Assignment 4[10].

Consider the following situation.

	Labs	Assignments	Final	Project	Decision
1	All	Poor	Good	Average	Pass
2	Some	Good	Good	Good	Pass
3	Some	Poor	Poor	Good	Fail
4	All	Good	Good	Good	Pass
5	Some	Poor	Good	Average	Pass
6	Some	Poor	Poor	Good	Fail
7	Some	Good	Poor	Good	Pass
8	Some	Good	Poor	Average	Fail
9	All	Poor	Poor	Poor	Pass

Draw a decision tree by selecting the attribute with the highest gain ratio at each step. Show all your work.

Entropy (D) = $-\frac{6}{9} \times \log_2(\frac{6}{9}) - \frac{3}{9} \times \log_2(\frac{3}{9}) = 0.918$

(6/9: # of passes; 3/9: # of fails)

SPLIT #1

A. Labs

LD_1: {1, 4, 9} "All"

LD_2: {2, 3, 5, 6, 7, 8} "Some"

entropy LD_1: $-\frac{3}{3} \times \log(\frac{3}{3}) = 0$

entropy LD_2: $-\frac{3}{6} \times \log(\frac{3}{6}) - \frac{3}{6} \times \log(\frac{3}{6}) = 1$

entropy L: $(\frac{3}{9} \times 0) + (\frac{6}{9} \times 1) = 0.67$

gain: $0.918 - 0.67 = 0.248$

gain ratio: $0.248 / [(-\frac{3}{9} \times \log_2(\frac{3}{9}) - \frac{6}{9} \times \log_2(\frac{6}{9}))] = 0.27$

B. Assignments

AD_1: {2, 4, 7, 8} "Good"

AD_2: {1, 3, 5, 6, 9} "Poor"

entropy AD_1: $-\frac{3}{4} \times \log(\frac{3}{4}) - \frac{1}{4} \times \log(\frac{1}{4}) = 0.811$

entropy AD_2: $-\frac{3}{5} \times \log(\frac{3}{5}) - \frac{2}{5} \times \log(\frac{2}{5}) = 0.97$

entropy A: $(\frac{4}{9} \times 0.811) + (\frac{5}{9} \times 0.97) = 0.8993$

gain: $0.918 - 0.8993 = 0.0186$

gain ratio: $0.0186 / ((-\frac{4}{9} \times \log_2(\frac{4}{9}) - \frac{5}{9} \times \log_2(\frac{5}{9})) = 0.018$

C. Final

FD_1: {1, 2, 4, 5} "Good"

FD_2: {3, 6, 7, 8, 9} "Poor"

entropy FD_1: $\frac{4}{4} \times \log\frac{4}{4} - \frac{0}{4} \times \log\frac{0}{4} = 0$

entropy FD_2: $\frac{2}{5} \times \log(\frac{3}{5}) - \frac{3}{5} \times \log(\frac{3}{5}) = 0.97$

entropy F: $(\frac{4}{9} \times 0) + (\frac{5}{9} \times 0.97) = 0.53$

gain: $0.918 - 0.53 = 0.3791$

gain ratio: $0.3791 / (-\frac{4}{9} \times \log(\frac{4}{9}) - \frac{5}{9} \times \log(\frac{5}{9})) = 0.38$ ← SPLIT

D. Project

PD_1: {2, 3, 4, 6, 7} "Good"

PD_2: {1, 5, 8} "Average"

PD_3: {9} "Poor"

entropy PD_1: $-\frac{3}{5} \times \log(\frac{3}{5}) - \frac{2}{5} \times \log(\frac{2}{5}) = 0.97$

entropy PD_2: $-\frac{2}{3} \times \log(\frac{2}{3}) - \frac{1}{3} \times \log(\frac{1}{3}) = 0.918$

entropy PD_3: $-\frac{1}{1} \times \log(\frac{1}{1}) - \frac{0}{1} \times \log(\frac{0}{1}) = 0$

entropy P: $(\frac{5}{9} \times 0.97) + (\frac{3}{9} \times 0.918) + (\frac{1}{9} \times 0) = 0.944$

gain: $0.918 - 0.944 = 0.074$

gain ratio: $0.074 / (-\frac{5}{9} \times \log_2(\frac{5}{9}) - \frac{3}{9} \times \log_2\frac{3}{9} - \frac{1}{9} \times \log_2(\frac{1}{9})) = 0.07$

