1.1

Q1.

Before pruning

Train Acc : $\frac{8}{9} = 88.9\%$

Validation Acc: $\frac{4}{14} = 28.6\%$

After pruning:

Training Acc: $\frac{8}{9} = 88.9\%$

Validation Acc $\frac{10}{14} = 71.4\%$

$\therefore$ The subtree should be pruned.

Q2.

1. $H(D) = -\sum_i p_i \log_2(p_i)$

$P(Yes) = \frac{9}{15} = 0.6$

$P(No) = \frac{6}{15} = 0.4$

$H(D) = -0.6 \log_2(0.6) - 0.4 \log_2(0.4)$

$= 0.442 + 0.529 = 0.971$

A1: age:

$H(D|A_1) = \frac{5}{15} H(Youth) + \frac{5}{15} H(Middle) + \frac{5}{15} H(Elderly)$

$\because H(Youth) = -\frac{2}{5}\log_2(\frac{2}{5}) - \frac{3}{5}\log_2(\frac{3}{5}) = 0.971$

$H(Middle) = -\frac{3}{5}\log_2(\frac{3}{5}) - \frac{2}{5}\log_2(\frac{2}{5}) = 0.971$

$H(Elderly) = -\frac{4}{5}\log_2(\frac{4}{5}) - \frac{1}{5}\log_2(\frac{1}{5}) = 0.722$

$\therefore H(D|A_1) = 0.888$

$\therefore Gain(A_1) = 0.971 - 0.888 = 0.083$

$H(A_1) = -3 \times \frac{5}{15}\log_2(\frac{5}{15}) = 1.585$

$\therefore Gain\ Ratio(A_1) = \frac{0.083}{1.585} = 0.052$

A2: Employed

$H(D|A_2) = \frac{10}{15} H(E=No) + \frac{5}{15} H(E=Yes)$

$\because H(E=No) = -\frac{4}{10}\log_2(\frac{4}{10}) - \frac{6}{10}\log_2(\frac{6}{10}) = 0.971$

$H(E=Yes) = -\frac{5}{5}\log_2(\frac{5}{5}) - \frac{0}{5}\log_2(\frac{0}{5}) = 0$

$\therefore H(D|A_2) = 0.647$

$\therefore Gain(A_2) = 0.971 - 0.647 = 0.324$

$H(A_2) = -\frac{10}{15}\log_2(\frac{10}{15}) - \frac{5}{15}\log_2(\frac{5}{15}) = 0.918$

$\therefore Gain\ Ratio(A_2) = \frac{0.324}{0.918} = 0.353$

A3: House

$H(D|A_3) = \frac{11}{15} H(O=No) + \frac{4}{15} H(O=Yes)$

$\because H(O=No) = -\frac{5}{11}\log_2(\frac{5}{11}) - \frac{6}{11}\log_2(\frac{6}{11}) = 0.994$

$H(O=Yes) = -\frac{4}{4}\log_2(\frac{4}{4}) - \frac{0}{4}\log_2(\frac{0}{4}) = 0$

$\therefore H(D|A_3) = 0.729$

$\therefore Gain(A_3) = 0.971 - 0.729 = 0.242$

$H(A_3) = -\frac{11}{15}\log_2(\frac{11}{15}) - \frac{4}{15}\log_2(\frac{4}{15}) = 0.837$

$\therefore Gain\ Ratio(A_3) = \frac{0.242}{0.837} = 0.289$

A4: Credit Status

$H(D|A_4) = \frac{5}{15} H(A) + \frac{6}{15} H(G) + \frac{4}{15} H(VG)$

$\because H(A) = -\frac{1}{5}\log_2(\frac{1}{5}) - \frac{4}{5}\log_2(\frac{4}{5}) = 0.722$

$H(G) = -\frac{4}{6}\log_2(\frac{4}{6}) - \frac{2}{6}\log_2(\frac{2}{6}) = 0.918$

$H(VG) = -\frac{4}{4}\log_2(\frac{4}{4}) - \frac{0}{4}\log_2(\frac{0}{4}) = 0$

$\therefore H(D|A_4) = 0.608$

$\therefore Gain(A_4) = 0.971 - 0.608 = 0.363$

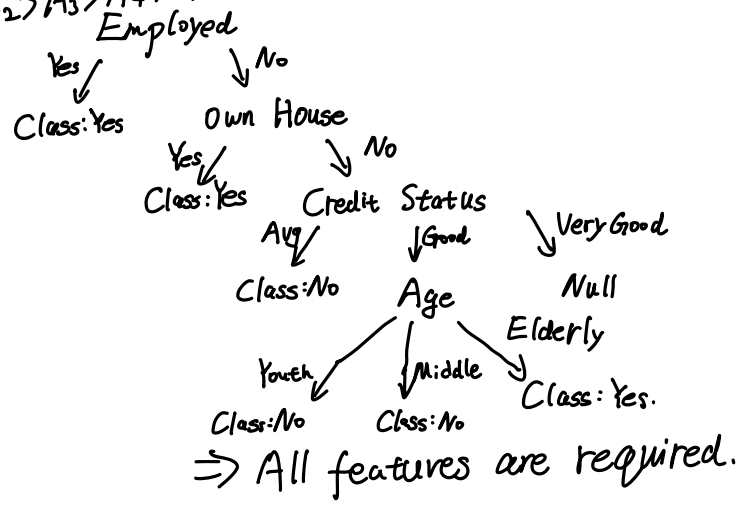
$H(A_4) = -\frac{5}{15}\log_2(\frac{5}{15}) - \frac{6}{15}\log_2(\frac{6}{15}) - \frac{4}{15}\log_2(\frac{4}{15}) = 1.566$

$\therefore$ Gain Ratio$(A_4) = \frac{0.363}{1.566} = 0.232$

2. $\because A_2$ Has the Highest Gain Ratio

$\therefore$ Select A_2: Employed as Root node.

3/4: $A_2 > A_3 > A_4 > A_1$



$\Rightarrow$ All features are required.

1.2.

Conv3(16): $3\times3\times3\times16+16 = 448$ Parameters

$(32-3+2\times1)/1+1=16$ (32, 32, 16) Activation Shape

Maxpool2: 0 Parameters

$(32-2)\div2+1=16$ (16, 16, 16)

Conv5(24): $24\times(5\times5\times16+1) = 9624$

$(16-5+2\times2)\div1+1=16$

Filters 24 $\Rightarrow$ (16, 16, 24)

Maxpool2: Para: 0

$(16-2)\div2+1=8 \Rightarrow (8, 8, 24)$

FC10: (10,)

$1536\times10+10 = 15370$

NO.	Layer	Activation Shape	# Parameters	Mark
1	Input Layer	(32,32,3)	0	-
2	Conv3(16)	(32, 32, 16)	448	[3+2] pts
3	Maxpool2	([6, 16, 16]	0	[3+2] pts
4	Conv5(24)	[16, 16, 24)	9624	[3+2] pts
5	Maxpool2	(8, 8, 24)	0	[3+2] pts
6	FC10	(10,)	15370	[3+2] pts