

Podstawowy schemat analizy bayesowskiej

W celu omówienia podstawowego schematu analizy bayesowskiej rozważmy sytuację, w której do modelowania danego wektora obserwacji, y , zbudowany jest tylko jeden model bayesowski (a nie kilka alternatywnych, których – na marginesie – stopień dobroci dopasowania do danych mógłby/powinien być dodatkowym przedmiotem badania). Podstawowy modelowanie bayesowskiego prezentuje się następująco:

1. Specyfikujemy **model próbkowy**, $p(y|\theta)$, tj. określamy postać rozkładu, z którego – przy danych parametrach – pochodzą modelowane obserwacje; tym samym uzyskujemy funkcję wiarygodności:

$$L_y(\theta) = p(y|\theta).$$

2. Specyfikujemy **rozkład a priori**, $p(\theta)$, który odzwierciedla naszą wstępną wiedzę (nasze przekonania) odnośnie parametrów modelu, czy – mówiąc nieformalnie – „kształtu niepewności” związanej z parametrami. Uwaga: Na razie jeszcze nie wiadomo, jak dobierać rozkład *a priori*, ale niebawem się dowiemy :)

3. Konstruujemy **model bayesowski**, zapisując funkcję gęstości łącznego rozkładu obserwacji i parametrów:

$$p(y, \theta) = p(y|\theta)p(\theta).$$

4. Wyznaczamy i poddajemy analizie **rozkład a posteriori**:

$$p(\theta|y) = \frac{p(y, \theta)}{p(y)} \propto p(y, \theta). \quad (1)$$

5. Wyznaczamy i poddajemy analizie rozkład predyktywny (ten etap możemy pominąć, jeśli model ma służyć tylko opisowi danego zjawiska, a nie predykcji):

$$p(y^f|y) = \int_{\Theta} p(y^f, \theta|y) d\theta = \int_{\Theta} \underbrace{p(y^f|\theta, y)}_{\substack{\text{próbkowa} \\ \text{gęstość} \\ \text{predyktywna}}} \cdot \underbrace{p(\theta|y)}_{\substack{\text{gęstość} \\ \text{a posteriori} \\ \text{parametrów} \\ \text{(zob.1)}}} d\theta =$$

Warto w tym miejscu zaznaczyć, że obliczenie wartości brzegowej gęstości wektora obserwacji dla danego wektora obserwacji, tj. $p(y)$ (zob. mianownik wyrażenia 1), zwykle nie jest konieczne do wykonania powyższego schematu. Wynika to z faktu, iż w praktyce, przystępując do wyznaczenia rozkładu *a posteriori*, wprzód rozeznajemy, czy jego *jądro* (← wyjaśnienie pojęcia w przypisie 2) jest jądrem któregoś ze znanych rozkładów prawdopodobieństwa (tylko względem θ , bo wektor danych, y – z punktu widzenia $p(\theta|y)$ – traktujemy jako ustalony):¹

- jeśli tak, to stałą normującą możemy dobrać „automatycznie” (znane rozkłady prawdopodobieństwa mają znane stałe normujące);
- jeśli nie, to w celu analizy rozkładu *a posteriori* musimy sięgnąć po metody numeryczne, w wyniku których uzyskujemy próbę pseudolosową wygenerowaną z rozkładu *a posteriori*, przy czym zadanie to daje się wykonać bez znajomości $p(y)$.

Tak więc, podsumowując, do wyznaczenia charakterystyk rozkładu *a posteriori* nie jest potrzebna znajomość wartości brzegowej gęstości wektora obserwacji, $p(y)$. Ta staje się koniecznością dopiero w sytuacji, gdy chcemy porównać (pod względem dobroci dopasowania do danych) różne modele bayesowskie zbudowane do wyjaśnienia tych samych obserwacji (tj. tego samego wektora y ; zob. kolejny podpunkt, 1.B). Niestety, analityczne wyprowadzenie formuły $p(y)$ jest możliwe tylko w prostych modelach bayesowskich; w tych bardziej zaawansowanych zachodzi konieczność numerycznej aproksymacji $p(y)$ (w każdym z rozważanych, alternatywnych modeli). Co gorsza, jak dowodzi praktyka, precyzyjne oszacowanie $p(y)$ nie należy do prostych zadań... Na chwilę obecną odrzucamy jednak ten problem.

Na zakończenie niniejszego podpunktu chciałbym uczynić jeszcze jedną, ważną uwagę – można by rzec, „meta-modelarską”. Otóż w powyższym schemacie powinien zostać umieszczony jeszcze jeden punkt, którego treść wydaje mi się adekwatna zarówno w kontekście wnioskowania bayesowskiego, jak i na gruncie „klasycznym”: Zachowaj postawę krytyczną wobec „swojego” modelu...

¹ Aby wyjaśnić, czym jest *jądro* funkcji gęstości $p(x)$ rozkładu prawdopodobieństwa zmiennej losowej x , zauważmy, że dla każdego rozkładu jego gęstość $p(x)$ możemy przedstawić jako iloczyn dwóch czynników:

$$p(x) = c\tilde{p}(x),$$

gdzie c jest tzw. *stałą normującą* (nie zależy funkcyjnie od x), gwarantującą, że całka z $p(x)$ po całej dziedzinie zmiennej x wynosi 1, natomiast $\tilde{p}(x)$ jest tym całym czynnikiem we wzorze na $p(x)$, który zależy funkcyjnie od x . Ten drugi czynnik, czyli $\tilde{p}(x)$, nazywamy właśnie *jądrem* rozkładu (albo gęstości). Przy okazji zauważmy, że

$$\int_R \tilde{p}(x) dx = \frac{1}{c}.$$

Dla przykładu rozważmy 1-wymiarowy rozkład normalny o gęstości: $f_N(x|\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$. (c.d. na nast. str.)

W powyższym wzorze stałą normującą c jest wyrażenie $\frac{1}{\sigma\sqrt{2\pi}}$ (nie zależy od argumentu x), zaś jądrem – czynnikiem $\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$ (zależy od x).