Module 8 Self Check

David Bishop

October 2024

1 Calculations

1.1 Entropy

1.1.1 Shape

$E(S) = -\sum_i p_i \log_2(p_i)$
i, can be round or square
$p_{round} = \frac{7}{15}$
$p_{square} = \frac{8}{15}$
$-((\frac{7}{15} \log_2 \frac{7}{15}) + (\frac{8}{15} \log_2 \frac{8}{15}))$
$E(S) \approx 0.997$
$E(S_{round}) = -((\frac{4}{7} log_2 \frac{4}{7}) + (\frac{3}{7} log_2 \frac{3}{7})) \approx 0.985$
$E(S_{square}) = -((\frac{3}{8} log_2 \frac{3}{8}) + (\frac{5}{8} log_2 \frac{5}{8})) \approx 0.954$

1.1.2 Size

$E(S) = -\sum_i p_i \log_2(p_i)$
i, can be small or large
$p_{small} = \frac{7}{15}$
$p_{large} = \frac{8}{15}$
$-((\frac{7}{15} \log_2 \frac{7}{15}) + (\frac{8}{15} \log_2 \frac{8}{15}))$
$E(S) \approx 0.997$
$E(S_{small}) = -((\frac{1}{7} log_2 \frac{1}{7}) + (\frac{6}{7} log_2 \frac{6}{7})) \approx 0.592$
$E(S_{large}) = -((\frac{2}{8} log_2 \frac{2}{8}) + (\frac{6}{8} log_2 \frac{6}{8})) \approx 0.811$

1.1.3 Color

$E(S) = -\sum_i p_i \log_2(p_i)$
i, can be blue, green, red
$p_{blue} = \frac{3}{15}$
$p_{green} = \frac{6}{15}$
$p_{red} = \frac{6}{15}$
$-((\frac{3}{15} \log_2 \frac{3}{15}) + (\frac{6}{15} \log_2 \frac{6}{15}) + (\frac{6}{15} \log_2 \frac{6}{15}))$
$E(S) \approx 0.464$
$E(S_{blue}) = -((\frac{3}{3} log_2 \frac{3}{3}) + (\frac{0}{3} log_2 \frac{0}{3})) = 0$
$E(S_{green}) = -((\frac{2}{6} log_2 \frac{2}{6}) + (\frac{4}{6} log_2 \frac{4}{6})) \approx 0.918$
$E(S_{red}) = -((\frac{3}{6} log_2 \frac{3}{6}) + (\frac{3}{6} log_2 \frac{3}{6})) \approx 1$

1.1.4 Safe?

$E(S) = -\sum_i p_i \log_2(p_i)$
i, can be yes or no
$p_{yes} = \frac{7}{15}$
$p_{no} = \frac{8}{15}$
$-((\frac{7}{15}\log_2\frac{7}{15}) + (\frac{8}{15}\log_2\frac{8}{15}))$
$E(S) \approx 0.997$

1.2 Information Gain

1.2.1 Shape

$IG(Safe?, Shape) = E(Safe?) - ((\frac{|S_{round}|}{|S|}E(S_{round}) + (\frac{|S_{square}|}{|S|}E(S_{square}))$
$IG = 0.997 - ((\frac{7}{15} * 0.985) + (\frac{8}{15} * 0.954)) \approx 0.0285$

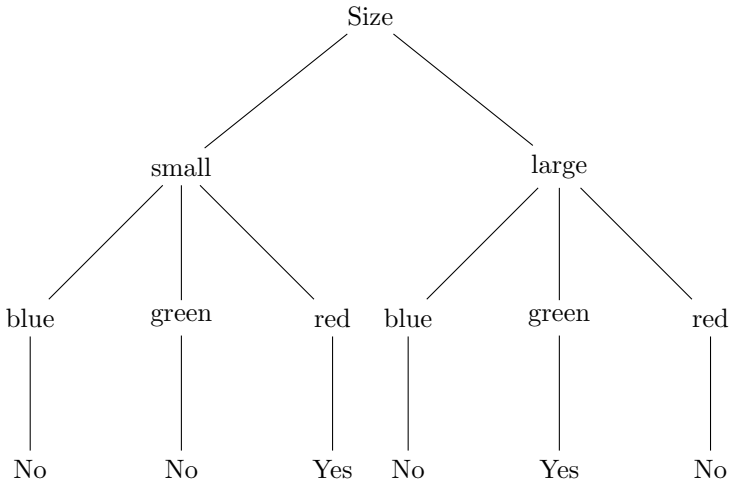
1.2.2 Size

$IG(Safe?, Size) = E(Safe?) - ((\frac{|S_{small}|}{|S|}E(S_{small}) + (\frac{|S_{large}|}{|S|}E(S_{large}))$
$IG = 0.997 - ((\frac{7}{15} * 0.592) + (\frac{8}{15} * 0.811)) \approx 0.288$

1.2.3 Color

$IG(Safe?, Color) = E(Safe?) - ((\frac{|S_{blue}|}{|S|}E(S_{blue}) + (\frac{|S_{green}|}{|S|}E(S_{green}) + (\frac{|S_{red}|}{|S|}E(S_{red}))$
$IG = 0.997 - ((\frac{3}{15} * 0) + (\frac{6}{15} * 0.918) + (\frac{6}{15} * 1)) \approx 0.230$

2 Decision Tree



Since shape has the lowest Information gain and we reach a conclusion at each of the size color combos we do not need to add it to the tree.