UCLA CS 145 Homework #1

Kevin Shih 504946083

1. Linear Regression

1.1



1.2 (Program’s outputs are showed below)

(a)

They are different. Closed form solution is using derivative to find Beta that minimize the cost function. Batch and Stochastic gradient descent are methods that minimize the cost function by moving down in the steepest direction. The results of gradient descent methods are nearly impossible to be exactly same (could be very close) as the closed form solution. Since we have to move Beta with a specific step size in every loop, we might either pass or not yet reach the optimized point every time.

(b)

After applying Z-score, beta is changed and the MSE value is slightly bigger. I think this result tells that the features in the dataset might not need to be normalized and maybe those features with larger scale should just weight more. Before applying Z-score to any dataset, we should think if this dataset really need a normalization, and just try. It is no guarantee that applying normalization can make the prediction better.

|  |  |  |
| --- | --- | --- |
| Closed  Form | /Users/kevinshih/Desktop/Screen Shot 2017-10-16 at 1.56.04 PM.png | MSE:  4.396098 |
| Batch Gradient | /Users/kevinshih/Desktop/Screen Shot 2017-10-16 at 2.12.01 PM.png | MSE:  (slightly different)  4.396098 |
| Stochastic Gradient | /Users/kevinshih/Desktop/Screen Shot 2017-10-16 at 2.16.02 PM.png | MSE:  4.397317 |

|  |  |  |
| --- | --- | --- |
| Closed  Form  With  Norm | ../../../../Desktop/Screen%20Shot%202017-10-16%20at%202.31.46%20P | MSE:  4.404546 |
| Batch  Gradient  With  Norm | ../../../../Desktop/Screen%20Shot%202017-10-16%20at%202.33.57%20P | MSE:  (slightly different)  4.404546 |
| Stochastic Gradient  With  Norm | ../../../../Desktop/Screen%20Shot%202017-10-16%20at%202.35.17%20P | MSE:  4.4041222 |

1. Logistic Regression and Model Selection

,

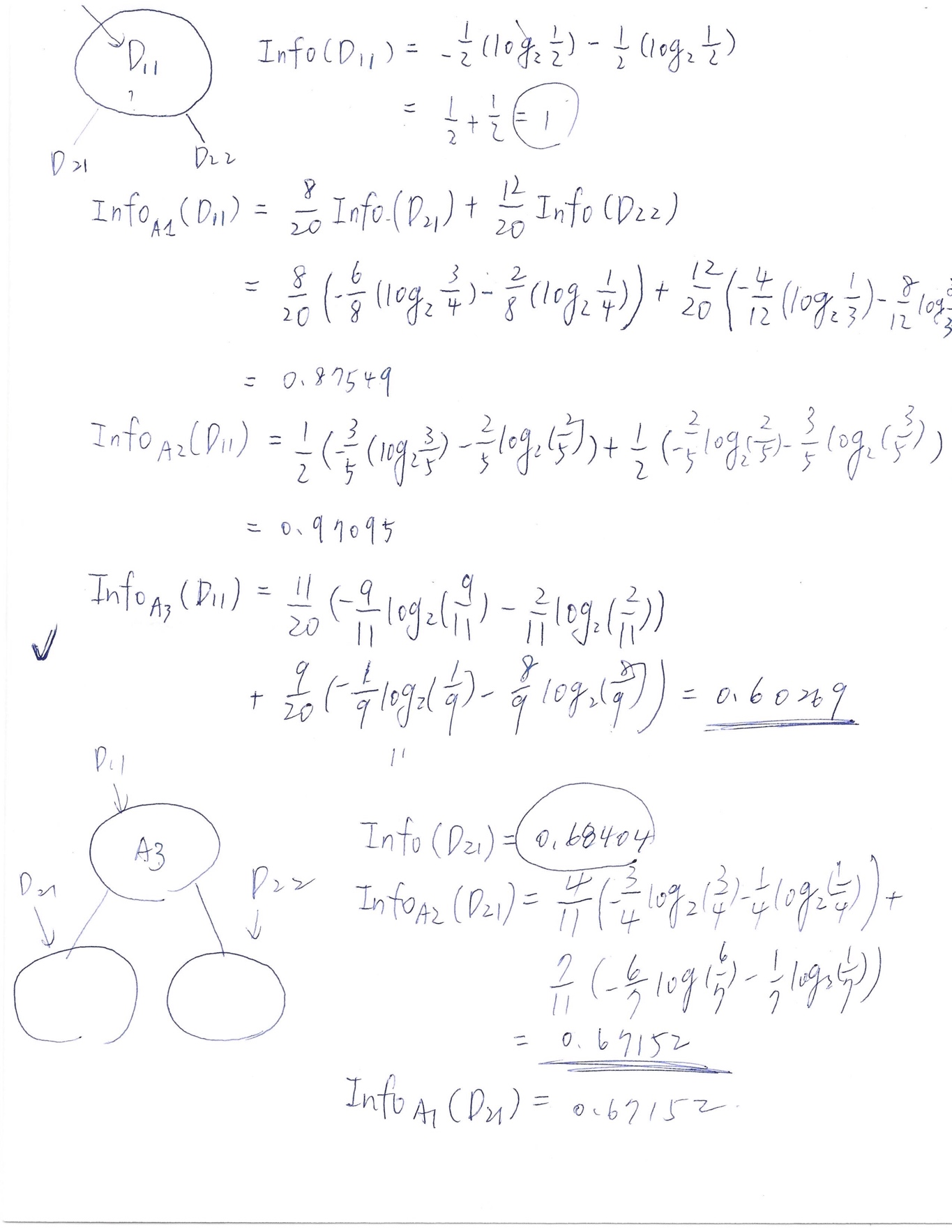
(a)

