Zestaw 1 - Zadanie 2

Metody probabilistyczne w uczeniu maszynowym

Łukasz Trzos

Treść zadania

Wykaż, że dla zmiennej X o rozkładzie $Beta(\alpha, \beta)$:

- $E[X] = \frac{\alpha}{\alpha+\beta}$,
- $M[X] = \frac{\alpha-1}{\alpha+\beta-2}$,
- $Var[X] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$,

gdzie $M[X]$ oznacza dominantę (modę) zmiennej X (ma sens dla $\alpha, \beta > 1$).

Rozwiązanie

Wykorzystywane własności

W rozwiązaniu skorzystamy z własności funkcji Beta. Dla $x, y \in \mathbb{R}+$:

$$B(x,y) = \int_0^1 t^{x-1}(1-t)^{y-1}\ dt = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$$

gdzie funkcja Γ jest rozszerzeniem pojęcia silnii na liczby rzeczywiste, w szczególności dla $x > 0$:

$$\Gamma(x+1) = x\Gamma(x)$$

Wartość oczekiwana

Z definicji wartości oczekiwanej dla zmiennej ciągłej:

$$E[X] = \int_{-\infty}^{\infty} x f_X(x)\ dx = \int_0^1 \frac{x^\alpha(1-x)^\beta}{B(\alpha,\beta)} = \frac{1}{B(\alpha,\beta)}\int_0^1 x^\alpha(1-x)^\beta = \frac{B(\alpha+1,\beta)}{B(\alpha,\beta)} =$$

$$= \frac{\Gamma(\alpha+1)\Gamma(\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+1)\Gamma(\alpha)\Gamma(\beta)} = \frac{\alpha}{\alpha+\beta}$$

$\square$

Wariancja

Korzystamy ze wzoru $Var[X] = E[X^2] - E[X]^2$:

$$E[X^2] = \int_{-\infty}^{\infty} x^2 f_X(x)\ dx = \int_0^1 \frac{x^{\alpha+1}(1-x)^\beta}{B(\alpha,\beta)} = \frac{1}{B(\alpha,\beta)}\int_0^1 x^{\alpha+1}(1-x)^\beta = \frac{B(\alpha+2,\beta)}{B(\alpha,\beta)} =$$

$$= \frac{\Gamma(\alpha+2)\Gamma(\beta)\Gamma(\alpha+\beta)}{\Gamma(\alpha+\beta+2)\Gamma(\alpha)\Gamma(\beta)} = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}$$

z czego wynika:

$$Var[X] = \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)} - \frac{\alpha^2}{(\alpha+\beta)^2} = \frac{(\alpha+1)\alpha(\alpha+\beta) - \alpha^2(\alpha+\beta+1)}{(\alpha+\beta)^2(\alpha+\beta+1)} = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

□

Dominanta

Dla funkcji o rozkładzie ciągłym, dominanta to argument, dla którego funkcja gęstości ma największą wartość. W celu jej wyznaczenia obliczamy pochodną po funkcji gęstości rozkładu Beta:

$$f_X(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}$$

$$f'_X(x) = \frac{1}{B(\alpha,\beta)}((\alpha-1)x^{\alpha-2}(1-x)^{\beta-1} - (\beta-1)x^{\alpha-1}(1-x)^{\beta-2}) = \frac{x^{\alpha-2}(1-x)^{\beta-2}}{B(\alpha,\beta)}((\alpha-1)(1-x)-(\beta-1)x)$$

Funkcja nie może posiadać maksimum globalnego w punkcie 0 lub 1, bo jej wartość wynosi w nich 0, a przy założeniach $\alpha, \beta > 1$, wszystkie inne wartości są dodatnie. Zatem maksimum globalne jest możliwe jedynie w punkcie, dla którego zachodzi:

$$(\alpha-1)(1-x) - (\beta-1)x = 0$$

czyli:

$$\alpha - 1 = (\beta-1)x + (\alpha-1)x = (\alpha+\beta-2)x$$

Jest to funkcja liniowa, jedynym punktem, który ją zeruje jest:

$$x = \frac{\alpha-1}{\alpha+\beta-2}$$

Musi to być maksimum lokalne, a więc również globalne. Gdyby było to minimum, to w połączeniu z faktem, że funkcja ma jedno ekstremum na przedziale $(0,1)$, jej największa wartość byłaby w 0 lub 1, co, jak wcześniej zauważyliśmy, jest niemożliwe Wiemy też, że $x \in (0,1)$, ponieważ z $\alpha > 1$ wnioskujemy, że licznik jest dodatni, a z $\beta > 1$ wynika, że mianownik jest większy od licznika □