A practical problem

Question 1:

We have the dataset with 4 Low and 4 High of Risk level:

$\Rightarrow S = \{4L; 4H\}$

$$\Rightarrow \begin{cases} p_L = 4/8 = 0,5 \\ p_H = 4/8 = 0,5 \end{cases}$$

$$\text{Entropy}(S) = \sum_j -p_j \log_2 p_j$$

$$= -0,5 \lg_2 0,5 - 0,5 \lg_2 0,5$$
$$= 1$$

Splitting Credit Score at 650 we have 2 group:

- ≤ 650: $\{0L; 4H\}$
- > 650: $\{4L; 0H\}$

$$\Rightarrow \begin{cases} \text{Entropy}(\leq 650) = -1 \lg_2 1 = 0 \\ \text{Entropy}(> 650) = -1 \lg_2 1 = 0 \end{cases}$$

Information Gain:

$$G = \text{Entropy}(S) - \sum \frac{|S_i|}{|S|} \text{Entropy}(S_i)$$

$$= 1 - \frac{4}{8} \cdot 0 - \frac{4}{8} \cdot 0$$

$$= 1$$

The Information Gain = 1 is perfectly separated "High" and "Low" with Credit Score at 650 as the ideal root node.

Question 2:

$$\text{Variance} = \frac{\sum (x_i - \overset{\text{mean}}{\bar{x}})^2}{n-1}$$

$$\text{mean } \bar{x} = \frac{720 + 650 + 750 + 600 + 780 + 630 + 710 + 640}{8}$$

$$= 685$$

$$\text{Variance} = \frac{(720-685)^2 + (650-685)^2 + (750-685)^2 + (600-685)^2 + (630-685)^2 + (710-685)^2 (640-685)^2}{8-1}$$

$$= 4085,7$$

Splitting on Age = 35

≤ 35: $\{720, 650, 600, 630, 640\}$

> 35: $\{750, 780, 710\}$

Group ≤ 35:

$$\text{Mean} = \frac{720 + 650 + 600 + 630 + 640}{5}$$

$$= 648$$

$$\text{Variance}(\leq 35) = \frac{(720-648)^2 + (650-648)^2 + (600-648)^2 + (630-648)^2 + (640-648)^2}{5-1}$$

$$= 1970$$

Group > 35:

$$\text{mean} = \frac{750 + 780 + 710}{3}$$

$$= 746,67$$

$$\text{variance} = \frac{(750-746,67)^2 + (780-746,67)^2 + (710-746,67)^2}{3-1}$$

$= 1233,3$

Weighted:

≤ 35 : $5/8$

> 35 : $3/8$

Weighted variance:

$$\left(\frac{5}{8} \cdot 1970\right) + \left(3/8 \cdot 1233,3\right)$$

$= 1693,74$

Variance Reduction:

= Variance − Weighted Vari

$= 4085,7 - 1693,74$

$= 2391,96$

Question 3:

We have T_2: Age $= 30$; Credit Score $= 645$

Credit Score $\leq 650 \to [2;4;6;8]$

$\Rightarrow$ the probability of High Risk

$P_1 = 1.$

Age $\leq 35 \to [1;2;4;6;8]$

$\Rightarrow$ the probability of High Risk:

$P_2 = 0,75$

P_1 and P_2 are independent.

$\Rightarrow P(\text{High} \mid T_2) = 1-(1-P_1).(1-P_2)$

$= 1-(1-1).(1-0,75)$

$= 1$

the T_2 has 100% probability of being HighRisk

Question 4:

We have:

$\hat{y} = \theta_0 + \theta_1 \cdot x$ | where x: Age

Compute $\hat{y}$:

x_i	$y_i = 500 + 5 \cdot x_i$
35	675
28	640
45	725
31	655
52	760
29	645
42	710
33	665

Cost function:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^{m} (\hat{y}_i - y)^2$$

$$J(\theta) = \frac{1}{2 \cdot 8}\left[(675-720)^2 + (640-650)^2 + \ldots (665-640)^2\right]$$

$= 439,06$

Compute Gradients

$$\frac{\partial J}{\partial \theta_0} = \frac{1}{m} \sum_{i=1}^{m} (\hat{y} - y_i)$$

$$= \frac{1}{8}\left[(675-720) + (640-650) + \ldots + (665-640)\right]$$

$= -0,625$

$$\frac{\partial J}{\partial \theta_1} = \frac{1}{m} \sum_{i=1}^{m} (\hat{y} - y_i) \cdot x_i$$

$$= \frac{1}{8}\left[(675-720).35 + \ldots \ldots (665-640).33\right]$$

$= -131,875$

Update Parameters:

$$\theta_0^+ = \theta_0 - \alpha \frac{\partial J}{\partial \theta_0} = 500 - 0{,}01(-0{,}625)$$
$$= 500{,}00625$$

$$\theta_1^+ = \theta_1 - \alpha \frac{\partial J}{\partial \theta_1} = 5 - 0{,}01(-131{,}875)$$
$$= 6{,}31875$$

the GD is moving in the direction that reduces error and improve prediction