Fundamentals of Data Science

Semester B 20-21

Tutorial 7

1. As an example, the class-conditional probability P(Home Owner=No|No) can be estimated by finding the fraction of training instances in the class No that take on the attribute value Home Owner=No, and the resulting value is 4/7. The other class-conditional probability values are estimated in a similar way.

   The sample mean and variance values of the attribute Annual Income for the class Yes are calculated as follows:

$$\bar{x}_{\text{Annual Income,Yes}} = \frac{95+85+90}{3} = 90$$

$$\sigma^2_{\text{Annual Income, Yes}} = \frac{1}{2}[(95-90)^2 + (85-90)^2 + (90-90)^2] = 25$$

2.a. The class-conditional probabilities for the attributes A, B and C are estimated as follows:

$$P(A=T\,|+) = \frac{3}{5} = 0.6, \quad P(A=F\,|+) = \frac{2}{5} = 0.4, \quad P(A=T\,|-) = \frac{2}{5} = 0.4, \quad P(A=F\,|-) = \frac{3}{5} = 0.6$$

$$P(B=T\,|+) = \frac{1}{5} = 0.2, \quad P(B=F\,|+) = \frac{4}{5} = 0.8, \quad P(B=T\,|-) = \frac{2}{5} = 0.4, \quad P(B=F\,|-) = \frac{3}{5} = 0.6$$

$$P(C=T\,|+) = \frac{2}{5} = 0.4, \quad P(C=F\,|+) = \frac{3}{5} = 0.6, \quad P(C=T\,|-) = 1, \quad P(C=F\,|-) = 0$$

b. Let $p = P(A=F, B=T, C=F)$

   The posterior probability $P(+|A=F, B=T, C=F)$ is estimated as follows:

$$\begin{aligned}
&P(+\,|\,A=F, B=T, C=F) \\
&= \frac{P(A=F, B=T, C=F\,|+)P(+)}{P(A=F, B=T, C=F)} \\
&= \frac{[P(A=F\,|+)P(B=T\,|+)P(C=F\,|+)]P(+)}{p} \\
&= \frac{[(0.4)(0.2)(0.6)](0.5)}{p} \\
&= \frac{0.024}{p}
\end{aligned}$$

The posterior probability $P(- | A=F, B=T, C=F)$ is estimated as follows:

$$\begin{aligned}
&P(- \mid A=F, B=T, C=F) \\
&= \frac{P(A=F, B=T, C=F \mid -)P(-)}{P(A=F, B=T, C=F)} \\
&= \frac{[P(A=F \mid -)P(B=T \mid -)P(C=F \mid -)]P(-)}{p} \\
&= \frac{[(0.6)(0.4)(0)](0.5)}{p} \\
&= \frac{0}{p}
\end{aligned}$$

The predicted class label is ‘+’.