

ASS3 Decision tree

⇒ Data set entropy

$$E(S) = \frac{-8}{14} \log_2 \frac{8}{14} - \frac{6}{14} \log_2 \frac{6}{14} = 0.985$$

* Early registration (S) (4)

$$E(S_0) = \frac{-4}{8} \log_2 \frac{4}{8} - \frac{4}{8} \log_2 \frac{4}{8} = 1$$

$$E(S_1) = \frac{-4}{6} \log_2 \frac{4}{6} - \frac{2}{6} \log_2 \frac{2}{6} = 0.918$$

$$IG(S, \text{early registration}) = 0.985 - \frac{8}{14} - \frac{6}{14} \rightarrow 0.918$$

$$= 0.0201$$

* Finished homework 2

$$E(S_0) = \frac{-3}{7} \log_2 \frac{3}{7} - \frac{4}{7} \log_2 \frac{4}{7} = 0.985$$

$$E(S_1) = \frac{-5}{7} \log_2 \frac{5}{7} - \frac{2}{7} \log_2 \frac{2}{7} = 0.863$$

$$IG(S, \text{finished homework 2}) = 0.985 - \frac{7}{14} 0.985 - \frac{7}{14} 0.863 \\ = 0.061$$

* Senior

$$E(S_0) = -\frac{3}{6} \log_2 \frac{3}{6} - \frac{3}{6} \log_2 \frac{3}{6} = 1$$

$$E(S_1) = -\frac{5}{8} \log_2 \frac{5}{8} - \frac{3}{8} \log_2 \frac{3}{8} = 0.954$$

$$IG(S, \text{Senior}) = 0.985 - \frac{6}{14} - \frac{8}{14} 0.954 = 0.0113$$

* Likes Coffee

$$E(S_0) = -\frac{5}{10} \log_2 \frac{5}{10} - \frac{5}{10} \log_2 \frac{5}{10} = 1$$

$$E(S_1) = -\frac{3}{4} \log_2 \frac{3}{4} - \frac{1}{4} \log_2 \frac{1}{4} = 0.811$$

$$IG(S, \text{Likes Coffee}) = 0.985 - \frac{10}{14} - \frac{4}{14} 0.811 = 0.039$$

* liked The last H.W

$$E(S_0) = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0.971$$

$$E(S_1) = -\frac{5}{8} \log_2 \frac{5}{8} - \frac{4}{8} \log_2 \frac{4}{8} = 0.991$$

$$IG(S, \text{liked The last H.W}) = 0.985 - \frac{5}{14} 0.971 - \frac{9}{14} 0.991 \\ = 0.00114$$

→ Maximum IG → Finished homework? (Tree Root)

* Subtree for left branch finished homework = 0
 $S = 0.985$

→ early registration

$$E(S_0) = -\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} = 1$$

$$E(S_1) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0.918$$

$$IG = 0.985 - \frac{4}{7} - \frac{3}{7} 0.918 = 0.0201$$

→ Senior

$$E(S_0) = -\frac{1}{4} \log_2 \frac{1}{4} - \frac{3}{4} \log_2 \frac{3}{4} = 0.811$$

$$E(S_1) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0.918$$

$$IG = 0.985 - \frac{4}{7} 0.811 - \frac{3}{7} 0.918 = 0.128$$

→ Likes coffee

$$E(S_0) = -\frac{1}{5} \log_2 \frac{1}{5} - \frac{4}{5} \log_2 \frac{4}{5} = 0.722$$

$$E(S_1) = 0$$

$$IG = 0.985 - \frac{5}{7} 0.722 = 0.469$$

→ Likes HW

$$E(S_0) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0.918$$

$$E(S_1) = 1$$

$$IG = 0.985 - \frac{4}{7} - \frac{3}{7} 0.918 = 0.0201$$

Maximum IG is Likes coffee

finished homework = 1 $S = 0.863$

* early registration

$$E(S_0) = 1$$

$$E(S_1) = 0$$

$$IG = 0.863 - \frac{4}{7} = 0.787$$

* Senior

$$E(S_0) = 0$$

$$E(S_1) = \frac{-3}{5} \log \frac{3}{5} - \frac{2}{5} \log \frac{2}{5} = 0.971$$

$$IG = 0.863 - \frac{5}{7} 0.971 = 0.169$$

* likes coffee

$$E(S_0) = \frac{-4}{5} \log \frac{4}{5} - \frac{1}{5} \log \frac{1}{5} = 0.722$$

