User 1	?	5	3	4	2
User 2	4	1	5	?	5
User 3	5	?	4	2	5
User 4	1	5	3	5	?

Determine missing information,

1. Using probabilistic methods

2. Using RF-Rec predictors

1. X = (Item2=5, Item3=3, Item4=4, Item5=2)

P(X|Item1=1) = P(Item2=5| Item1=1) x P(Item3=3| Item1=1) x P(Item4=4| Item1=1) x P(Item5=2| Item1=1)=1x1x0x0=0

P(X|Item1=2) = P(Item2=5| Item1=2) x P(Item3=3| Item1=2) x P(Item4=4| Item1=2) x P(Item5=2| Item1=2)

P(X|Item1=3)

P(X|Item1=4)

P(X|Item1=5)

2.

f(user1, 1) = Count value 1 in User1 = 0	f(item1, 1) = Count value 1 in Item1 = 1
f(user1, 2) = Count value 2 in User1 = 1	f(item1, 2) = Count value 2 in Item1 = 0
f(user1, 3) = Count value 3 in User1 = 1	f(item1, 3) = Count value 3 in Item1 = 0
f(user1, 4) = Count value 4 in User1 = 1	f(item1, 4) = Count value 1 in Item1 = 1
f(user1, 5) = Count value 5 in User1 = 1	f(item1, 5) = Count value 1 in Item1 = 1

argmax f(user1, x) x f(item1, x) → 4 or 5

f(user2, 1) = Count value 1 in User2 =	f(item4, 1) = Count value 1 in Item4 =
f(user2, 2) = Count value 2 in User2 =	f(item4, 2) = Count value 2 in Item4 =
f(user2, 3) = Count value 3 in User2 =	f(item4, 3) = Count value 3 in Item4 =
f(user2, 4) = Count value 4 in User2 =	f(item4, 4) = Count value 1 in Item4 =
f(user2, 5) = Count value 5 in User2 =	f(item4, 5) = Count value 1 in Item4 =

	Item1	Item2	Item3	Item4	Item5	
Alice	5	3	4	4	?	
User1	3	1	2	3	3	sim = 0,85
User2	4	3	4	3	5	sim = 0,00
User3	3	3	1	5	4	sim = 0,70
User4	1	5	5	2	1	sim = -0,79

User filtering

$$\bar{r}_{Alice} = \frac{5+3+4+4}{4} = 4$$

$$\bar{r}_{User1} = \frac{3+1+2+3}{4} = 2.25$$

$$sim(Alice, User1) = \frac{(5-4)(3-2.25)+(3-4)(1-2.25)+\ldots}{\sqrt{(5-4)^2+(3-4)^2+\ldots} \times \sqrt{(3-2.25)^2+(1-2.25)^2+\ldots}}$$

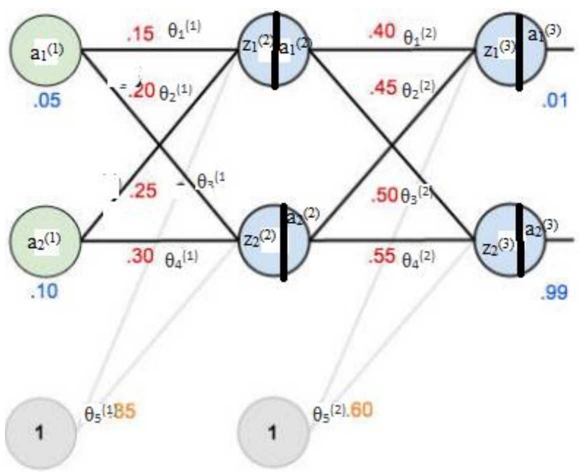
$$pred(Alice, Item5) = \bar{r}_{Alice} + \frac{0.85(3-2.25)+0\times\left(5-\frac{4+3+4+3}{4}\right)+\ldots}{0.85+0+\ldots}$$

Item filtering

$$\bar{r}_{Item5} = \frac{3+5+4+1}{4} = 3.25$$

$$\bar{r}_{Item1} = \frac{3+4+3+1}{4} = 2.75$$

$$sim(\text{Item}5, \text{Item}1) = \frac{(3-3.25)(3-2.75)+(5-3.25)(4-2.75)+\dots}{\sqrt{(3-3.25)^2+(5-3.25)^2+\dots} \times \sqrt{(3-2.75)^2+(4-2.75)^2+\dots}}$$



