

دسته بند Bayes

Hasan Hammad
@Hasanhammad13

دسته بندی Bayes

- ▶ دسته بندی Bayes کلاسی را انتخاب می کند که بیشترین احتمال پسین را داشته باشد.
- ▶ بیز ساده را می توان یک مدل بر مبنای **احتمال شرطی** در نظر گرفت. فرض کنید $X=(x_1,...,x_n)$ برداری از n ویژگی را بیان کند که به صورت متغیرهای **مستقل** هستند و دارای **توزیع نرمال** هستند. به این ترتیب می توان احتمال رخداد C_k یعنی $P(C_k|x_1,...,x_n)$ را به عنوان یکی از حالت های کلاس رخدادهای مختلف به ازای k های متفاوت، به شکل زیر نمایش داد.

$$p(C_k | X) = \frac{p(C_k) p(X | C_k)}{p(X)}$$

کلاس

«پیشین» (Prior)

«درست نمایی» (Likelihood)

$P(C_k \text{ and } X)$

محاسبه نمی شود در دسته بندی bayes

دیتا برداری از ویژگی ها

مثال

- ▶ در این مثال می خواهیم یک رشته از DNA دسته بندی بکنیم با استفاده از دسته بند bayes
- ▶ اطلاعات جدول زیر در نظر بگیرید:

نوع داده	رشته DNA	کلاس
آموزش	C G T	0
آموزش	G C T	0
آموزش	A G C	1
آموزش	A C A T	1
تست	T G C	؟

- ▶ حال می خواهیم احتمال تعلق داده ی تست به هر کلاس محاسبه بکنیم.

مثال

$$D=X$$

کلاس

$$P(C|D) = \frac{P(C)P(D|C)}{P(D)}$$

نمونه تست

نمونه تست (D): T G C

کلاس 0

$$P(C = 0|D) = \frac{P(C = 0)P(D|C = 0)}{P(D)}$$

کلاس 1

$$P(C = 1|D) = \frac{P(C = 1)P(D|C = 1)}{P(D)}$$

دقت کنید که مخرج کسر برای این دو تا کلاس یکسان است
و نیازی به محاسبه آن ندارد

مثال

نمونه تست $T G C : (D)$

کلاس

$$P(C|D) = \frac{P(C)P(D|C)}{P(D)}$$

نمونه تست

کلاس 0

$$P(C = 0|D) = \underbrace{P(C = 0)P(D|C = 0)}$$

$$P(C = 0) = \frac{1}{2} = 2/4$$

احتمال انتخاب کلاس 0

کلاس 1

$$P(C = 1|D) = \underbrace{P(C = 1)P(D|C = 1)}$$

$$P(C = 1) = \frac{1}{2} = 2/4$$

احتمال انتخاب کلاس 1

مثال

نمونه تست (D): T G C

کلاس

$$P(C|D) = \frac{P(C)P(D|C)}{P(D)}$$

نمونه تست

کلاس 0

$$P(C = 0|D) = P(C = 0)P(D|C = 0)$$

$$P(C = 0) = \frac{1}{2}$$

$$P(D|C = 0) = P(D1|0) \times P(D2|0) \times P(D3|0)$$

$$P(D|C = 0) = P(T|0) \times P(G|0) \times P(C|0)$$

ویژگی های نمونه ی
تست

کلاس 1

$$P(C = 1|D) = P(C = 1)P(D|C = 1)$$

$$P(C = 1) = \frac{1}{2}$$

$$P(D|C = 1) = P(D1|1) \times P(D2|1) \times P(D3|1)$$

$$P(D|C = 1) = P(T|1) \times P(G|1) \times P(C|1)$$

ویژگی های نمونه ی
تست

مثال

نوع داده	رشته DNA	کلاس
آموزش	C G T	0
آموزش	G C T	0
آموزش	A G C	1
آموزش	A C A T	1
تست	T G C	?