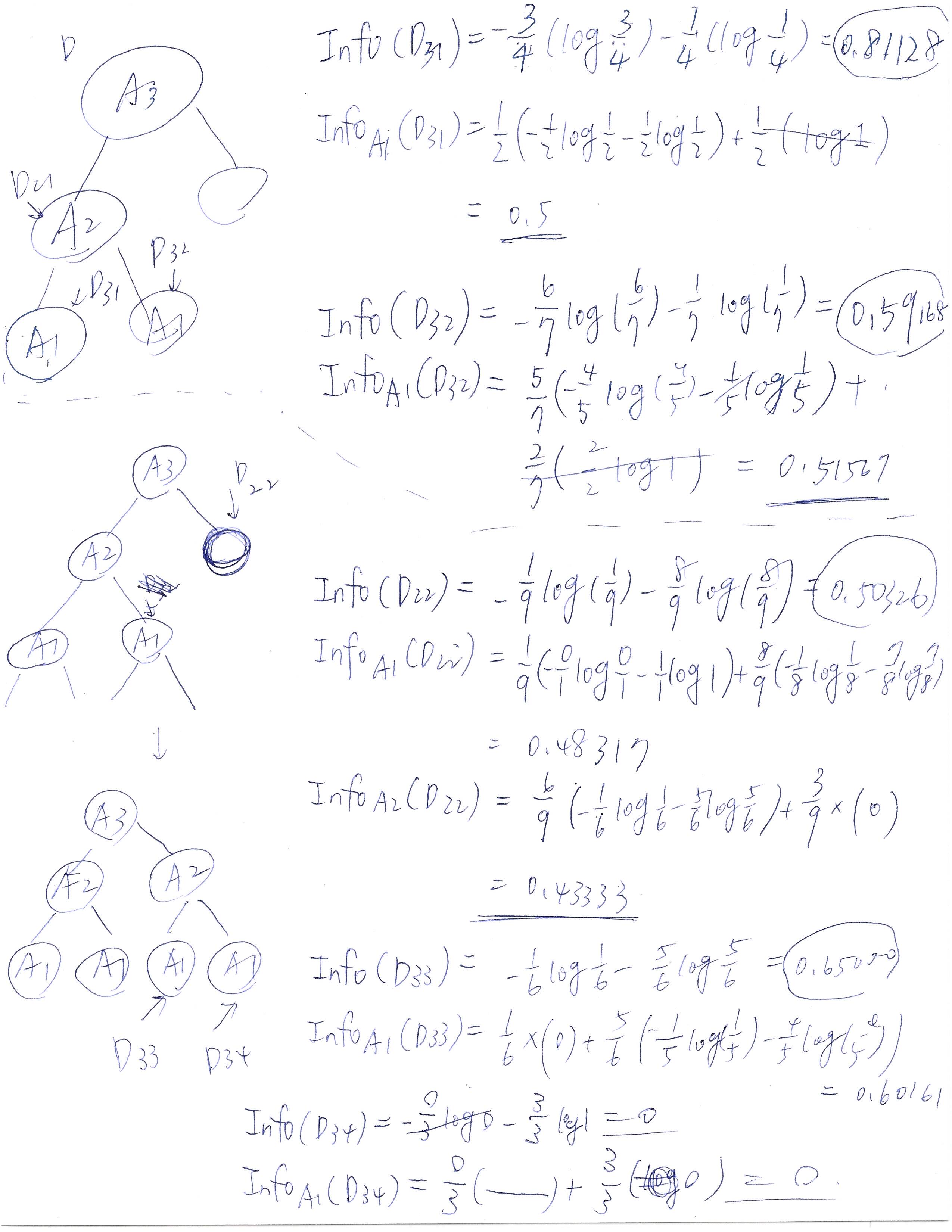
(b)

