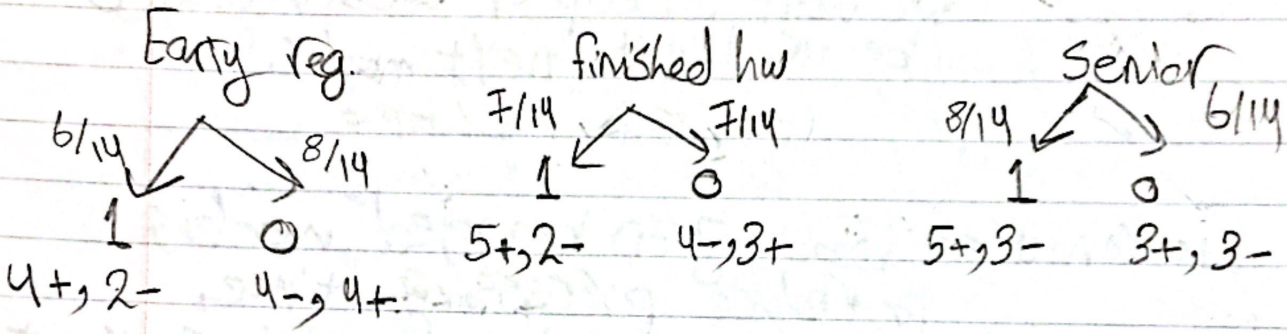
→ rules:

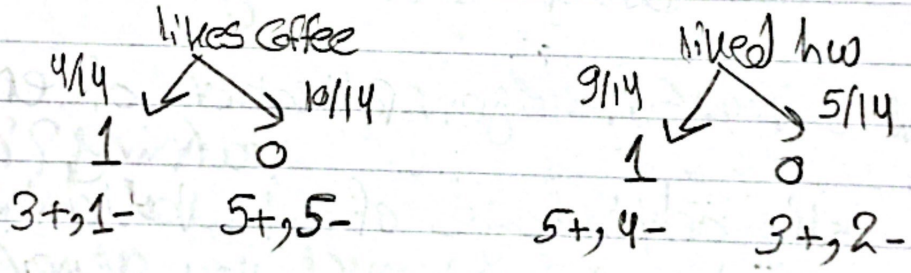
$$E(\text{DataSet}) = -[P_+ \log P_+ + P_- \log P_-]$$

$$IG = E(D) - \sum \frac{S_v}{S} E(S_v)$$

→ level 1)

$$E(D) = -\left[\frac{8}{14}\log\frac{8}{14} + \frac{6}{14}\log\frac{6}{14}\right] = 0.985$$





$$IG(\text{Early reg.}) = 0.985 - \left[\frac{6}{14} * -\left(\frac{4}{6}\log\frac{4}{6} + \frac{2}{6}\log\frac{2}{6}\right) + \frac{8}{14} * (1)\right] = 0.02$$

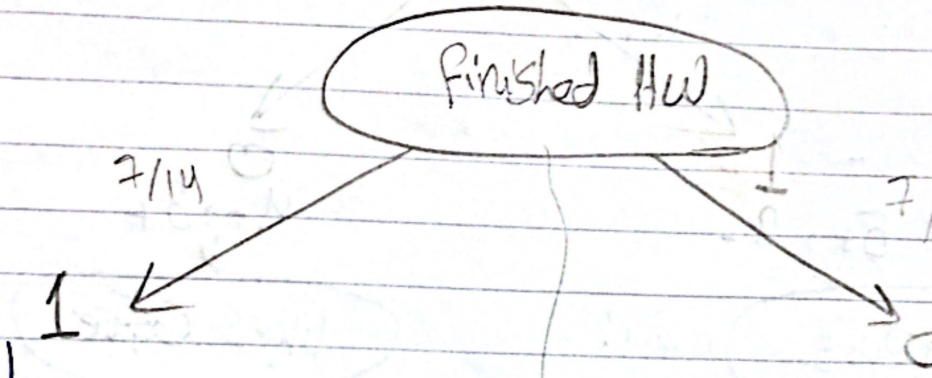
$$IG(\text{finished hw}) = 0.985 - \left[\frac{7}{14} * -\left(\frac{5}{7}\log\frac{5}{7} + \frac{2}{7}\log\frac{2}{7}\right) + \frac{7}{14} * -\left(\frac{4}{7}\log\frac{4}{7} + \frac{3}{7}\log\frac{3}{7}\right)\right] =$$
$$0.985 - [0.43 + 0.49] = 0.062$$