- Step 1: The forward pass

$$\begin{cases} z_1^{(2)} = \theta_1^{(1)} a_1^{(1)} + \theta_2^{(1)} a_2^{(1)} + \theta_5^{(1)} = 0.15 * 0.05 + 0.2 * 0.1 + 0.85 = 0.3775 \\ a_1^{(2)} = \frac{1}{1+e^{-z_1^{(2)}}} = \frac{1}{1+e^{-0.3775}} = 0.592 \end{cases}$$

$$\begin{cases} z_2^{(2)} = \theta_3^{(1)} a_1^{(1)} + \theta_4^{(1)} a_2^{(1)} + \theta_5^{(1)} \\ a_2^{(2)} = \frac{1}{1+e^{-z_2^{(2)}}} = 0.596 \end{cases}$$

$$\begin{cases} z_1^{(3)} = \theta_1^{(2)} a_1^{(2)} + \theta_2^{(2)} a_2^{(2)} + \theta_5^{(2)} \\ a_1^{(3)} = \frac{1}{1+e^{-z_1^{(3)}}} = 0.751 \end{cases}$$

$$\begin{cases} z_2^{(3)} = \theta_3^{(2)} a_1^{(2)} + \theta_4^{(2)} a_2^{(2)} + \theta_5^{(2)} \\ a_2^{(3)} = \frac{1}{1+e^{-z_2^{(3)}}} = 0.772 \end{cases}$$

- Step 2: Calculating Cost function

$$J\left(\mathbf{a}^{(3)}\right) = \sum \frac{1}{2}\left(target - a^{(3)}\right)^2 = \frac{1}{2}\left(target - a_1^{(3)}\right)^2 + \frac{1}{2}\left(target - a_2^{(3)}\right)^2 = \frac{1}{2}(0.01-0.751)^2 + \frac{1}{2}(0.99-0.772)^2 = 0.298$$

- Step 3: The Back Propagation
  - Now, you need to calculate $\frac{\partial J}{\partial \theta_1^{(2)}}$ to update $\theta_1^{(2)}$ by gradient descent
  - $$\frac{\partial J}{\partial \theta_1^{(2)}} = \frac{\partial J}{\partial a_1^{(3)}} \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} \frac{\partial z_1^{(3)}}{\partial \theta_1^{(2)}}$$
    - $$\frac{\partial J}{\partial a_1^{(3)}} = -\frac{1}{2} 2\left(target - a_1^{(3)}\right) = -(0.01-0.751) = 0.74$$

- $a_1^{(3)} = \frac{1}{1+e^{-z_1^{(3)}}} \rightarrow \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} = a_1^{(3)}\left(1-a_1^{(3)}\right) = 0.75(1-0.75) = 0.186$
- $z_1^{(3)} = \theta_1^{(2)} a_1^{(2)} + \theta_2^{(2)} a_2^{(2)} + \theta_5^{(2)} \rightarrow \frac{\partial z_1^{(3)}}{\partial \theta_1^{(2)}} = a_1^{(2)} = 0.592$

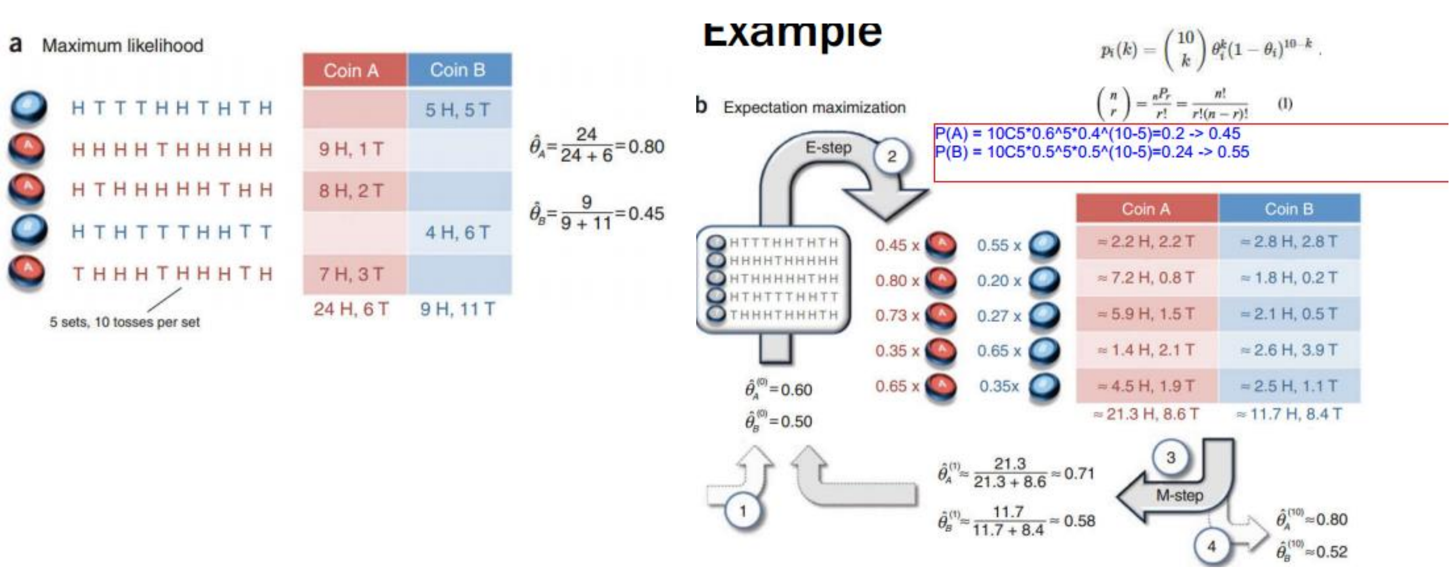
$$\rightarrow \frac{\partial J}{\partial \theta_1^{(2)}} = \frac{\partial J}{\partial a_1^{(3)}} \frac{\partial a_1^{(3)}}{\partial z_1^{(3)}} \frac{\partial z_1^{(3)}}{\partial \theta_1^{(2)}} = 0.74 * 0.186 * 0.592 = 0.082$$