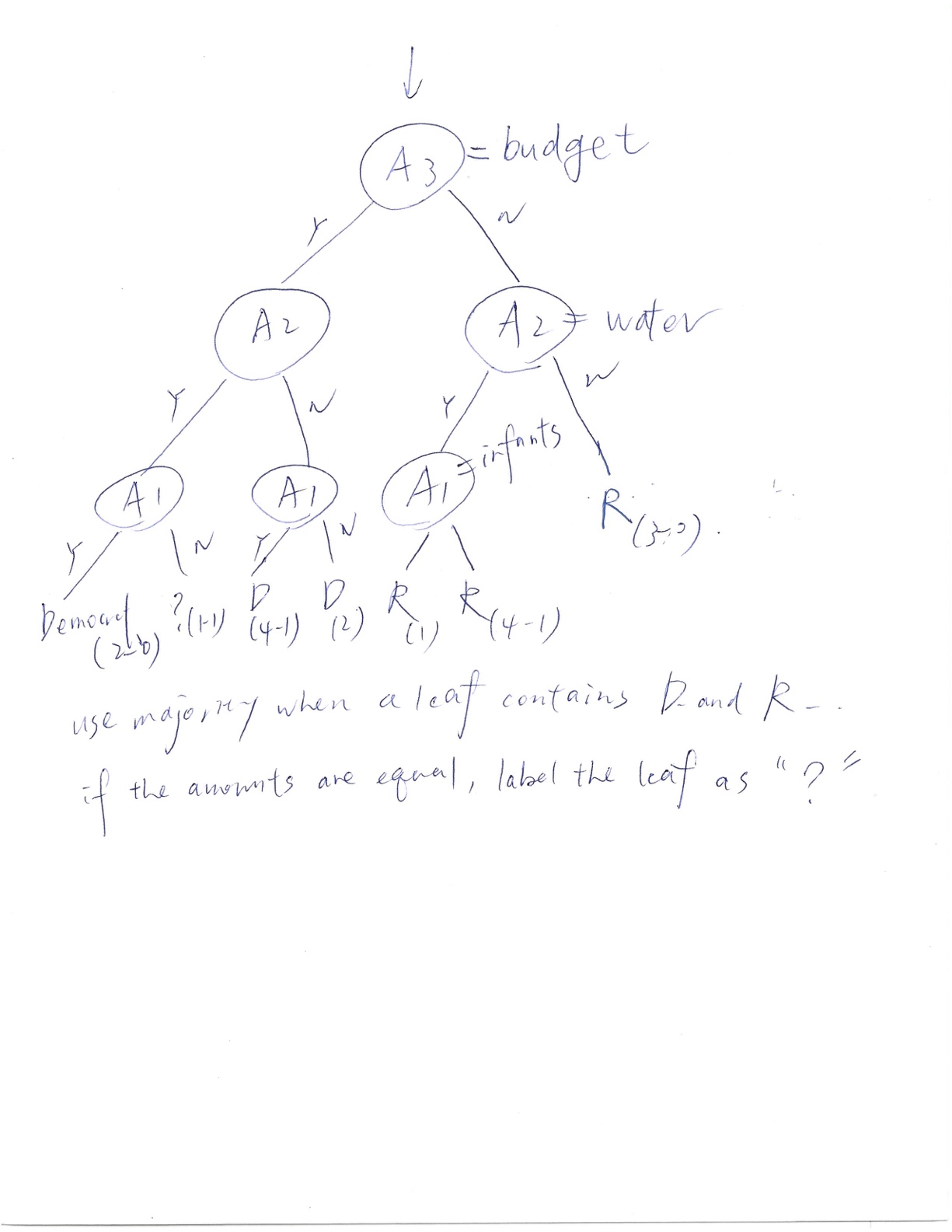
(c)

3. Decision Tree

3.1







(b)

For house-votes-84, I pick the information gain method. Although there are “y, n, ?” in attributes, “?”s are defined as missing values, it would be ok if it makes the measure biased.(since the values already lost in the first place)

For tic-tac-toe, information gain measure might be biased since almost every attribute has 3 values, so use the gain ratio measure instead.