« استنتاج بیزی و الگوریتم MCMC »

درس کیهان شناسی

بهار ۱۴۰۲

استنتاج بیزی (Bayesian) :

در این مبحث قصد داریم به کاربردهای استنتاج بیزی در کیهان شناسی بپردازیم. پیش از هر چیز خوب است بدانید که در آمار و احتمالات ما می توانیم دو دید گاه به مسائل داشته باشیم: دید گاه اول که دید گاه بیزی (Bayesian) نام دارد، بر این دلالت دارد که احتمال نشان دهنده سطح دانش ما نسبت به یک پدیده است و نه نشان دهنده احتمال وجود آن پدیده. درواقع در این دید گاه احتمال را یک پدیده فیزیکی نمی بینیم. اما دید گاه دیگری نیز وجود دارد که دید گاه frequentist نام دارد و ادعا می کند احتمال یک علم است که تصادفی بودن پدیده ها را توصیف می کند. دید گاهی که ما از آن بهره خواهیم جست، دید گاه اول یعنی بیزی می باشد. معادله کلی تئوری بیز برای استنتناج پارامترهای مسئله به شکل زیر است:

$$P(hypothesis|data, model) = \frac{\mathcal{L}(data|hypothesis, model) \times \ pr(hypothesis|model)}{E(data|model)}$$

هر کدام از کمیت های به کار رفته در رابطه بالا به شرح زیر است:

- $\mathcal{L}(data|hypothesis, model)$: این کمیت نشان دهنده تابع likelihood می باشد. این تابع سطح دانش ما درباره درست بودن فرضیه ای که ارائه دادیم را با توجه به دیتا و مدل مورد نظر نشان می دهد. این تابع یک توزیع احتمالی نیست بنابراین نسبت دادن آن به داده ها کار درستی نیست. درواقع می توان آن را به پارامترهایی که می خواهیم پیش بینی کنیم نسبت داد.

- $pr(hypothesis|model)$: این کمیت تابع prior را معرفی می کند. این تابع نشان دهنده سطح دانش ما نسبت به درستی فرضیه مورد نظر است پیش از آن که دیتایی که در اختیار داریم را بررسی کنیم.

- $P(hypothesis|data, model)$: این کمیت تابع posterior را نشان می دهد که تابع احتمال prior را پس از تاثیر دادن اطلاعاتی که از مدل و دیتا در اختیار داریم به روز رسانی می کند.

- $E(data|model)$: این کمیت که evidence نام دارد، در بسیاری از مسائل تحلیل داده نادیده گرفته می شود چرا که به سادگی می توان آن را یک ضریب بهنجارش دانست که به طور مشخص به فرضیه ما وابسته نمی باشد. در بعضی از مسائل مانند مسائل انتخاب مدل، این کمیت بسیار مهم می باشد اما در مسائلی که مربوط به تخمین پارامتر می شوند معمولا نقش مهمی ایفا نمی کند.

بنابراین می توان نوشت :

$$P(hypothesis|data, model)$$
$$\propto \mathcal{L}(data|hypothesis, model) \times\ pr(hypothesis|model)$$

اگرچه بیشتر با شکل لگاریتمی فرمول بالا کار می کنند. بنابراین می توان نوشت:

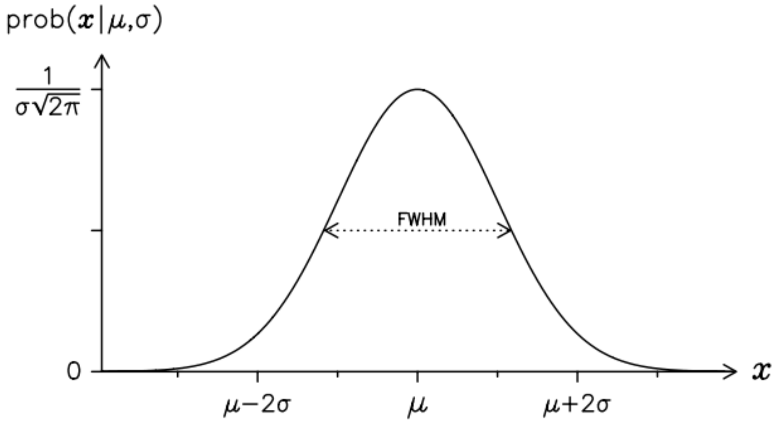
$$\log{(P)} \propto \log{(\mathcal{L})} \times \log{(pr)}$$

به عنوان مثال، اگر تابع likelihood را به صورت تابعی گوسی در نظر بگیریم، می توان برای آن نوشت:

$$\mathcal{L} \propto \prod_i^n \frac{1}{2\pi\sigma_i^2} \exp{(-\frac{(x_i-\mu)^2}{2\sigma_i^2})}$$

$$log(\mathcal{L}) = -log\left(2\pi\sigma_i^2\right) - \frac{(x_i-\mu)^2}{2\sigma_i^2}$$

که در رابطه فوق، x_i نقاط دیتا، σ_i^2 واریانس دیتای x_i و μ نقطه x_i می باشد وقتی که آن را در مدل مورد نظر خود قرار می دهیم.



شکل۱. توزیع گوسی یک تابع که در $x=\mu$ نسبت به ماکزیمم تابع متقارن می باشد.

الگوریتم متروپولیس (MCMC)

اکنون می توانیم به بررسی الگوریتم متروپولیس بپردازیم. این الگوریتم یکی از الگوریتم های به کار رفته در روش های مونت کارلو می باشد. شگرد این روش به این صورت است که ابتدا یک نقطه آغازین x_0 در نظر میگیریم و سپس به صورت رندوم یک نقطه دیگر را انتخاب می کنیم که آن را در اینجا x^{trial} می نامیم. حال باید تصمیم بگیریم که آیا نقطه رندومی که انتخاب کرده ایم قابل قبول است یا خیر. و اگر قابل قبول است، با چه احتمالی؟

$$p_{accept} = \min\left(1, {P\left(x^{trial}\right)}\Big/{P(x_i)}\right)$$

احتمال فوق، احتمال قبول شدن نقطه انتخابی می باشد که به تابع posterior در نقطه ای که هستیم یعنی x_i و در نقطه رندوم یعنی x^{trial} بستگی دارد. اکنون با توجه به عدد به دست آمده برای p_{accept} ، میتوان سه راه متفاوت را در برخورد با x^{trial} انتخاب کرد:

- اگر که $p_{accept} \geq 1$ باشد، این نقطه را انتخاب خواهیم کرد و به جای x^{trial} نقطه x_i را قرار خواهیم داد.
- اگر $0 < p_{accept} < 1$ باشد، نقطه x^{trial} با احتمال p_{accept} انتخاب خواهد شد و به نقطه x_i خواهد رفت یا این که با احتمال $1 - p_{accept}$ رد خواهد شد و به نقطه x_{i-1} خواهد رفت.

این الگوریتم را از $i = 1$ تا یک تعداد دلخواه تکرار می کنیم تا بتوانیم تعداد خوبی از جواب های مورد قبول را به دست آوریم. (دقت کنید که هر چه این تعداد بیشتر باشد، نمونه گیری شما دقیق تر خواهد بود.)

منابع برای مطالعه بیشتر:

- کتاب های مرجع :

1. David J.C. MacKay : Information Theory, Inference, and Learning algorithms

2. D.S.Sivia and J.Skilling : data analysis – a Bayesian tutorial

- ویدئوهای مفید :

1. Understanding Metropolis-Hastings algorithm

2. Posterior Predictive Distribution - Proper Bayesian Treatment!