Entropy (D) = $-\frac{2}{5} \times \log_2(\frac{2}{5}) - \frac{3}{5} \times \log_2(\frac{3}{5}) = 0.9709$

(3/5: # of fails)

Entropy (D) = - 2/5 (# of passes) × $\log_2$(2/5) - 3/5 (# of fails) × $\log_2$(3/5) = 0.9709

SPLIT #2

A. Labs

LD_1: {3, 6, 7, 8} "All"

LD_2: {9} "Some"

entropy LD_1: -1/4 × log(1/4) - 3/4 × log(3/4) = 0.811

entropy LD_2: - 1/1 × log(1/1) = 0

entropy L: (4/5 × 0.811) + (1/5 × 0) = 0.6488

gain: 0.970 - 0.6488 = 0.3221

gain ratio: 0.3221 / (-4/5 × $\log_2$(4/5) - (1/5) × log(1/5)) = 0.44 ↑ SPLIT

B. Assignments

AD_1: {7, 8} "Good"

AD_2: {3, 6, 9} "Poor"

entropy AD_1: -1/2 × log(1/2) - 1/2 × log(1/2) = 1

entropy AD_2: - 1/3 × log(1/3) - 2/3 × log(2/3) = 0.918

entropy A: (2/5 × 1) + (3/5 × 0.918) = 0.95

gain: 0.970 - 0.95 = 0.02

gain ratio: 0.02 / (-2/5 × $\log_2$(2/5) - 3/5 × $\log_2$(3/5)) = 0.02

C. Project

PD_1: {3, 6, 7} "Good"

PD_2: {8} "Average"

PD_3: {9} "Poor"

entropy PD_1: -1/3 × log(1/3) - 2/3 × log 2/3 = 0.918

entropy PD_2: - 0/1 × log(0/1) - 1/1 × log(1/1) = 0

entropy PD_3 = -1/1 × log(1/1) - 0/1 × log(0/1) = 0

entropy P: (3/5 × 0.918) + (1/5 × 0) + (1/5 × 0) = 0.5508

gain: 0.97 - 0.5508 = 0.4192

gain ratio: 0.4192 / (-3/5 × $\log_2$(3/5) - 0 - 0) = 0.43

Entropy (D) = - 1/4 (# of passes) × $\log_2$(1/4) - 3/4 (# of fails) × $\log_2$(3/4) = 0.8112

SPLIT #3

A. Assignments

AD_1: {7, 8} "Good"

AD_2: {3, 6} "Poor"

entropy AD_1: -1/2 × log(1/2) - 1/2 × log(1/2) = 1

entropy AD_2: - 0/2 × log(0/2) - 2/2 × log(2/2) = 0

entropy A: (2/4 × 1) + (2/4 × 0) = 0.5

gain: 0.8112 - 0.5 = 0.3112

gain ratio: 0.3112 / (-2/4 × $\log_2$(2/4) + 0) = 0.622 ↑ SPLIT

B. Project

PD_1: {3, 6, 7} "Good"

PD_2: {8} "Average"

