CS2109S Tutorial 3

AY 25/26 Sem 1 — github/omgeta

A. 1. At root:

$$H(Y) = 1$$
$$H(Y \mid Education) = \frac{4}{10}H(\frac{1}{4}, \frac{3}{4}) + \frac{3}{10}H(\frac{1}{3}, \frac{2}{3}) + \frac{3}{10}H(\frac{3}{3}, 0) = 0.6$$
$$\implies IG(Y; Education) = H(Y) - H(Y \mid Education) = 1 - 0.4 = 0.6$$
$$H(Y \mid Age) = \frac{5}{10}H(\frac{3}{5}, \frac{2}{5}) + \frac{5}{10}H(\frac{2}{5}, \frac{3}{5}) = 0.971$$
$$\implies IG(Y; Age) = H(Y) - H(Y \mid Age) = 1 - 0.971 = 0.029$$
$$H(Y \mid Experience) = \frac{6}{10}H(\frac{3}{6}, \frac{3}{6}) + \frac{4}{10}H(\frac{2}{4}, \frac{2}{4}) = 1$$
$$\implies IG(Experience) = H(Y) - H(Y \mid Experience) = 0$$

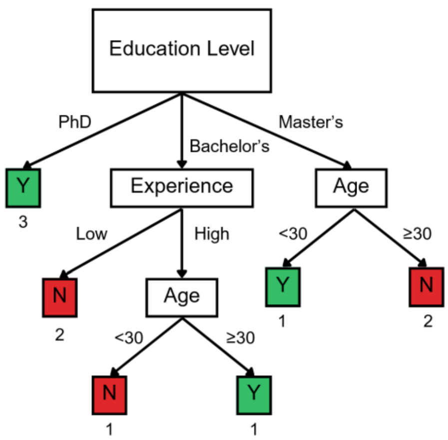
Since Education has the highest information gain, we choose it as root. Since all $Education = PhD$ have same classification, we only build a new subtree for $Education = Masters, Bachelors$. For $Education = Masters$:

$$H(Y \mid Education = Masters) = H(\frac{1}{3}, \frac{2}{3}) = 0.918$$
$$H(Y \mid Education = Masters, Experience) = \frac{2}{3}H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{3}H(1, 0) = 0.667$$
$$\implies IG(Y \mid Education = Masters; Experience) = 0.918 - 0.667 = 0.251$$
$$H(Y \mid Education = Masters, Age) = \frac{1}{3}H(\frac{1}{1}, \frac{0}{1}) + \frac{2}{3}H(\frac{0}{2}, \frac{2}{2}) = 0$$
$$\implies IG(Y \mid Education = Masters; Age) = 0.918 - 0 = 0.918$$

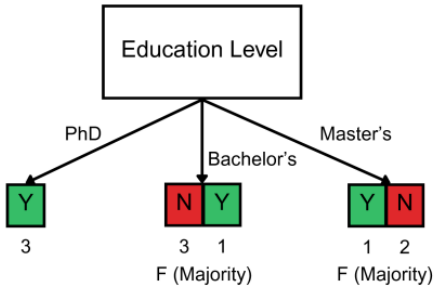
Since Age has the highest information gain, we choose it as subroot. For $Education = Bachelors$:

$$H(Y \mid Education = Bachelors) = H(\frac{1}{4}, \frac{3}{4}) = 0.811$$
$$H(Y \mid Education = Bachelors, Experience) = \frac{2}{4}H(\frac{2}{2}, \frac{0}{2}) + \frac{2}{4}H(\frac{1}{2}, \frac{1}{2}) = 0.5$$
$$\implies IG(Y \mid Education = Bachelors; Experience) = 0.811 - 0.5 = 0.311$$
$$H(Y \mid Education = Bachelors, Age) = \frac{2}{4}H(\frac{0}{2}, \frac{2}{2}) + \frac{2}{4}H(\frac{1}{2}, \frac{1}{2}) = 0.5$$
$$\implies IG(Y \mid Education = Bachelors; Age) = 0.811 - 0.5 = 0.311$$

With equal information gain, we tiebreak choosing Experience as subroot. We are left with splitting by Age in $Experience = High$



2. Outliers would be the applicant with Bachelors, High Experience and Age ≥ 30, as well as the applicant with Masters and age < 3



3. Based on the predicted and actual decisions, $TP = 3, TN = 1, FP = 1, FN = 2$ so that $Accuracy = \frac{4}{7}$, $Precision = \frac{3}{4}$, $Recall = \frac{3}{5}$, $F1 = 0.667$ where $F1 = 0.667 > 0.6$ which indicates the model achieves a stronger balance than expected providing more reliable predictions for the positive class.

B. 1. Construct $X = \begin{pmatrix} 1 & 2 & 1 & 4 & 2 & 1 \\ 1 & 3 & 2 & 9 & 6 & 4 \\ 1 & 5 & 3 & 25 & 15 & 9 \\ 1 & 7 & 4 & 49 & 28 & 16 \\ 1 & 8 & 5 & 64 & 40 & 25 \\ 1 & 9 & 6 & 81 & 54 & 36 \end{pmatrix}$ then $w = \begin{pmatrix} 7.5 \\ -4 \\ 6.5 \\ -9.5 \\ 33.5 \\ -28 \end{pmatrix}$, giving $\hat{y} = 7.5 - 4x_1 + 6.5x_2 - 9.5x_1^2 + 33.5x_1x_2 - 28x_2^2$

2. Column $x_3 = x_1^2 + 2x_1x_2 + x_2^2$ is a linear combination of the existing columns $\implies X$ loses full column rank $\implies X^TX$ is singular. We can drop one of the dependent columns.

C. 1. For $\hat{y} = 2$, $L_{MSE} = 0.005, L_{MAE} = 0.05$;
For $\hat{y} = 4$, $L_{MSE} = 0.405, L_{MAE} = 0.45$

2. MSE magnifies large outliers

D. 1. Given $y = x^2$, $\frac{dy}{dx} = 2x$, then $x_{t+1} = x_t - \gamma 2x_t = (1 - 2\gamma)x_t$

γ	t	x_t	$y_t = x_t^2$
all	0	5	25
	1	-95	9025
	2	1805	3258025
10	3	-34295	1176147025
	4	651605	424589076025
	5	-12380495	153276656445025
	1	-5	25
	2	5	25
1	3	-5	25
	4	5	25
	5	-5	25
	1	4	16
	2	3.2	10.24
0.1	3	2.56	6.5536
	4	2.048	4.1943
	5	1.6384	2.6844
	1	4.9	24.01
	2	4.802	23.0592
0.01	3	4.706	22.1461
	4	4.6118	21.2691
	5	4.5196	20.4268

$\therefore \gamma = 0.1$ converges the fastest

2. Add learning rate decay or an adaptive learning rate