

۹۹۵۲۱۲۷۱

تبرین هشتم هوش مصنوعی

فرزان رحمانی

الف ۱

$$P(F = \text{true}) = \sum_s P(F = \text{true}, S = s)$$

$$= \sum_s P(S = s, F = \text{true}) P(S = s) = P(F = \text{true} | S = \text{پاییز}) P(S = \text{پاییز}) + P(F = \text{true} | S = \text{بهار}) P(S = \text{بهار}) = 0,4 \times 0,5 + 0,1 \times 0,5 = 0,2 + 0,05 = 0,25$$

(در گراف رسم شده، F به طور مستقیم وابسته به S است و متغیرهای دیگر اثر نمی گذارند.)

ب) طبق اطلاعات جدول داده شده:

$$P(F = \text{true} | S = \text{پاییز}) = 0,4$$

ج) می دانیم رابطه زیر در Bayes Net برقرار است پس با توجه به آن به حل می پردازیم:

$$P(X_1, X_2, \dots, X_n) = \prod_i P(X_i | \text{parents}(X_i))$$

$$P(F = \text{true} | S = \text{پاییز}, H = \text{true}, D = \text{true}) = \frac{P(F = \text{true}, S = \text{پاییز}, H = \text{true}, D = \text{true})}{P(S = \text{پاییز}, H = \text{true}, D = \text{true})}$$

$$= \frac{P(S = \text{پاییز}) P(D = \text{true} | S = \text{پاییز}) P(F = \text{true} | S = \text{پاییز}) P(H = \text{true} | F = \text{true}, D = \text{true})}{\sum_f P(S = \text{پاییز}) P(D = \text{true} | S = \text{پاییز}) P(F = f | S = \text{پاییز}) P(H = \text{true} | F = f, D = \text{true})}$$

$$= \frac{0,5 \times 0,1 \times 0,4 \times 0,9}{0,5 \times 0,1 \times 0,4 \times 0,9 + 0,5 \times 0,1 \times 0,4 \times 0,1} = \frac{0,018}{0,018 + 0,002} \approx 0,4286$$

۲) قسمت الف و ب مطالعه نشر و عملیات مختلف روی دیتاست تحلیل نشر.

ج)

$$P(\text{Petal Width} = XS | \text{Species} = \text{Iris-setosa}) = \frac{\overset{\text{تعداد}}{n}(\text{Petal Width} = XS, \text{Species} = \text{Iris-setosa})}{n(\text{Species} = \text{Iris-setosa})}$$

$$P(\text{Petal Width} = XS | \text{Species} = \text{Iris-setosa}) = \frac{۳۴}{۵۹} = 0,۵۷۶۱$$

* کو محاسبه تعداد در نوت بوک درج شد.

$$P(\text{Species} = \text{Iris-Virginica} | \text{Sepal Length} = L) = \frac{n(\text{Species} = \text{Iris-Virginica}, \text{Sepal Length} = L)}{n(\text{Species} = \text{Iris-Virginica})}$$

$$= \frac{31}{42} \approx 0,7381$$

(>) خیر نمی توان. جدول توزیع توام احتمال توام این دیتاست دارای ۷ ویژگی است. ویژگی نوع گیاه دارای سه مقدار متفاوت است و بقیه ویژگی ها دارای ۴

مقدار مختلف می باشند. پس در مجموع این جدول دارای $3 \times 4 \times 4 \times 4 \times 4 \times 4$

خانه (حالت مختلف) می باشد. $3 \times 4^4 = 768$ خانه مختلف در این

جدول موجود است. در حالی که دیتاست فقط دارای ۱۵۰ نمونه می باشد.

یعنی در بهترین حالت پیش یکسان داده ها $768 - 150 = 618$ خانه از جدول

خالی می ماند و حتی یک نمونه هم از این حالت در جدول موجود نیست. این نشان می دهد

که حدوداً ۸۰٪ $\frac{618}{768} \approx 80\%$ از خانه های جدول توزیع توام خالی هستند و دارای

احتمال صفر هستند. لذا ۱۵۰ نمونه برای این جدول ۷۶۸ نمی تواند کافی باشد و ~~برآورد~~

هایی با خطای زیادی تولید می کنند و قابل اتکا نیست. برای پیدا کردن حداقل

تعداد نمونه با ضریب اطمینان ^(وضعی) مطلوب می توان از جداول آماری توزیع های مختلف

و همچنین فرمول های آماری مختلف کمک بگیریم.

مثالی از یک فرمول آماری

$$n = \left(\frac{z^*}{M} \right)^2 \hat{p} (1 - \hat{p})$$

(Finding Sample Size for estimating a Population proportion)

M is the margin of error

\hat{p} is the estimated value of the proportion

(if has no idea about \hat{p} then we use $\hat{p} = 0,5$ that is most conservative)