CZ4041: Tutorial Week 4

Due on February 4, 2021 at 8:30am

Assoc Prof Pan, Sinno Jialin - CS4

Pang Yu Shao
U1721680D

04/02/2021

Problem 1

1. Estimate the conditionaal probabilities for $P(A = 1|+)$, $P(B = 1|+)$, $P(C = 1|+)$, $P(A = 1|-)$, $P(B = 1|-)$, $P(C = 1|-)$
2. Use the estimate of conditional probabilities given in the previous question to predict the class label for a test example $(A = 1, B = 1, C = 1)$ using the naive Bayes approach.

Solution

1.

$$
\begin{aligned}
P(A = 1|+) &= 0.5 \\
P(B = 1|+) &= 0.5 \\
P(C = 1|+) &= 1.0 \\
P(A = 1|-) &= 0.333 \\
P(B = 1|-) &= 0.333 \\
P(C = 1|-) &= 0.333
\end{aligned}
$$

2.

$$
\begin{aligned}
P(+|A = 1, B = 1, C = 1) &= \frac{P(+) * P(A = 1|+) * P(B = 1|+) * P(C = 1|+)}{P(A = 1, B = 1, C = 1)} \\
&= \frac{P(+) * P(A = 1|+) * P(B = 1|+) * P(C = 1|+)}{P(A = 1, B = 1, C = 1)} \\
&= \frac{0.4 * 0.5 * 0.5 * 1}{P(A = 1, B = 1, C = 1)} \\
&= \frac{\mathbf{0.1}}{\mathbf{P(A = 1, B = 1, C = 1)}}
\end{aligned}
$$

$$
\begin{aligned}
P(-|A = 1, B = 1, C = 1) &= \frac{P(-) * P(A = 1|-) * P(B = 1|-) * P(C = 1|-)}{P(A = 1, B = 1, C = 1)} \\
&= \frac{P(-) * P(A = 1|-) * P(B = 1|-) * P(C = 1|-)}{P(A = 1, B = 1, C = 1)} \\
&= \frac{0.6 * 0.333 * 0.333 * 0.333}{P(A = 1, B = 1, C = 1)} \\
&= \frac{0.0222}{P(A = 1, B = 1, C = 1)}
\end{aligned}
$$

Problem 2

On the 28th page of the lecture notes "Lecture 3", recalculate the likelihoods using m-estimate. Compare the m-estimate method and the original method shown on the 25th page for estimating probabilities. Which method is better and why?

Solution

$$\begin{aligned}
P(HomO = Yes|No) &= \frac{3 + 3 * 2/3}{7 + 3} \\
&= 0.5 \\
P(HomO = No|No) &= \frac{4 + 3 * 2/3}{7 + 3} \\
&= 0.6 \\
P(HomO = Yes|Yes) &= \frac{0 + 3 * 1/3}{3 + 3} \\
&= 1/6 \\
P(HomO = No|Yes) &= \frac{3 + 3 * 1/3}{3 + 3} \\
&= 2/3 \\
P(MaritalStatus = Single|No) &= \frac{2 + 3 * 2/3}{7 + 3} \\
&= 0.4 \\
P(MaritalStatus = Divorced|No) &= \frac{1 + 3 * 2/3}{7 + 3} \\
&= 0.3 \\
P(MaritalStatus = Married|No) &= \frac{4 + 3 * 2/3}{7 + 3} \\
&= 0.6 \\
P(MaritalStatus = Single|Yes) &= \frac{2 + 3 * 1/3}{3 + 3} \\
&= 0.5 \\
P(MaritalStatus = Divorced|Yes) &= \frac{1 + 3 * 1/3}{3 + 3} \\
&= 1/3 \\
P(MaritalStatus = Married|Yes) &= \frac{0 + 3 * 1/3}{3 + 3} \\
&= 1/6
\end{aligned}$$

This is not complete, this needs to be normalized.

