

VeriMadenciliği – Data Mining

HW – 4 – çözüm

1. Bir A ülkesinde XX hastalığının görülme yaygınlığı %25 imiş. Bu hastalığı teşhiste kullanılan bir YY yöntemi var ama mükemmel çalışmıyor. Hasta birisi için hasta diye yaptığı teşhis %99 başarı ile doğru oluyor. Hasta olmayan birisi için hasta değil diye yaptığı teşhis ise %93 ile doğru oluyor. Buna göre test yaptıran birisi hastasın sonucunu aldıysa gerçekten hasta olma olasılığı nedir?

H hasta, !H hasta değil anlamına gelsin. + testin pozitif çıkması, - de negative çıkması olsun.

Verilenler: $P(H) = 0.25$ $P(!H) = 0.75$ $P(+ | H) = 0.99$ $P(- | !H) = 0.93$

Bulmak istediğimiz olasılık test pozitif çıktı ise hasta olma olasılığı yani $P(H | +)$

$$P(H | +) = P(+ | H) P(H) / P(+)$$

Burdatek bulmamız gereken $P(+)$. Test iki türlü + olur, ya hastadır doğru bilir ya da hasta değildir yanlışlıkla hasta der.

$$\text{Hastayken + çıkması} = \text{hasta olma olasılığı} \times \text{hasta olduğunu bilirken + çıkması} = 0.25 \times 0.99 = 0.2475$$

$$\text{Sağlıklıyken + çıkması} = \text{sağlıklı olma olasılığı} \times \text{sağlıklıyken + çıkması} = 0.75 \times 0.07 = 0.0525$$

$$P(+) = 0.247 + 0.052 = 0.3$$

$$P(H | +) = 0.99 \times 0.25 / 0.3 = \mathbf{0.825}$$

2. Aşağıda verilen tabloyu kullanarak, $x=\{E, 3, 3\}$ (cinsiyet, boy,kilo) olarak verilen verinin hangi beden olduğunu Bayesian sınıflandırma yaparak bulun.

Gender/cinsiyet	Length/boy	Weight/kilo	Size/beden
K	1	2	Small
K	2	2	Medium
K	2	3	Medium
K	1	1	Small
E	3	3	Medium
E	3	4	Large
E	1	2	Small
E	3	5	Large
E	2	3	Medium
E	3	3	Large

$$P(\text{small}) = 3/10$$

$$P(\text{medium}) = 4/10$$

$$P(\text{large}) = 3/10$$

$$P(E \mid \text{small}) = 1 / 3$$

$$P(E \mid \text{medium}) = 2 / 4$$

$$P(E \mid \text{large}) = 3 / 3$$

$$P(\text{boy} = 3 \mid \text{small}) = 0 / 3$$

$$P(\text{boy} = 3 \mid \text{medium}) = 1 / 4$$

$$P(\text{boy} = 3 \mid \text{large}) = 3 / 3$$

$$P(\text{kilo} = 3 \mid \text{small}) = 0 / 3$$

$$P(\text{kilo} = 3 \mid \text{medium}) = 3 / 4$$

$$P(\text{kilo} = 3 \mid \text{large}) = 1 / 3$$

$$P(E, 3, 3 \mid \text{small}) = 1/3 * 0/3 * 0/3 = 0$$

$$P(E, 3, 3 \mid \text{medium}) = 2/4 * 1/4 * 3/4 = 3/32$$

$$P(E, 3, 3 \mid \text{large}) = 2/2 * 2/2 * 1/3$$

$$P(\text{small} \mid E, 3, 3) = 0 * 3/10 = 0$$

$$P(\text{medium} \mid E, 3, 3) = 3/32 * 4/10 = 0.037$$

$$P(\text{large} \mid E, 3, 3) = 1/3 * 3/10 = 0.1$$

(E,3,3) verisi LARGE beden olarak sınıflanır.

3. Aşağıdaki vize ve final sınav notlarından oluşan öğrenme kümesini kullanarak, vize notu verildiğinde final notunu tahmin edecek bir denklemi linear regression ile bulun.

Midterm (x)	Final (y)
72	84
50	63
81	78
74	78
94	90
86	78
59	49
83	88
65	66

$$\bar{x} = 73.78 \quad \bar{y} = 74.89 \quad w_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

$$w_1 = \frac{(-1.78 * 9.11) + (-23.78 * -11.89) + (7.22 * 3.11) + \dots + (9.22 * 13.11)(-8.78 * 8.89)}{(-1.78^2) + (-23.78^2) + (7.22^2) + \dots + (9.22^2) + (-8.78^2)}$$

$$\sim = 1058.5 / 1551 \sim = 0.68$$

$$w_0 = 74.89 - (0.68) * 73.78 = 24.7$$

$$y = 0.68 x + 24.7$$