MCMC method

بسم الله الرحمن الرحیم

۹۹/۲/۲۱

Markov Chain Monte Carlo (MCMC)

یک الگوریتم جستجو برای بهینه سازی

مقدمه: در مباحث مدل سازی نظری مبتنی بر اصول موضوعه و رهیافت های پدیده شناختی همیشه درجات آزادی وجود دارند که توسط خود نظریه این درجات آزادی تعیین نمی شوند، پارامترهای آزادی وجود دارند که در مقام مقایسه با رخداد متناسب با آن در آزمایشگاه یا طبیعت، قابل تعیین هستند.

$\{D\}$: Observed Data

$\{\theta\}$: Free Parameters

$P(\{\theta\}|\{D\}) =$ احتمال اینکه دسته پارامترهای $\{\theta\}$ درست باشند بشرط اینکه $\{D\}$ که همه داده‌های اندازه‌گیری شده باشند

Posterior Function

اگر ادعا کنیم که مدل به طور مطلوبی توانسته نتایج حاصل از تجربه و آزمایش یک پدیده را توصیف کند یعنی اینکه تابع پسینی بیشینه شده است.

$$P(\underbrace{\{\theta\}=\{\theta_{Best}\}}_{***}|\{D\}) \equiv \text{Maximized}$$

Bayse Theorem

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

پسین ← Posterior

Likelihood درست‌نمایی

Prior اطلاعات پیشین

مقدار حقیقی است

$\sqrt{\theta_1}$

$\theta_1 \geq 0$

$p(\{\theta\})$: احتمال اینکه توزیعی از $\{\theta\}$ با مقدار مشخص داشته باشیم

$$\begin{cases} p(\theta_1 < 0) = 0 \\ p(\theta_1 \geq 0) \neq 0 \end{cases}$$ اطلاعات قبلی در خصوص نظریه داریم.

$$\underbrace{P_{Max}(\{\theta\}|\{D\})} \longrightarrow \left[\mathcal{L}(\{D\}|\{\theta\})\, p(\{\theta\})\right]_{Max}$$

اگر $p(\{\theta\}) = cts$

$$P_{Max} \longrightarrow \mathcal{L}_{Max}$$

$\mathcal{L}(\{D\}|\{\theta\})$: احتمال اینکه اندازه‌گیری‌های $\{D\}$ به دست آید به شرط اینکه مدل $\{\theta\}$ داده شده باشد.

Likelihood

فرض کنید یک مدل داریم

$Y_{Theory}(x_i, \{\theta\})$

و به ازای آن داده تجربی

$Y_{obs}(x_i)$ $\quad i=1, \dots, N$

تعداد نقاط اندازه‌گیری شده $\quad \{D_i\}, \; i=1 \dots N$

$$\mathcal{L}(\{D\}|\{\theta\}) = \text{N-Joint Conditional Probability}$$

Density function

فرض استقلال اندازه‌های مختلف
فرض قضیه حد مرکزی

$$\mathcal{L}(\{D\}|\{\theta\}) = \prod_{i=1}^{N} P_i(D_i|\{\theta\})$$

Gaussian Distribution

$$e^{-\frac{(x-\bar{x})^2}{2\sigma^2}}$$

$$\mathcal{L}(\{D\}|\{\theta\}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_i^2}}\, e^{-\frac{[Y_{obs}(x_i) - Y_{the}(x_i,\{\theta\})]^2}{2\sigma_i^2}}$$

در حالت حدی قضیه حد مرکزی در واقع گفته ایم که کمیت مشاهده پذیر Y حول مقدار حاصل از نظریه به صورت گوسی توزیع شده است

$$\mathcal{L}(\{D\}|\{\theta\}) \simeq \exp\left[-\sum_{i=1}^{N} \frac{[Y_{obs}(x_i) - Y_{the}(x_i,\{\theta\})]^2}{2\sigma_i^2}\right]$$