$$IG(\text{Senior}) = 0.985 - \left[\frac{8}{14} * -\left(\frac{5}{8}\log\frac{5}{8} + \frac{3}{8}\log\frac{3}{8}\right) + \frac{6}{14} * 1\right] = 0.011$$

$$IG(\text{liked hw}) = 0.985 - \left[\frac{9}{14} * -\left(\frac{5}{9}\log\frac{5}{9} + \frac{4}{9}\log\frac{4}{9}\right) + \frac{5}{14} * -\left(\frac{3}{5}\log\frac{3}{5} + \frac{2}{5}\log\frac{2}{5}\right)\right]$$
$$= 0.985 - [0.637 + 0.346] = 0.002$$

$$IG(\text{likes Coffee}) = 0.985 - \left[\frac{4}{14} * -\left(\frac{3}{4}\log\frac{3}{4} + \frac{1}{4}\log\frac{1}{4}\right) + \frac{10}{14}\right] = 0.039$$

- Since "finished hw" had the largest IG at level 1, we consider it our root node.
It divides our dataset into two subsets.



→ level 2)

$E(S_1) = -\left(\frac{5}{7}\log\frac{5}{7} + \frac{2}{7}\log\frac{2}{7}\right) = 0.863$

$E(S_2) = -\left(\frac{4}{7}\log\frac{4}{7} + \frac{3}{7}\log\frac{3}{7}\right) = 0.985$

IG(early_reg) =
$0.863 - \left[\frac{3}{7}*0 + \frac{4}{7}*1\right] = 0.292$

IG(early_reg) =
$0.985 - \left[\frac{4}{7}*+(1) + \frac{3}{7}*-\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right)\right]$
$= 0.02$

IG(senior) =
$0.863 - \left[\frac{5}{7}*-\left(\frac{3}{5}\log\frac{3}{5} + \frac{2}{5}\log\frac{2}{5}\right) + \frac{2}{7}*0\right] = 0.169$

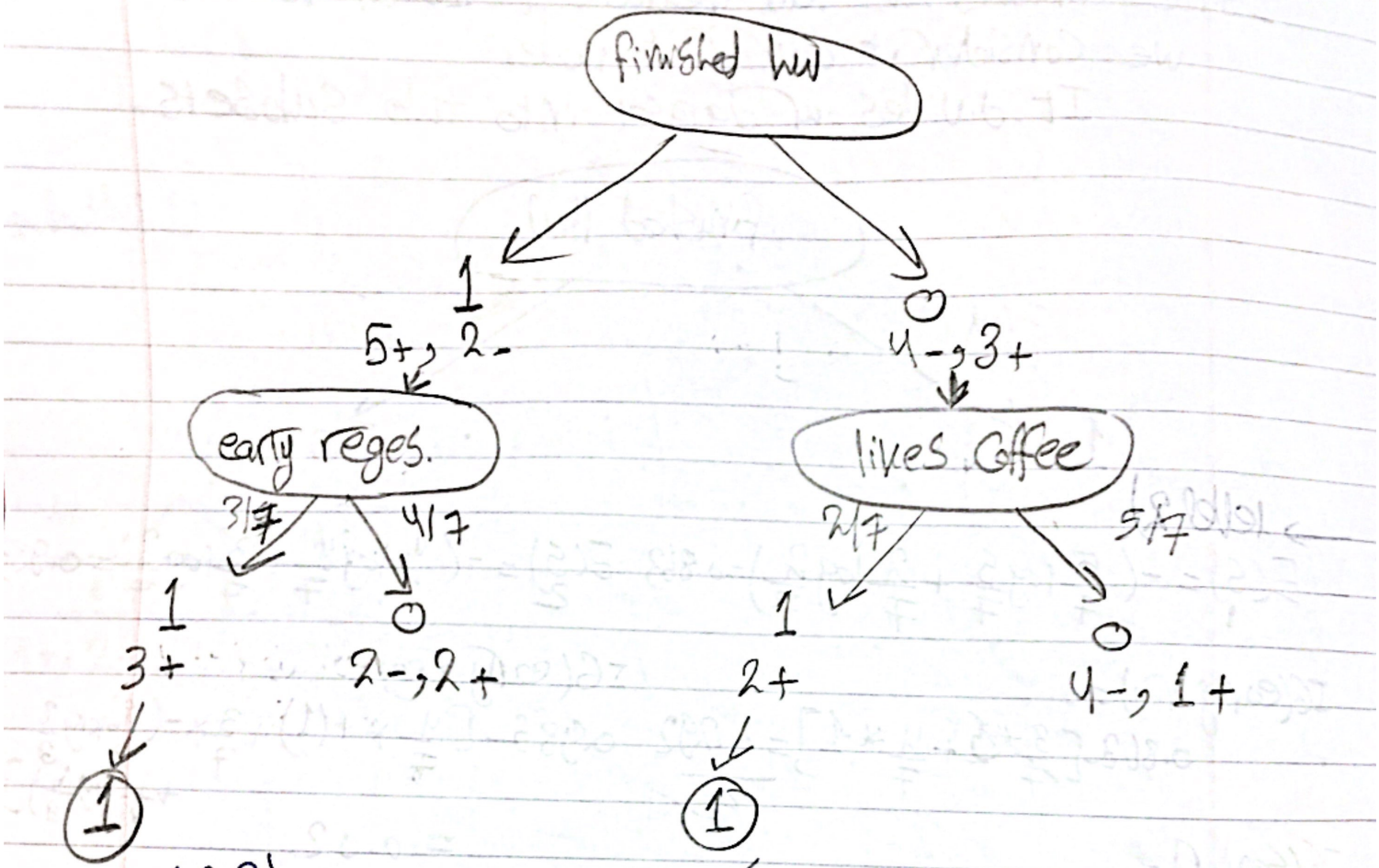
IG(senior) =
$0.985 - \left[\frac{3}{7}*-\left(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}\right) + \frac{4}{7}*-\left(\frac{3}{4}\log\frac{3}{4} + \frac{1}{4}\log\frac{1}{4}\right)\right]$
$= 0.985 - [0.394 + 0.464] = 0.127$