$$E(S_1) = 1$$

$$IG = 0.863 - 0.722 \left(\frac{5}{7} \right) - \frac{2}{7} = 0.062$$

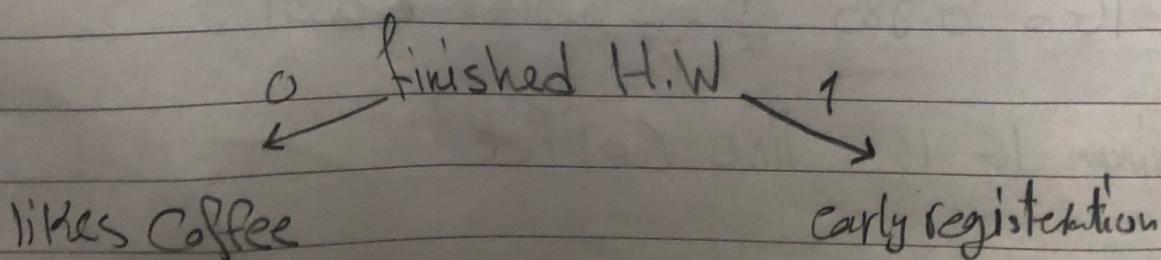
* liked last H.W

$$E(S_0) = 0$$

$$E(S_1) = \frac{-3}{5} \log \frac{3}{5} - \frac{2}{5} \log \frac{2}{5} = 0.971$$

$$IG = 0.863 - \frac{5}{7} 0.971 = 0.169$$

Maximum IG is early registration



→ Finished H.W = 0 likes Coffee = 0 $S = 0.722$

* early registration

$$E(S_0) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0.918$$

$$E(S_1) = 0$$

$$IG = 0.722 - \frac{3}{5} 0.918 = 0.1712$$

* Senior

$$E(S_0) = 0$$

$$E(S_1) = 1$$

$$IG = 0.722 - \frac{2}{5} = 0.322$$

* like H.W

$$E(S_0) = 0$$

$$E(S_1) = -\frac{2}{3} \log_2 \frac{2}{3} - \frac{1}{3} \log_2 \frac{1}{3} = 0.918$$

$$IG = 0.722 - \frac{3}{5} 0.918 = 0.1712$$

Maximum IG is senior

→ Finished H.W = 0 Likes Coffee = 1 $S = 0$
pure Class $A = 1$

→ Finished H.W = 1 early registration = 0

* Senior

$$E(S_0) = 0$$

$$E(S_1) = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0.918$$

$$IG = 1 - \frac{3}{4} 0.918 = 0.315$$

* likes Coffee

$$E(S_0) = 1$$

$$E(S_1) = 1$$

$$IG = 1 - \frac{2}{4} - \frac{2}{4} = 0$$

* likes H.W

$$E(S_0) = 0$$

$$E(S_1) = 0.918$$

$$IG = 0.315$$

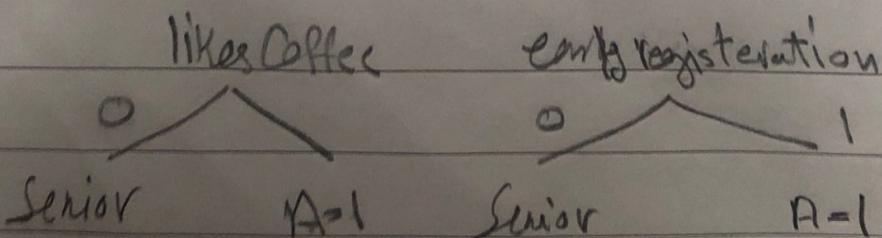
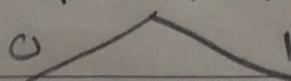
Maximum at: Senior or likes H.W

I'll choose Senior

→ Finished H.W = 1 early registrations 1

Pure Class A = 1

finished HW



→ finished H.W = 0 Likes Coffee = 0 Senior = 0

pure Class A = 0

→ Finished H.W = 0 Likes Coffee = 0 Senior = 1

* Likes Coffee

$$E(S_0) = 1$$

$$E(S_1) = 0$$

$$IG = 1 - \frac{1}{2} = \frac{1}{2}$$

* Liked H.W

$$E(S_0) = 0$$

$$E(S_1) = 0$$

$$IG = 1$$

Maximum IG is liked H.W

→ finished H.W = 0 Likes Coffee = 0 Senior = 1 Likes H.W = 0

pure Class A = 0

→ finished H.W = 0 Likes Coffee = 0 Senior = 1 Likes H.W = 1

pure Class A = 1

DATE
finished HW=1 early registration=0 Senior=0
pure Class A=1

finished H.W=1 early registration=0 Senior=1
Likes Coffee

$$E(S_0) = 1$$

$$E(S_1) = 0$$

$$IG = 0.918 - \frac{2}{3} = 0.251$$

Likes H.W

$$E(S_0) = 0$$

$$E(S_1) = 0$$

$$TG = 0.918$$

Maximum IG Likes H.W

finished H.W=1 early registration=0 Senior=1
Likes H.W=1

Pure Class A=0

finished H.W=1, early registration=0, Senior=1
Likes H.W=0

Pure Class A=1

Tree

finished H.W