$$
\begin{aligned}
P(HomO = Yes|No) &= 0.5/(0.5 + 0.6) \\
&= 0.455 \\
P(HomO = No|No) &= 0.6/(0.5 + 0.6) \\
&= 0.545 \\
P(HomO = Yes|Yes) &= (1/6)/(1/6 + 2/3) \\
&= 1/5 \\
P(HomO = No|Yes) &= (2/3)/(1/6 + 2/3) \\
&= 4/5 \\
P(MaritalStatus = Single|No) &= 0.4/(0.4 + 0.3 + 0.6) \\
&= 0.308 \\
P(MaritalStatus = Divorced|No) &= 0.3/(0.4 + 0.3 + 0.6) \\
&= 0.231 \\
P(MaritalStatus = Married|No) &= 0.6/(0.4 + 0.3 + 0.6) \\
&= 0.461 \\
P(MaritalStatus = Single|Yes) &= 0.5/(0.5 + 1/3 + 1/6) \\
&= 0.5 \\
P(MaritalStatus = Divorced|Yes) &= (1/3)/(0.5 + 1/3 + 1/6) \\
&= 1/3 \\
P(MaritalStatus = Married|Yes) &= (1/6)/(0.5 + 1/3 + 1/6) \\
&= 1/6
\end{aligned}
$$

Problem 3

Bayesian Belief Networks (Week 4)
If the person has high blood pressure, but exercises regularly and eats a healthy diet, diagnose whether he has heart disease.
Solution

$$P(HD = Yes|BP = High, D = Healthy, E = Yes) = \frac{P(HD = Yes, BP = High, D = Healthy, E = Yes)}{P(BP = High, D = Healthy, E = Yes)}$$

$$= \frac{P(BP = High|HD = Yes, D = Healthy, E = Yes)P(HD = Yes, D = Healthy, E = Yes)}{P(BP = High, D = Healthy, E = Yes)}$$

$$= \frac{P(BP = High|HD = Yes)P(HD = Yes, D = Healthy, E = Yes)}{P(BP = High, D = Healthy, E = Yes)}$$

$$= \frac{0.85 * P(HD = Yes|D = Healthy, E = Yes) * P(D = Healthy, E = Yes)}{P(BP = High|D = Healthy, E = Yes) * P(D = Heahtly, E = Yes)}$$

$$= \frac{0.85 * 0.25}{P(BP = High|D = Healthy, E = Yes)}$$

$$P(HD = No|BP = High, D = Healthy, E = Yes) = \frac{P(HD = No, BP = High, D = Healthy, E = Yes)}{P(BP = High, D = Healthy, E = Yes)}$$

$$= \frac{P(BP = High|HD = No, D = Healthy, E = Yes)P(HD = No, D = Healthy, E = Yes)}{P(BP = High, D = Healthy, E = Yes)}$$

$$= \frac{P(BP = High|HD = No)P(HD = No, D = Healthy, E = Yes)}{P(BP = High, D = Healthy, E = Yes)}$$

$$= \frac{0.2 * P(HD = No|D = Healthy, E = Yes) * P(D = Healthy, E = Yes)}{P(BP = High|D = Healthy, E = Yes) * P(D = Heahtly, E = Yes)}$$

$$= \frac{0.2 * 0.75}{P(BP = High|D = Healthy, E = Yes)}$$

Problem 3 continued on next page. . .

$$\frac{0.85 * 0.25}{P(BP = High|D = Healthy, E = Yes)} + \frac{0.2 * 0.75}{P(BP = High|D = Healthy, E = Yes)} = 1$$

$$\frac{0.3625}{P(BP = High|D = Healthy, E = Yes)} = 1$$

$$P(BP = High|D = Healthy, E = Yes) = 0.3625$$

Therefore,

$$P(HD = Yes|BP = High, D = Healthy, E = Yes) = \frac{0.85 * 0.25}{0.3625}$$

$$= \mathbf{0.5862}$$

$$P(HD = No|BP = High, D = Healthy, E = Yes) = \frac{0.2 * 0.75}{0.3625}$$

$$= \mathbf{0.4138}$$