

- (b) Given the information in part (a), is a randomly chosen college student more likely to be a graduate or undergraduate student?
- (c) Repeat part (b) assuming that the student is a smoker.
- (d) Suppose 30% of the graduate students live in a dorm but only 10% of the undergraduate students live in a dorm. If a student smokes and lives in the dorm, is he or she more likely to be a graduate or undergraduate student? You can assume independence between students who live in a dorm and those who smoke.

7. Consider the data set shown in Table 5.10

Table 5.10. Data set for Exercise 7.

Record	A	B	C	Class
1	0	0	0	+
2	0	0	1	-
3	0	1	1	-
4	0	1	1	-
5	0	0	1	+
6	1	0	1	+
7	1	0	1	-
8	1	0	1	-
9	1	1	1	+
10	1	0	1	+

- (a) Estimate the conditional probabilities for $P(A|+)$, $P(B|+)$, $P(C|+)$, $P(A|-)$, $P(B|-)$, and $P(C|-)$.
 - (b) Use the estimate of conditional probabilities given in the previous question to predict the class label for a test sample ($A = 0$, $B = 1$, $C = 0$) using the naïve Bayes approach.
 - (c) Estimate the conditional probabilities using the m-estimate approach, with $p = 1/2$ and $m = 4$.
 - (d) Repeat part (b) using the conditional probabilities given in part (c).
 - (e) Compare the two methods for estimating probabilities. Which method is better and why?
8. Consider the data set shown in Table 5.11.
- (a) Estimate the conditional probabilities for $P(A = 1|+)$, $P(B = 1|+)$, $P(C = 1|+)$, $P(A = 1|-)$, $P(B = 1|-)$, and $P(C = 1|-)$ using the same approach as in the previous problem.