IG(likes Coffee) =
$0.863 - \left[\frac{2}{7}*1 + \frac{5}{7}*-\left(\frac{4}{5}\log\frac{4}{5} + \frac{1}{5}\log\frac{1}{5}\right)\right] = 0.062$

IG(likes Coffee) =
$0.985 - \left[\frac{5}{7}*-\left(\frac{4}{5}\log\frac{4}{5} + \frac{1}{5}\log\frac{1}{5}\right) + \frac{2}{7}*0\right] = 0.469$

IG(liked hw) =
$0.863 - \left[\frac{5}{7}*-\left(\frac{3}{5}\log\frac{3}{5} + \frac{2}{5}\log\frac{2}{5}\right) + \frac{2}{7}*0\right] = 0.169$

IG(liked hw) =
$0.985 - \left[\frac{3}{7}*-\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right) + \frac{4}{7}*+(1)\right]$
=
$0.985 - \left[0.39 + \frac{4}{7}\right] = 0.02$



→ level 3)

$E(S_3) = 1$

IG(senior) =

$1 - \left[\frac{3}{4} * -\left(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}\right) + \frac{1}{4} * (0)\right] = 0.311$

IG(lives Coffee) $= 1 - \left[\frac{2}{4} * (1) + \frac{2}{4} * 1\right] = 0$

IG(liked hw) $= 1 - \left[\frac{3}{4} * -\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right) + \frac{1}{4} * 0\right] = 0.311$

$E(S_4) = -\left(\frac{4}{5}\log\frac{4}{5} + \frac{1}{5}\log\frac{1}{5}\right) = 0.723$

IG(early) =

$0.72 - \left[\frac{3}{5} * -\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right) + \frac{2}{5} * (0)\right] = 0.169$

IG(senior) $= 0.72 - \left[\frac{2}{5} * (1) + \frac{3}{5} * (0)\right] = 0.32$

IG(liked hw) =

$0.72 - \left[\frac{2}{5} * [0] + \frac{3}{5} * -\left(\frac{2}{3}\log\frac{2}{3} + \frac{1}{3}\log\frac{1}{3}\right)\right] = 0.169$

level 4)

$$E(S) = -\left(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}\right)$$

$$= 0.918$$

IG(Senior) =

$$0.918 - \left[\frac{2}{3} * (0) + \frac{1}{3} * (1)\right]$$

$$= 0.584$$

IG(likes coffee) =

$$0.918 - \left[\frac{1}{3} * (0) + \frac{2}{3} * 1\right] = 0.25$$

$E(S) = 1$

IG(early reg) = 1st level

$$1 - [1 * 1 + 0] = 0$$

IG(liked last) =

$$1 - \left[\frac{1}{2} * (0) + \frac{1}{2} * 0\right] = 0$$

→ final tree