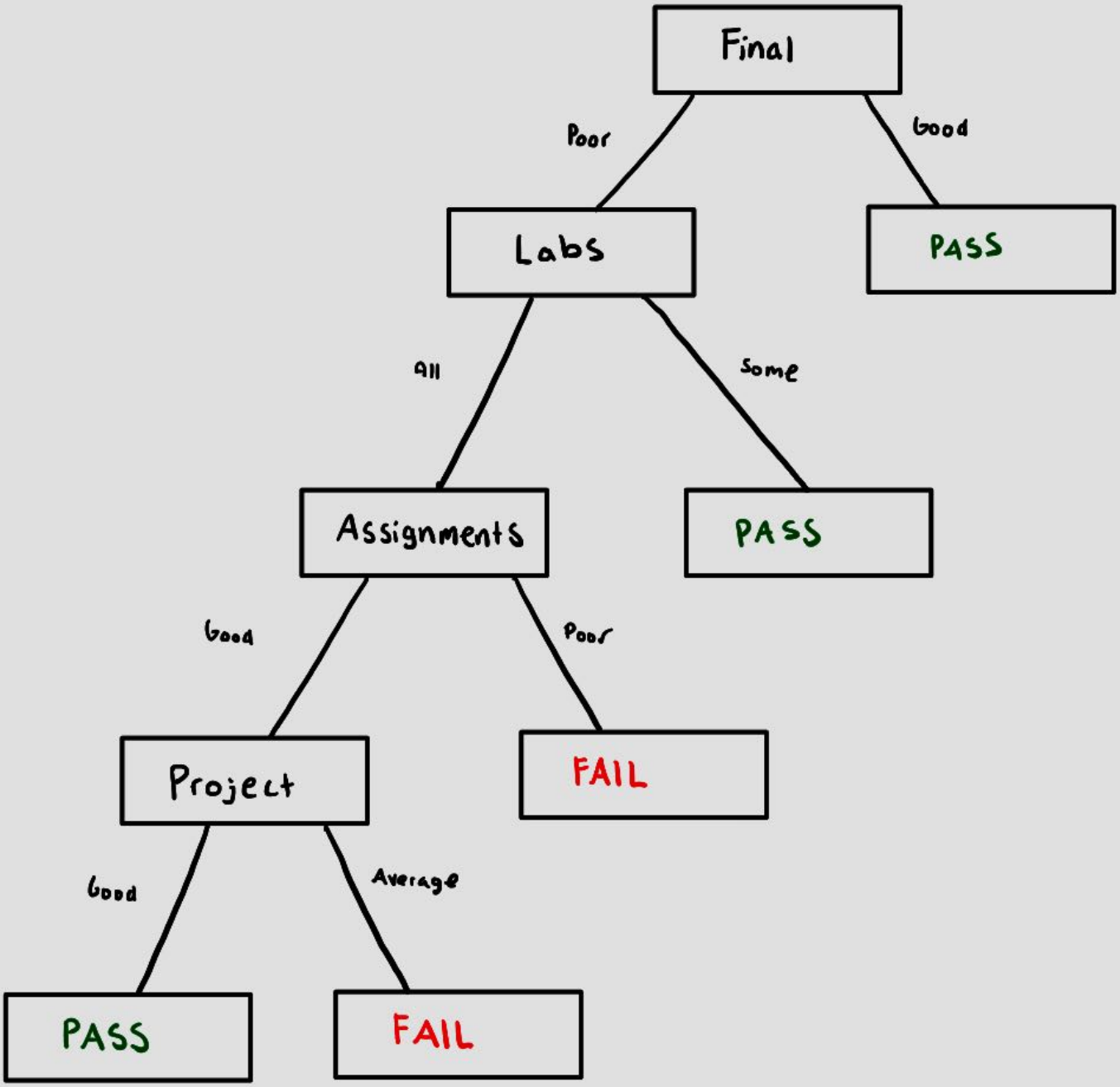
entropy PD_1: -1/3 × log(1/3) - 2/3 × log 2/3 = 0.918

entropy PD_2: - 0/1 × log(0/1) - 1/1 × log(1/1) = 0

entropy P: (3/4 × 0.918) + (1/4 × 0) = 0.6885

gain: 0.8112 - 0.6885 = 0.1227

gain ratio: 0.1227 / (-3/4 × $\log_2$(3/4) - 0) = 0.038


Final
Poor
Good
Labs
PASS
All
Some
Assignments
PASS
Good
Poor
Project
FAIL
Good
Average
PASS
FAIL