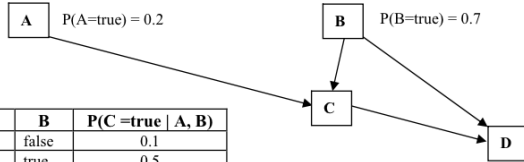


CSE 485 Midterm

Olcay Taner YILDIZ

I. QUESTION (18 POINTS)

Consider the following Bayesian Network, where variables A-D are all Boolean-valued:



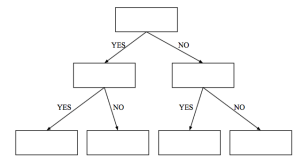
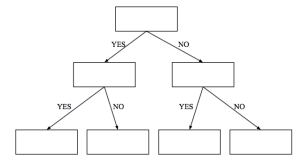
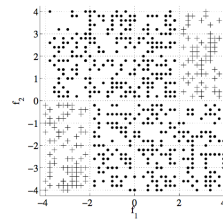
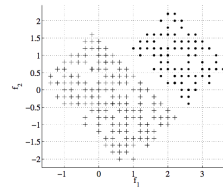
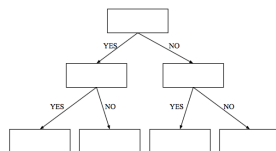
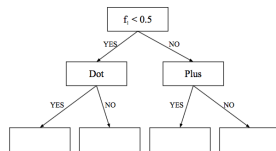
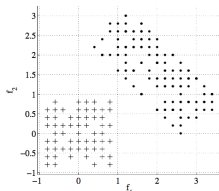
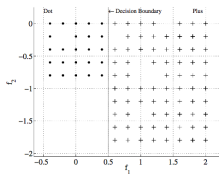
| A | B | P(C=true A, B) |
|-------|-------|------------------|
| false | false | 0.1 |
| false | true | 0.5 |
| true | false | 0.4 |
| true | true | 0.9 |

| B | C | P(D=true B, C) |
|-------|-------|------------------|
| false | false | 0.8 |
| false | true | 0.6 |
| true | false | 0.3 |
| true | true | 0.1 |

- What is the probability that all four of these Boolean variables are false?
- What is the probability that C is true, D is false, and B is true?
- What is the probability that C is true given that D is false and B is true?

II. QUESTION (16 POINTS)

For the following data, fill the decision nodes with appropriate decision conditions.



III. QUESTION (18 POINTS)

Assume you are given the following three features of TV shows, with the possible values shown, and wish to learn how to predict future top-10 shows.

$$Type \in \{Comedy, Drama, News, Sports\}$$

$$Location \in \{LA, NYC, Various\}$$

$$Duration \in \{30min, 60min\}$$

Given the data above, apply rule induction algorithm. Show the resulting ruleset.

| Type | Location | Duration | ClassInfo |
|------|----------|----------|-----------|
| C | NYC | 30 | + |
| C | LA | 60 | - |
| N | Var | 60 | + |
| N | LA | 30 | + |
| S | Var | 60 | - |
| D | LA | 60 | - |

IV. QUESTION (18 POINTS)

Assume you wish to use the K-nearest neighbors algorithm on this dataset and set aside the last two examples as a tuning set. Would you prefer $K = 1$ or $K = 3$?

| A | B | C | D | Output |
|---|---|---|---|--------|
| T | T | F | T | T |
| T | T | T | T | F |
| F | F | T | F | T |
| F | T | F | T | F |
| T | T | F | F | T |
| T | F | T | T | F |

V. QUESTION (15 POINTS)

Given the following error rates for a four feature problem, what will be the result of the following attribute subset selection procedures?

- Stepwise forward selection
- Stepwise backward elimination
- Floating search

Error Rates: (F_1) , 0.23; (F_2) , 0.15; (F_3) , 0.16; (F_4) , 0.24; (F_1, F_2) , 0.18; (F_1, F_3) , 0.19; (F_1, F_4) , 0.20; (F_2, F_3) , 0.16; (F_2, F_4) , 0.29; (F_3, F_4) , 0.22; (F_1, F_2, F_3) , 0.19; (F_1, F_2, F_4) , 0.25; (F_1, F_3, F_4) , 0.24; (F_2, F_3, F_4) , 0.26; (F_1, F_2, F_3, F_4) , 0.29.

VI. QUESTION (16 POINTS)

The following table shows the midterm and final exam grades obtained for students in a database course.

| Midterm Exam | Final Exam |
|--------------|------------|
| 72 | 84 |
| 50 | 63 |
| 81 | 77 |
| 74 | 78 |
| 94 | 90 |
| 86 | 75 |
| 59 | 49 |
| 83 | 79 |
| 65 | 77 |
| 33 | 52 |
| 88 | 74 |
| 81 | 90 |

- Use the method of least squares to find an equation for the prediction of a student's final exam grade based on the student's midterm grade in the course.
- Predict the final exam grade of a student who received an 86 on the midterm exam.