$$\rightarrow \theta_1^{(2)} = \theta_1^{(2)} - \frac{\partial J}{\partial \theta_1^{(2)}}$$

- Similarly, you can calculate all $\theta^{(2)}$
- However, When calculating the $\theta^{(1)}$, you need more calculation
- $\frac{\partial J}{\partial \theta_1^{(1)}} = \frac{\partial J}{\partial a_1^{(2)}} \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \frac{\partial z_1^{(2)}}{\partial \theta_1^{(1)}}$
  - $\frac{\partial J}{\partial a_1^{(2)}} = \frac{\partial J\left(a_1^{(3)}\right)}{\partial a_1^{(2)}} + \frac{\partial J\left(a_2^{(3)}\right)}{\partial a_1^{(2)}} = \left(\frac{\partial J\left(a_1^{(3)}\right)}{\partial z_1^{(3)}} \frac{\partial z_1^{(3)}}{\partial a_1^{(2)}} + \frac{\partial J\left(a_2^{(3)}\right)}{\partial z_2^{(3)}} \frac{\partial z_2^{(3)}}{\partial a_1^{(2)}}\right) = \left(0.13 * 0.4 + (-0.019)\right) = 0.0363$
  - $a_1^{(2)} = \frac{1}{1+e^{-z_1^{(2)}}} \rightarrow \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} = a_1^{(2)}\left(1-a_1^{(2)}\right) = 0.59(1-0.759) = 0.241$
  - $z_1^{(2)} = \theta_1^{(1)} a_1^{(1)} + \theta_2^{(1)} a_2^{(1)} + \theta_5^{(1)} \rightarrow \frac{\partial z_1^{(2)}}{\partial \theta_1^{(1)}} = a_1^{(1)} = 0.05$

$$\rightarrow \frac{\partial J}{\partial \theta_1^{(1)}} = \frac{\partial J}{\partial a_1^{(2)}} \frac{\partial a_1^{(2)}}{\partial z_1^{(2)}} \frac{\partial z_1^{(2)}}{\partial \theta_1^{(1)}} = 0.0363 * 0.241 * 0.05 = 0.0004$$

$$\rightarrow \theta_1^{(1)} = \theta_1^{(1)} - \frac{\partial J}{\partial \theta_1^{(1)}}$$



- Given the dataset, calculate Naïve Bayes to find the class of **last row**

Example	Sky	AirTemp	Humidity	Wind	Water	Forecast	EnjoySport
1	Sunny	Warm	Normal	Strong	Warm	Same	Yes
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rainy	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes
5	Cloudy	Warm	High	Weak	Cool	Same	Yes
6	Cloudy	Cold	High	Weak	Cool	Same	No

