ELE 364: Assignment #1 Solutions

1. (10 pts) The entropy is:

$$
\begin{aligned}
&= -\left(\frac{a}{a+b}\log_2\frac{a}{a+b} + \frac{b}{a+b}\log_2\frac{b}{a+b}\right) \\
&= -\frac{1}{a+b}\left(a\log_2\frac{a}{a+b} + b\log_2\frac{b}{a+b}\right) \\
&= -\frac{1}{a+b}\left(\log_2\left(\frac{a}{a+b}\right)^a + \log_2\left(\frac{b}{a+b}\right)^b\right) \\
&= -\frac{1}{a+b}\left(\log_2\frac{a^a}{(a+b)^a} + \log_2\frac{b^b}{(a+b)^b}\right) \\
&= -\frac{1}{a+b}\log_2\frac{a^a b^b}{(a+b)^{a+b}} \\
&= \frac{1}{a+b}\log_2\frac{(a+b)^{a+b}}{a^a b^b}
\end{aligned}
$$

2. (10 pts)

(a) $p(25\text{-}50) = p(100\text{-}150) = 0.3$ and $p(75\text{-}100) = 0.4$.
Entropy $= -\sum p \ log_2(p) = -\{2 \times 0.3 \times \log_2(0.3) + 0.4 \times \log_2(0.4)\} = 1.571$ bits.
Gini index $= 1 - \sum p^2 = 1 - \{2 \times 0.3^2 + 0.4^2\} = 0.660$.

(b) Potential mid-point threshold options are 29, 31, 34.5, 39.5, 46, 57, and 67. Table 1 shows the information gain when plotting the data based on these threshold values.

Table 1: Information gain for each of the candidate age thresholds

Split by Threshold	Number of Instances	Partition entropy	Rem.	IG
$\geq$29	2/8	1.000 / 1.561	1.449	0.122
$\geq$31	3/7	0.918 / 1.557	1.365	0.206
$\geq$34.5	4/6	1.000 / 1.459	1.275	0.295
$\geq$39.5	5/5	0.971 / 0.971	0.971	0.600
$\geq$46	7/3	1.449 / 0.918	1.289	0.281
$\geq$57	8/2	1.406 / 0.000	1.125	0.446
$\geq$67	9/1	1.530 / 0.000	1.377	0.194

Age $\geq$39.5 has the highest information gain.

(c) Information gain for the education feature using entropy:
$H(Education = BS) = -\{(1/4) \times \log_2(1/4) + (3/4) \times \log_2(3/4)\} = 0.811$ bits
$H(Education = MS) = -\{(2/3) \times \log_2(2/3) + (1/3) \times \log_2(1/3)\} = 0.918$ bits
$H(Education = PhD) = -\{(1/3) \times \log_2(1/3) + (2/3) \times \log_2(2/3)\} = 0.918$ bits
$rem(Education) = \{0.4 \times 0.811 + 0.3 \times 0.918 + 0.3 \times 0.918\} = 0.875$ bits
Thus, $IG(Education) = H - rem(Education) = 1.571 - 0.875 = 0.696$ bits.

Information gain for the occupation feature using entropy:
$H(Occupation = farmer) = -\{(1/2) \times \log_2(1/2) + (1/2) \times \log_2(1/2)\} = 1.000$ bits

$H(Occupation = professional) = -\{(2/5) \times \log_2(2/5) + (3/5) \times \log_2(3/5)\} = 0.971$ bits
$H(Occupation = teacher) = -\{(2/3) \times \log_2(2/3) + (1/3) \times \log_2(1/3)\} = 0.918$ bits
$rem(Occupation) = \{0.2 \times 1.000 + 0.5 \times 0.971 + 0.3 \times 0.918\} = 0.961$ bits
Thus, $IG(Occupation) = H - rem(Occupation) = 1.571 - 0.961 = 0.610$ bits.

(d) First, we need to calculate the entropy of the dataset with respect to the education feature.
$H(Education) = -\{0.4 \times \log_2(0.4) + 2 \times 0.3 \times \log_2(0.3)\} = 1.571$ bits.
Information gain ratio for education feature:
$GR(Education) = IG(Education)/H(Education) = 0.695/1.571 = 0.443$.

The entropy of the dataset with respect to the occupation feature.
$H(Occupation) = -\{0.2 \times \log_2(0.2) + 0.5 \times \log_2(0.5) + 0.3 \times \log_2(0.3)\} = 1.485$ bits.
Information gain ratio for occupation feature:
$GR(Occupation) = IG(Occupation)/H(Occupation) = 0.610/1.485 = 0.411$.

(e) Information gain for the education feature using the Gini index:
$Gini(Education = BS) = 1 - \{(1/4)^2 + (3/4)^2\} = 0.375$
$Gini(Education = MS) = 1 - \{(2/3)^2 + (1/3)^2\} = 0.444$
$Gini(Education = PhD) = 1 - \{(1/3)^2 + (2/3)^2\} = 0.444$
$rem(Education) = \{0.4 \times 0.375 + 0.3 \times 0.444 + 0.3 \times 0.444\} = 0.416$
Thus, $IG(Education) = Gini - rem(Education) = 0.660 - 0.414 = 0.246$

Information gain for the occupation feature using the Gini index:
$Gini(Occupation = farmer) = 1 - \{(1/2)^2 + (1/2)^2 = 0.500$
$Gini(Occupation = professional) = 1 - \{(2/5)^2 + (3/5)^2 = 0.480$
$Gini(Occupation = teacher) = 1 - \{(2/3)^2 + (1/3)^2 = 0.444$
$rem(Occupation) = \{0.2 \times 0.500 + 0.5 \times 0.480 + 0.3 \times 0.444\} = 0.473$
Thus, $IG(Occupation) = H - rem(Occupation) = 0.660 - 0.473 = 0.187$

3. (10 pts) $\binom{5}{3} \times 0.8^3 \times 0.2^2 + \binom{5}{4} \times 0.8^4 \times 0.2 + \binom{5}{5} \times 0.8^5 = 0.94208$.

4. (10 pts)

(a) Current weights $= .1$ and $\epsilon = .1$. New weights of misclassified instances $= \frac{.1}{2 \times .1} = \frac{1}{2}$.

(b) New weights of correctly classified instances $= \frac{.1}{2 \times .9} = \frac{1}{18}$.

(c) Confidence factor $\alpha = \frac{1}{2} \ln(\frac{.9}{.1}) = 1.0986$.