتمال پذیرش $= 0.9$ — 90%

احتمال پذیرش $= 0.1 - 10\%$ ← $\frac{\mathcal{L}_{ij}}{\mathcal{L}_i} = 0.1$ معادل

نمونه الگوریتم MCMC با احتمال متروپلیس

تعداد اندازه‌گیری N $\quad \{D\}: \{D_i\},\ i=1,\dots$

تعداد پارامترهای آزاد مدل $\underline{M}$ $\quad \{\theta\}: \{\theta_j\},\ j=1,\dots$

Import Data set

Select $\{\theta\}_{old} =$ مقدار اولیه دلخواه برای پارامترهای مدل

Initial State

اگر بر اساس اطلاعات اولیه مقدار
شبیه نزدیک‌تر باشیم سریع‌تر الگوریتم
همگرا ← مقدار بهینه سریع‌تر پیدا می‌شود.

Compute $\chi^2_{old}(\{\theta\}_{old}) = \checkmark \leftarrow \mathcal{L}_{old} = \checkmark$

تعداد قدم‌ها $\boxed{N_{MC}}$ از مکان شروع حرکت کنیم

loop on MCMC

$\textcircled{A}\ \{\theta\}_{old} \xrightarrow{\Pi_{old-New}} \{\theta\}_{New}$

Compute $\chi^2_{New}(\{\theta\}_{New}) = \checkmark$

[illegible]

$$\mathcal{L}_{old} = e^{-\frac{\chi^2_{old}}{2}}$$
$$\mathcal{L}_{New} = e^{-\frac{\chi^2_{New}}{2}}$$
$$\Rightarrow \frac{\mathcal{L}_{New}}{\mathcal{L}_{old}} = e^{-\frac{[\chi^2_{old} - \chi^2_{New}]}{2}}$$

$$\Delta\chi^2 = \chi^2_{old} - \chi^2_{New}$$

Ⓑ check the Acceptance Rate

$$\boxed{AR = \min\left\{1, \frac{\mathcal{L}_{New}}{\mathcal{L}_{old}}\right\} = \min\left\{1, e^{\frac{-\chi^2_{New} + \chi^2_{old}}{2}}\right\}}$$

حساب می‌کنیم

$$= \min\left\{1, e^{\frac{\Delta\chi^2}{2}}\right\}$$

Ⓒ Write $\underset{\uparrow}{\underline{\{\theta\}_{old}}}$, $\underbrace{e^{-\chi^2_{old}/2}}_{\text{دمای حال جدید}}$

End loop

چند بار قدم بزنم ؟

Subroutine Ⓐ

$$\{\theta\}_{old} \longrightarrow \{\theta\}_{New}$$

$$\{\theta\}_{New} = \{\theta\}_{old} + \{\Delta\theta\}$$

انتخاب می‌کنیم

اعداد تصادفی با توزیع گوسی

$$\{\Delta\theta\} \longleftarrow N(0, \sigma_{\{\theta\}})$$

از اعداد تصادفی کامپیوتر استخراج می‌کنیم

$$\theta^{(1)}_{New} = \theta^{(1)}_{old} + \underline{\Delta\theta^{(1)}}$$

$$\vdots$$

$$\theta^{(M)}_{New} = \theta^{(M)}_{old} + \underline{\Delta\theta^{(M)}}$$

اگر خیلی کوچک شد

دیر به نتیجه می‌رسد

اگر خیلی بزرگ شد

باز دیر به مقصد می‌رسید

آزمون خطا.

Subroutine Ⓑ

$\xi =$ Call Random Number $\in (0,1)$

تابع توزیع کت

$\min\left(1, \frac{L_{New}}{L_{old}}\right)$

if $\xi \leqslant e^{\frac{\Delta\chi^2}{2}}$ Then

$\{\theta\}_{old} = \{\theta\}_{New}$

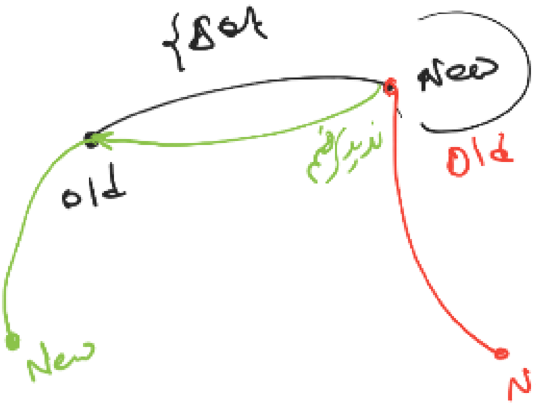
$\boxed{\chi^2_{old} = \chi^2_{New}}$

End if

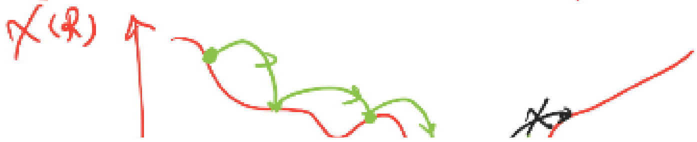
① اگر $\chi^2_{old} < \chi^2_{New} \Rightarrow \Delta\chi > 0 \Rightarrow e^{\frac{\Delta\chi^2}{2}} > 1 \Rightarrow$ ξ حتی کوچکتر از 1 پس حتما این شرط ارضا می‌شود پس حتما موقعیت جدید انتخاب و به عنوان موقعیت قدیمی نام گذاری می‌شود.

