6434476923 มัทธี อาคาปันธรรม

Naïve Bayesian classifier

class lable is status attribute

$P(C_i|X) = P(X|C_i)P(C_i)$

department	status	age	salary
sales	senior	31...35	46K...50K
sales	junior	26...30	26K...30K
sales	junior	31...35	31K...35K
systems	junior	21...25	46K...50K
systems	senior	31...35	66K...70K
systems	junior	26...30	46K...50K
systems	senior	41...45	66K...70K
marketing	senior	36...40	46K...50K
marketing	junior	31...35	41K...45K
secretary	senior	46...50	36K...40K
secretary	junior	26...30	26K...30K

a) Data sample : X = (marketing , 31 ... 35 , 46K ... 50K)

$P(C_i)$: P (status = " junior") = 6 / 11 = 0.5455

P (status = " senior") = 5 / 11 = 0.4545

junior = 6

senior = 5

Compute $P(X|C_i)$ for each class

P (department = " marketing" | status = " junior") = 1/6 = 0.1667

P (department = " marketing" | status = " senior") = 1/5 = 0.2

P (age = " 31 ... 35" | status = " junior") = 2/6 = 0.3333

P (age = " 31 ... 35" | status = " senior") = 2/5 = 0.4

P (salary = " 46 K ... 50k " | status = " junior") = 2/6 = 0.3333

P (salary = " 46 K ... 50k " | status = " senior") = 2/5 = 0.4

$P(X|C_i)$: P (X | status = " junior") = 0.1667 x 0.3333 x 0.3333 = 0.0185

P (X | status = " senior") = 0.2 x 0.4 x 0.4 = 0.032

$P(X|C_i) * P(C_i)$: P (X | status = " junior") * P (status = " junior") = 0.0185 x 0.5455 = 0.0101

P (X | status = " senior") * P (status = " senior") = 0.032 x 0.4545 = 0.0145 → highest

Therefore, X belongs to class (status = " senior") #

6434476923 พัทธ์ อาคาปิยะธรรม

Naïve Bayesian classifier

class lable is status attribute

$P(C_i|X) = P(X|C_i)P(C_i)$

department	*status*	*age*	*salary*
sales	senior	31...35	46K...50K
sales	junior	26...30	26K...30K
sales	junior	31...35	31K...35K
systems	junior	21...25	46K...50K
systems	senior	31...35	66K...70K
systems	junior	26...30	46K...50K
systems	senior	41...45	66K...70K
marketing	senior	36...40	46K...50K
marketing	junior	31...35	41K...45K
secretary	senior	46...50	36K...40K
secretary	junior	26...30	26K...30K

b) Data sample : X = (sale, 31...35, 66K...70K)

$P(C_i)$: P(status = "junior") = 6/11 = 0.5455

P(status = "senior") = 5/11 = 0.4545

junior = 6

senior = 5

Compute $P(X|C_i)$ for each class

P(department = "sale" | status = "junior") = 2/6 = 0.3333

P(department = "sale" | status = "senior") = 1/5 = 0.2

P(age = "31...35" | status = "junior") = 2/6 = 0.3333

P(age = "31...35" | status = "senior") = 2/5 = 0.4

~~P(salary = "66K...70K" | status = "junior") = 0/6 = 0~~ → prob เท่ากับ 0 ใช้ Laplace estimate

Generalized Laplace estimate :

Salary has 6 values (26K...30K, 31K...35K, 36K...40K, 41K...45K, 46K...50K, 66K...70K).

So, $s_i = 6$

$$P(A_i = v_j | C_K) = \frac{n_{ijk} + 1}{n_K + s_i}$$

Thus,

P(salary = "66K...70K" | status = "junior") = 0+1 / 6+6 = 1/12 = 0.0833

P(salary = "66K...70K" | status = "senior") = 2+1 / 5+6 = 3/11 = 0.2727

$P(X|C_i)$: P(X | status = "junior") = 0.3333 x 0.3333 x 0.0833 = 0.0093

P(X | status = "senior") = 0.2 x 0.4 x 0.2727 = 0.0218

$P(X|C_i) * P(C_i)$: P(X | status = "junior") * P(status = "junior") = 0.0093 x 0.5455 = 0.0051

P(X | status = "senior") * P(status = "senior") = 0.0218 x 0.4545 = 0.0099 → highest

Therefore, X belongs to class (status = "senior") #