7	Sunny	Warm	Normal	Strong	Warm	Same	?
8	Sunny	Warm	Low	Strong	Cool	Same	?

- $P(C_i)$: $\begin{cases} P(class='yes') = \frac{4}{6} = \frac{2}{3} \\ P(class='no') = \frac{2}{6} = \frac{1}{3} \end{cases}$
- P(Sky= 'Sunny' | class = 'Yes') = $\frac{3}{4}$
- P(Sky= 'Sunny' | class = 'No') = 0
- P(Sky= 'Rainy'| class= 'Yes')=0
- P(Sky= 'Rainy'| class= 'No')= 1
- Notice: value 'Low' is not in 'Humidity', so P(Humidity= 'Low'| class= 'Yes') =0
- P(Humidity= 'Low'| class= 'No')=0

- Put X = (Sunny, Warm, Low, Strong, Cool, Same)

- P(X | Class = 'Yes') = P(Sky= 'Sunny' | class = 'Yes') * P(AirTemp= 'Warm' | class = 'Yes') * P(Humidity= 'Low' | class = 'Yes') * P(Wind= 'Strong' | class = 'Yes') * P(Water= 'Cool' | class = 'Yes') * P(Forecast= 'Same' | class = 'Yes')

- P(X | Class = 'No) = P(Sky= 'Sunny' | class = 'No) * P(AirTemp= 'Warm' | class = 'No) * P(Humidity= 'Low' | class = 'No) * P(Wind= 'Strong' | class = 'No) * P(Water= 'Cool' | class = 'No) * P(Forecast= 'Same' | class = 'No)
- Compare: P(X | Class = 'Yes') * P(Class = 'Yes) and P(X | Class = 'No') * P(Class = 'No'). Choose Class max

mpg	cylinders	displacement	horsepower	weight	acceleration	model-year	maker
good	4	low	low	low	high	75to78	Asia
bad	6	medium	medium	medium	medium	70to74	America
bad	4	medium	medium	medium	low	75to78	Europe
bad	8	high	high	high	low	70to74	America
bad	6	medium	medium	medium	medium	70to74	America
bad	4	low	medium	low	medium	70to74	America
bad	4	low	medium	low	low	70to74	Asia
bad	8	high	high	high	low	75to78	America
bad	8	high	high	high	low	70to74	America
good	8	high	medium	high	high	79to83	America
bad	8	high	high	high	low	70to74	America
good	4	low	low	low	low	79to83	America
bad	6	medium	medium	medium	high	75to78	America
good	4	medium	low	low	low	79to83	America
good	4	low	low	medium	high	79to83	America
bad	8	high	high	high	low	70to74	America
good	4	low	medium	low	medium	75to78	Europe
bad	5	medium	medium	medium	medium	75to78	Europe

Iteration #1:

Entropy(S) = $-p_{good}\log_2(p_{good}) - p_{bad}\log_2(p_{bad}) = -\frac{6}{18}\log_2\left(\frac{6}{18}\right) - \frac{12}{18}\log_2\left(\frac{12}{18}\right) = 0.91$

Gain(S, cylinders) =

$$Entropy(S) - \sum_{v\in(4,5,6,8)} (S_v/S)Entropy(S_v) = Entropy(S) - \left[\frac{8}{18}Entropy(S_4) + \frac{1}{18}Entropy(S_5) + \frac{3}{18}Entropy(S_6) + \frac{6}{18}Entropy(S_8)\right]$$

$$= 0.91 - \left[\frac{8}{18}\times\left(-\frac{5}{8}\log_2\frac{5}{8} - \frac{3}{8}\log_2\frac{3}{8}\right) + 0 + 0 + \frac{6}{18}\times\left(-\frac{5}{6}\log_2\frac{5}{6} - \frac{1}{6}\log_2\frac{1}{6}\right)\right] = 0.27$$

Gain(S, displacement) =

$$Entropy(S)-\sum_{v\in(low.medium,high)}\frac{S_v}{S}Entropy(S_v)=0.91-\left[\frac{6}{18}Entropy(S_{low})+\frac{6}{18}Entropy(S_{medium})+\frac{6}{18}Entropy(S_{high})\right]=0.17$$

Gain(S, horsepower) =

$$Entropy(S)-\sum_{v\in(low.medium,high)}\frac{S_v}{S}Entropy(S_v)=0.91-\left[\frac{4}{18}Entropy(S_{low})+\frac{9}{18}Entropy(S_{medium})+\frac{5}{18}Entropy(S_{high})\right]=0.52$$

Gain(S, weight) =

$$Entropy(S)-\sum_{v\in(low.medium,high)}\frac{S_v}{S}Entropy(S_v)=0.91-\left[\frac{6}{18}Entropy(S_{low})+\frac{6}{18}Entropy(S_{medium})+\frac{6}{18}Entropy(S_{high})\right]=0.17$$