نمونه تست (D): T G C

کلاس 0

کلاس 1

$$P(C = 0|D) = P(C = 0) \underbrace{P(D|C = 0)}$$

$$P(C = 0) = \frac{1}{2}$$

$$P(D|C = 0) = P(D1|0) \times P(D2|0) \times P(D3|0)$$

$$P(D|C = 0) = P(T|0) \times P(G|0) \times P(C|0)$$

$$P(T|0) = \frac{\text{تعداد آن در کلاس 0}}{\text{تعداد کل ویژگی ها در کلاس 0}} = \frac{2}{6}$$

$$P(G|0) = \frac{\text{تعداد آن در کلاس 0}}{\text{تعداد کل ویژگی ها در کلاس 0}} = \frac{2}{6}$$

$$P(C|0) = \frac{\text{تعداد آن در کلاس 0}}{\text{تعداد کل ویژگی ها در کلاس 0}} = \frac{2}{6}$$

$$P(C = 1|D) = P(C = 1) \underbrace{P(D|C = 1)}$$

$$P(C = 1) = \frac{1}{2}$$

$$P(D|C = 1) = P(D1|1) \times P(D2|1) \times P(D3|1)$$

$$P(D|C = 1) = P(T|1) \times P(G|1) \times P(C|1)$$

$$P(T|1) = \frac{\text{تعداد آن در کلاس 1}}{\text{تعداد کل ویژگی ها در کلاس 1}} = \frac{1}{7}$$

$$P(G|1) = \frac{\text{تعداد آن در کلاس 1}}{\text{تعداد کل ویژگی ها در کلاس 1}} = \frac{1}{7}$$

$$P(C|1) = \frac{\text{تعداد آن در کلاس 1}}{\text{تعداد کل ویژگی ها در کلاس 1}} = \frac{2}{7}$$

مثال

نمونه تست (D): T G C

کلاس

$$P(C|D) = \frac{P(C)P(D|C)}{P(D)}$$

نمونه تست

کلاس 0

$$P(C = 0|D) = P(C = 0) \underbrace{P(D|C = 0)}$$

$$P(C = 0) = \frac{1}{2}$$

$$P(D|C = 0) = P(D1|0) \times P(D2|0) \times P(D3|0)$$

$$P(D|C = 0) = P(T|0) \times P(G|0) \times P(C|0)$$

$$P(D|C = 0) = \frac{2}{6} \times \frac{2}{6} \times \frac{2}{6}$$

کلاس 1

$$P(C = 1|D) = P(C = 1) \underbrace{P(D|C = 1)}$$

$$P(C = 1) = \frac{1}{2}$$

$$P(D|C = 1) = P(D1|1) \times P(D2|1) \times P(D3|1)$$

$$P(D|C = 1) = P(T|1) \times P(G|1) \times P(C|1)$$

$$P(D|C = 1) = \frac{1}{7} \times \frac{1}{7} \times \frac{2}{7}$$

مثال

نمونه تست (D): T G C

کلاس

$$P(C|D) = \frac{P(C)P(D|C)}{P(D)}$$

نمونه تست

کلاس 0

$$P(C = 0|D) = P(C = 0)P(D|C = 0)$$

$$P(C = 0) = \frac{1}{2}$$

$$P(D|C = 0) = \frac{2}{6} \times \frac{2}{6} \times \frac{2}{6}$$

$$P(C = 0|D) = \frac{1}{2} \times \frac{2}{6} \times \frac{2}{6} \times \frac{2}{6}$$

$$P(C = 0|D) = \frac{4}{216} = 0.0185$$

کلاس 1

$$P(C = 1|D) = P(C = 1)P(D|C = 1)$$

$$P(C = 1) = \frac{1}{2}$$

$$P(D|C = 1) = \frac{1}{7} \times \frac{1}{7} \times \frac{2}{7}$$

$$P(C = 1|D) = \frac{1}{2} \times \frac{1}{7} \times \frac{1}{7} \times \frac{2}{7}$$

$$P(C = 1|D) = \frac{1}{343} = 0.0029$$

مثال

نمونه تست (D): T G C

کلاس

$$P(C|D) = \frac{P(C)P(D|C)}{P(D)}$$

نمونه تست

کلاس 0

$$P(C = 0|D) = P(C = 0)P(D|C = 0)$$

$$P(C = 0|D) = 0.0185 \quad 18.6\%$$

کلاس 1

$$P(C = 1|D) = P(C = 1)P(D|C = 1)$$

$$P(C = 1|D) = 0.0029 \quad 0.0214 \Rightarrow 14\%$$

احتمال کلاس 0 بیشتر از احتمال کلاس 1
پس نمونه تست به کلاس 0 تعلق دارد.

$$P(D) = 0.0185 + 0.0029 = 0.0214$$