خطای اندازه‌گیری i ام

$$\mathcal{L}(\{D\}|\{\theta\}) \sim e^{-\frac{\chi^2(\{\theta\})}{2}}$$

$$\boxed{\chi^2(\{\theta\}) \equiv \sum_{i=1}^{N} \frac{[Y_{obs}(x_i) - Y_{the}(x_i,\{\theta\})]^2}{\sigma_i^2}}$$

$$\mathcal{L}_{Max} \longrightarrow \chi^2_{min} = \chi^2(\{\theta\} = \{\theta\}_{Best})$$

جستجوی کمینه در فضای پارامتری $\{\theta\}$ که نهایتا $\{\theta\}_{Best}$ بدست آید که کمینه سراسری

χ^2 بدست آید که منجر به بیشینه سراسری توزیع درست نمایی شود که نهایتا منجر به

بیشینه سراسری توزیع پسینی شود (در غیاب توزیع Prior)

ایده اصلی MCMC در واقع جستجو در فضای پارامتری به منظور یافتن کمینه سراسری (نه کمینه موضعی)

Global Minima

در فضای $\chi^2(\{\theta\})$ یا به طور کلی به منظور یافتن بیشینه سراسری توزیع پسینی

که نهایتا منجر به یافتن $\{\theta\}_{Best}$ می‌شود

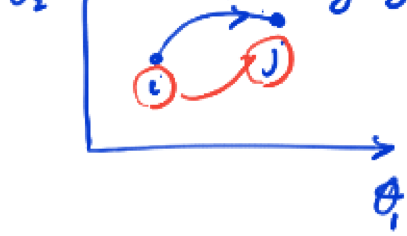
{جستجو در فضای پارامتری به منظور ارضای یک شرط، به کمک اعداد تصادفی}

از آنجایی که این جستجو به صورت پی در پی انجام می‌شود و انتخاب مقادیر جدید پارامترها فقط به این بستگی دارد که در مرحله قبل کجا بوده، بنابراین با یک فرآیند مارکوف یا زنجیره مارکوف مواجه هستیم.

در عملیات جستجو

$$\Pi_{ij} = \begin{cases} ① \ \alpha_{ij} & \text{احتمال جهش از نقطه } i \text{ به نقطه } j \\ ② \ p_{ij} & \text{احتمال قبول این جهش} \end{cases}$$

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



احتمال جهش از نقطه i به هر نقطه دیگر ممکن $\qquad \sum \Pi_{i\cdot} = 1$

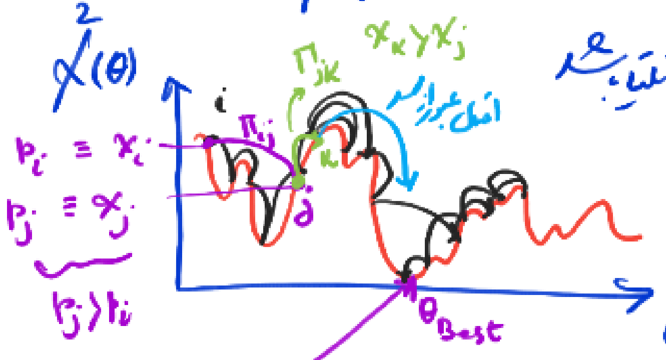
قبول رفتن به بالاتر باید وجود داشته باشد.

همچنین انتظار داریم که

$$T^{\overset{?}{n}}(\text{initial state}) = \text{final State}$$

همچنین انتظار داریم که final State تکرار شود

فرض کنید که یک پارامتر را داریم



$\longrightarrow$ landscape

Global minima $\equiv$ final State

یکی از مهم‌ترین الگوریتم‌هایی که شکل عناصر ماتریس گذار Π_{ij} را ارائه می‌دهد

الگوریتم Metropolis است

$$\Pi_{ij} = \alpha_{ij} \times \underbrace{\min\left\{1, \frac{P_j}{P_i}\right\}}_{p_{ij}}$$

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

به بیان دیگر

$$\begin{cases} \Pi_{ij} = \alpha_{ij} \times 1 & \text{for} \quad p_j \geq p_i \quad i \neq j \\ \Pi_{ij} = \alpha_{ij} \times \dfrac{p_j}{p_i} & \text{for} \quad p_j < p_i \quad i \neq j \end{cases}$$

احتمال اینکه در کمینه‌های موضعی به دام بیفتیم را کاهش می‌دهد.