Gain(S, acceleration) =

$$Entropy(S)-\sum_{v\in(low.medium,high)}\frac{S_v}{S}Entropy(S_v)=0.91-\left[\frac{9}{18}Entropy(S_{low})+\frac{5}{18}Entropy(S_{medium})+\frac{4}{18}Entropy(S_{high})\right]=0.15$$

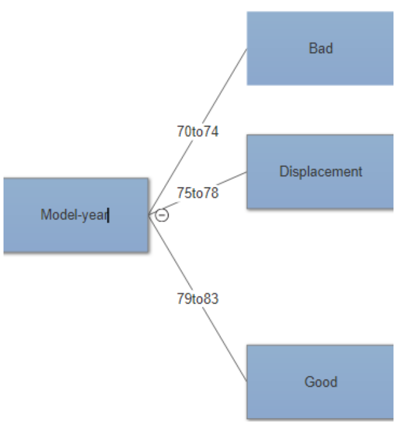
Gain(S, model-year) =

$$Entropy(S)-\sum_{v\in(70\text{to}74.75\text{to}78,79\text{to}83)}\frac{S_v}{S}Entropy(S_v)=0.91-\left[\frac{8}{18}Entropy(S_{70to74})+\frac{6}{18}Entropy(S_{75to78})+\frac{4}{18}Entropy(S_{79to83})\right]=0.6$$

Gain(S, maker) =

$$Entropy(S)-\sum_{v\in(Asia,\text{Europe},\text{America})}\frac{S_v}{S}Entropy(S_v)=0.91-\left[\frac{2}{18}Entropy(S_{Asia})+\frac{3}{18}Entropy(S_{Europe})+\frac{13}{18}Entropy(S_{America})\right]=0.005$$

➔ Choose model-year to split



Dataset after iteration #1:

mpg	cylinders	displacement	horsepower	weight	acceleration	model-year	maker
good	4	low	low	low	high	75to78	Asia
bad	4	medium	medium	medium	low	75to78	Europe
bad	8	high	high	high	low	75to78	America
bad	6	medium	medium	medium	high	75to78	America
good	4	low	medium	low	medium	75to78	Europe
bad	5	medium	medium	medium	medium	75to78	Europe

Iteration #2:

Entropy(S) = $-p_{good}\log_2(p_{good})-p_{bad}\log_2(p_{bad})=0.91$

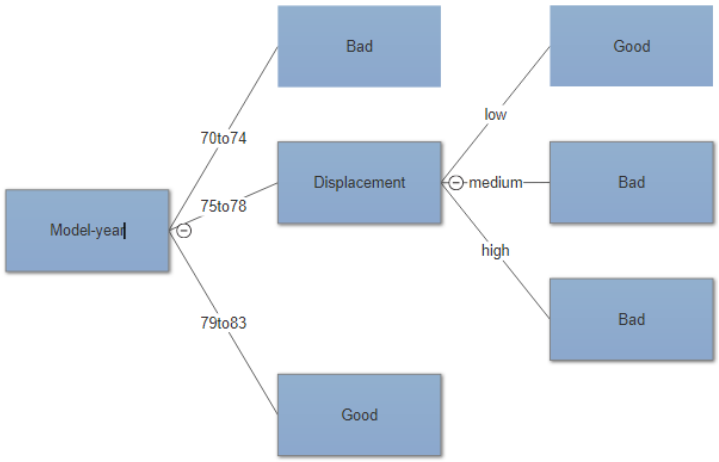
Gain(S, cylinders) =

$$Entropy(S)-\sum_{v\in(4,5,6,8)}(S_v/S)Entropy(S_v)=Entropy(S)-\left[\frac{3}{6}Entropy(S_4)+\frac{1}{6}Entropy(S_5)+\frac{1}{6}Entropy(S_6)+\frac{1}{6}Entropy(S_8)\right]=0.45$$

Gain(S, displacement) =

$$Entropy(S) - \sum_{v \in (low.medium,high)} \frac{S_v}{S} Entropy(S_v) = 0.91 - \left[\frac{2}{6} Entropy(S_{low}) + \frac{3}{6} Entropy(S_{medium}) + \frac{1}{6} Entropy(S_{high}) \right] = 0.91$$

➔ Don't need to calculate another features because Gain(S, displacement) is the highest Gain you can obtain, so choose Displace to split



- The decision tree is not overfitting because Model-year is categorized rather than continuous. Therefore, you don't need to use Pruning technique

However, if test data point has model-year values which is less than 70 or greater than 83, your decision tree built above cannot predict