② اگر $\chi^2_{old} < \chi^2_{New} \Rightarrow \Delta\chi < 0 \Rightarrow e^{\frac{\Delta\chi^2}{2}} < 1 \Rightarrow$ پس

بسته به اینکه وضعیت جدید چقدر خراب باشد امکان پذیرش پیدا می‌شود



حلقه MCMC یعنی تعداد پرش چقدر باشد؟ ←



متفاوت

جواب همگرا شده است

معیارهای همگرایی را بررسی ← قاعده سرانگشتی ←

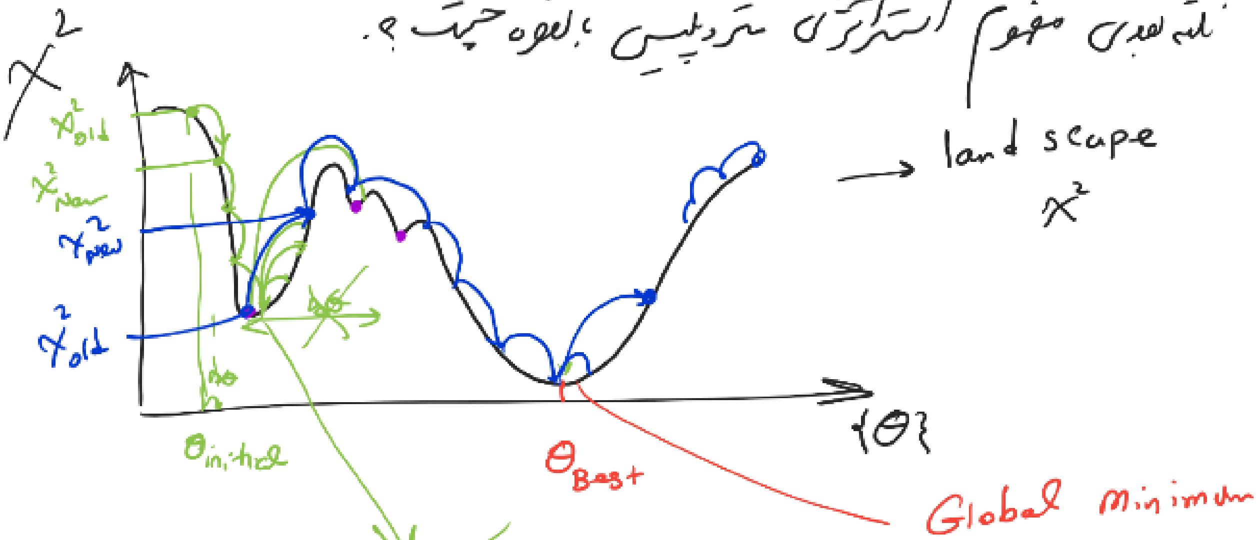
$N_{MC} = 10^5$ ← R_s مقدار نهایی چاپ شود

$N_{MC} = 3\times10^5$ ← R_s

شبیه هم هستند

همگرا شده است

نکته بعدی مفهوم استراتژی متروپلیس بالقوه چیست ؟



گیر می‌افتیم (خصوصاً اگر $\Delta\theta$ کوچک باشد)

برای اینکه بتوانم از کمینه‌های موضعی خلاص شوم

حالت جدید را حتی اگر بدتر باشد بپذیرم

ایده متروپلیس با درنظر گرفتن احتمال پذیرش حالت بدتر امکان خلاص از کمینه‌های

موضعی فراهم می‌کند

$\Delta\theta$ خیلی کوچک نگیریم

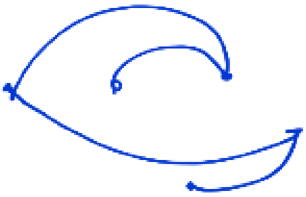
خیلی بزرگ نگیریم

تکرار در سیستم

از شرایط اولیه متفاوت شروع کنیم

جواب پایان یک

ولی جستجو به روش وقت تصادفی



(ولی در موارد قبلی به صورت تعیینی جستجو کردیم)

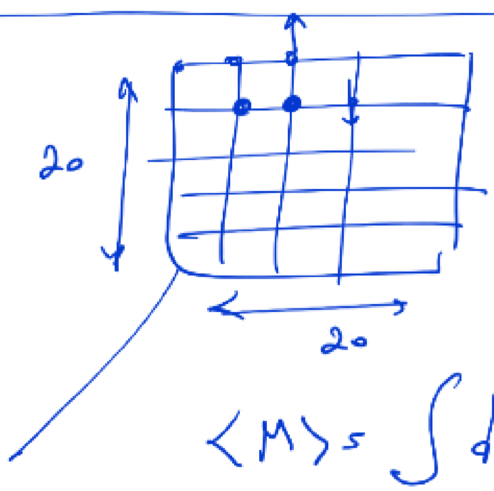
بسیار موثر MCMC $\leftarrow \{\theta\}$ $\underset{\text{بزرگ}}{M}$

تعداد کل حالت هایی که باید چک کنیم

$10^{9} - 10^{10}$

$M = 100$

$-100 \leq \theta \leq +100 \quad \Delta\theta = 10^{-5} \quad - \quad 10^{7}$



$\boxed{\text{Ising}}$ موضوع مدل

20

$\boxed{S = \pm 1}$

$$\langle M \rangle = \int d\Gamma \, \rho(\Gamma) \, M(\Gamma)$$

N.. 400

[illegible]

$$\Gamma = 2^{400} = \frac{2.58 \times 10^{120}}{10^{9}} = \left(\frac{8.8 \times 10^{103}\ \text{Yr}}{10^{9}\ \text{ن}} \right)$$

$$\sim 10^{94}\ \text{year}$$