

Statistics 202  
Data Mining & Analysis  
Assignment #5  
Due Wednesday, December 9

Prof. J. Taylor

IDM REFERS TO THE COURSE TEXTBOOK. PROVIDE COPIES OF YOUR CODE  
IN THE ASSIGNMENT.

Q. 1) (*IDM*, 6.10.1)

Q. 2) (*IDM*, 6.10.2)

Q. 3) (*IDM*, 6.10.11) You can use the figure in this file rather than redrawing it.

Q. 4) (*IDM*, 6.10.15, a & b only)

Q. 5) (*IDM*, 7.8.10)

Q. 6) (*IDM*, 10.7.8)

Q. 7) (*IDM*, 10.7.9)

Q. 8) Consider the `Income` data set in the `arules` package. Use these commands to load the data

```
> package(arules)
```

```
> data(Income)
```

- (a) Use the `arules` function to find rules with support at least 30% and confidence of at least 90%. Print the first 10 rules sorted by confidence.
- (b) Repeat (a), but force `arules` to only include rules where `dual incomes` and `language in home` appear only on the left hand side, i.e. in the antecedent.
- (c) Repeat (a), decreasing the confidence level to 60%, thus finding rules in which only income is allowed to be in the consequent.
- (d) Repeat (a), decreasing the support threshold to 20% and the confidence threshold to 50%, thus finding rules in which income greater than 40,000\$ is the only allowed variable in the consequent.

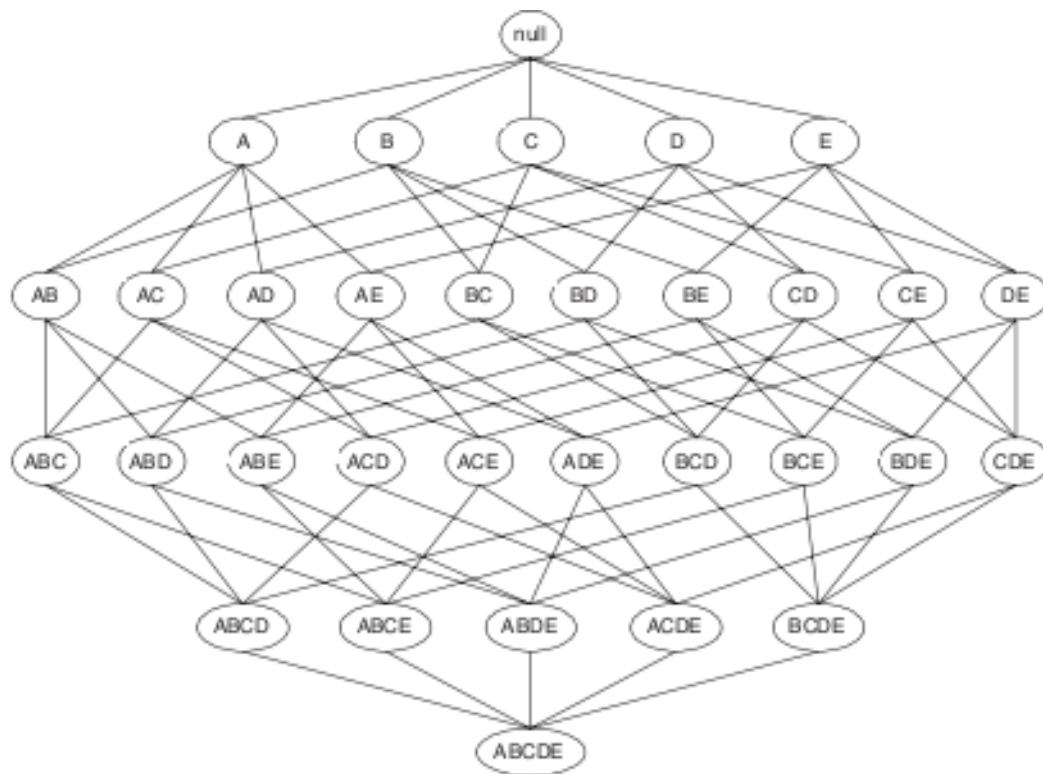


Figure 1: Lattice for 6.10.11