Zestaw 1 - Zadanie 12

Metody probabilistyczne w uczeniu maszynowym

Łukasz Trzos

Treść zadania

Chcemy dokonać zakupu na znanym portalu aukcyjnym. Jeden ze sprzedawców ma 90 ocen pozytywnych i 10 negatywnych, natomiast drugi 2 oceny pozytywne i żadnej negatywnej. Przeprowadź analizę bayesowską i na tej podstawie zdecyduj, u którego sprzedawcy dokonasz zakupu.

Rozwiązanie

Rozkładem a priori jest $Beta(1, 1)$, który jest równoważny rozkładowi jednostajnemu $Uni([0, 1])$ Rozkład parametru wiarygodności sprzedawcy a posteriori obliczamy ze wzoru:

$$p(\theta|D) = \frac{p(D|\theta)p(\theta)}{p(D)}$$

gdzie:

- $p(\theta)$ to rozkład a priori naszego parametru, o którym wiemy, że jest równy $Uni([0, 1])$, zatem $p(\theta) = 1$
- $p(D|\theta)$ to prawdopodobieństwo otrzymania dokładnie takich opinii, jakie mamy dane w zadaniu, zakładając że faktyczna wiarygodność sprzedawcy wynosi θ. Jeśli w D wystąpiło w jakiejś kolejności x pozytywnych oraz y negatywnych opinii, to prawdopodobieństwo tego zdarzenia wynosi $\theta^x(1-\theta)^y$

Zatem:

$$p(\theta|D) = \frac{1}{p(D)}\theta^x(1-\theta)^y$$

gdzie $p(D)$ jest pewną stałą, która normalizuje funkcję gęstości, aby jej suma na przedziale $[0, 1]$ wynosiła 1. Widzimy, że część zależna od θ jest taka sama, jak w rozkładzie $Beta(x + 1, y + 1)$, zatem stała również musi dopasować się do tego rozkładu, inaczej gęstość nie byłaby poprawna. Zatem rozkładem parametru wiarygodności a posteriori jest $Beta(x + 1, y + 1)$.

Wiarygodności sprzedawców mają więc kolejno rozkłady $Beta(91, 11)$ i $Beta(3, 1)$ W celu wyznaczenia prawdopodobieństwa, że pierwszy sprzedawca jest bardziej wiarygodny od drugiego, skorzystamy z kodu na rysunku Figure 1

Po jego uruchomieniu z $z = 10^9$ otrzymujemy przybliżenie $P(X > Y) \approx 71,26\%$. Zatem lepszą decyzją jest dokonanie zakupu u pierwszego sprzedawcy.

Możemy też wyznaczyć to samo prawdopodobieństwo licząc odpowiednią całkę:

$$P(X > Y) = \int_0^1 \int_0^x f_X(x)\ f_Y(y)\ dy\ dx$$

gdzie $f_X(x) = \frac{x^{90}(1-x)^{10}}{B(91,11)} = (\frac{101!}{90!\ 10!})x^{90}(1-x)^{10}$
oraz $f_Y(y) = \frac{y^2(1-y)^0}{B(3,1)} = 3y^2$
W celu policzenia ostatniej całki skorzystamy z kalkulatora internetowego:

$$P(X > Y) = (\frac{101!}{90!\ 10!})\int_0^1 x^{90}(1-x)^{10}\int_0^x 3y^2\ dy\ dx = (\frac{101!}{90!\ 10!})\int_0^1 x^{93}(1-x)^{10}\ dx \approx 0,7126$$

```python
import random

z = int(input("Enter the number of tries:"))

counter = 0

for _ in range(z):
    X = random.betavariate(91,11)
    Y = random.betavariate(3,1)

    if X > Y:
        counter += 1

print(f"X was larger in {counter*100/z}% of all tests")
```

Rysunek 1: Kod wykorzystany do obliczenia $P(X > Y)$