School of Data and Computer Science, Sun Yat-sen University

T04 Machine Learning

17341137 Zhenpeng Song

December 21, 2019

Contents

1 Q1

To Judge which attribute to select, we should calculate these two important indices:

- Entropy: $Ent(D) = -\sum_{k=1}^{|Y|} p_k log_2 p_k$, Y stands for the possible values for the label(In this case: User Action, —Y— = 2).
- Info_Gain: $\text{Gain}(D, a) = \text{Ent}(D) - \sum_{v=1}^{V} \frac{|D^v|}{|D|} \text{Ent}\left(D^v\right)$

Therefore, I produce the graph below:

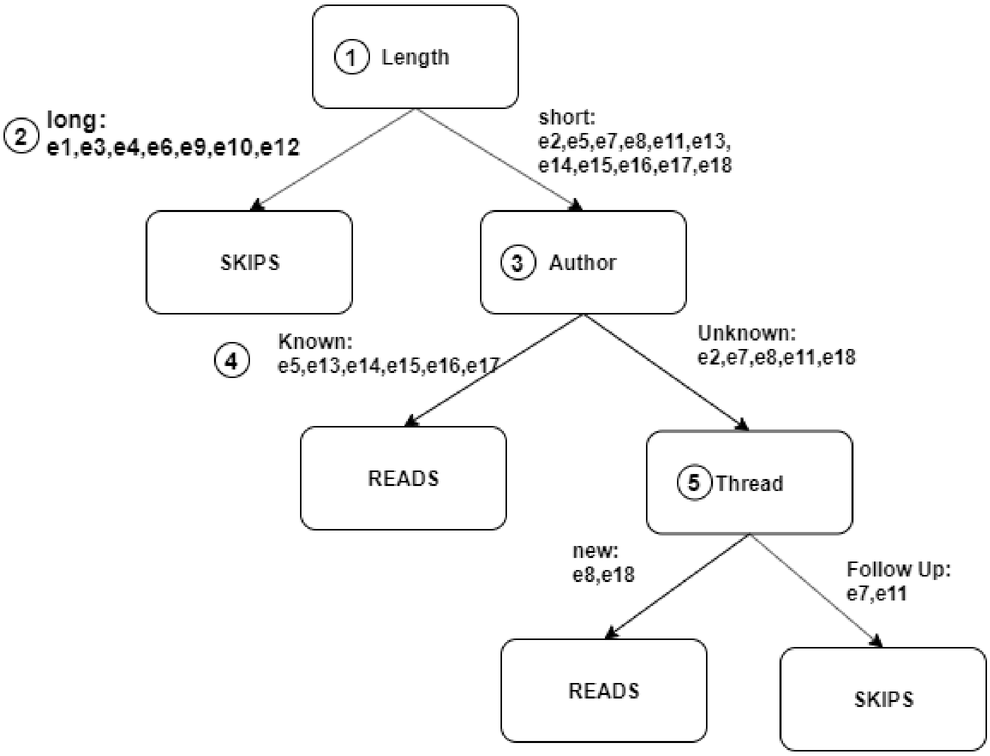

Figure 1: Decision Tree

Where steps 1 to 5 are:

1. Chosen attribute: **Length**. $D \rightarrow D1, D2$.
2. Calculate Info_Gain: $Ent(D1) = 0, Ent(D2) = -\frac{2}{11} \times log_2(\frac{2}{11}) - \frac{9}{11} \times log_2(\frac{9}{11}) = 0.6840$.
   - $Author = 0.6840 - \frac{5}{11} \times 0.9710 - \frac{6}{11} \times 0 = 0.2426$
   - $Thread = 0.6840 - \frac{4}{11} \times 1 - \frac{7}{11} \times 1 = -0.3160$
   - $WhereRead = 0.6840 - \frac{5}{11} \times 0.7219 - \frac{6}{11} \times 0.6500 = 0.0013$

3. Choose attribute: **Author**.$D2 \rightarrow D3, D4$.

4. Calculate Info_Gain: $Ent(D3) = 0, Ent(D4) = -\frac{2}{5} \times log_2(\frac{2}{5}) - \frac{3}{5} \times log_2(\frac{3}{5}) = 0.9710$

   - $Thread = 0.9710 - \frac{3}{5} \times 0 - \frac{2}{5} \times 0 = 0.9710$
   - $WhereRead = 0.9710 - \frac{1}{5} \times 0 - \frac{4}{5} \times 0.8113 = 0.3219$

5. Choose attribute: **Thread**.

6. Done.

2 Q2

2.1 Bayesian

According to the description:

- $h =< 0.1, 0.2, 0.4, 0.2, 0.1 >$
- $d =< lime, cherry, cherry, lime, lime >$

hypothesis	h1	h2	h3	h4	h5
P(lime|hi)	0	0.25	0.5	0.75	1
P(hi)	0.1	0.2	0.4	0.2	0.1
P(d|hi)	0	0.578125	0.03125	0.484375	0

To predict the 6th candy, calculate:

$$\begin{aligned} P(lime\big|d) &= \sum_i P(lime\big|h_i)P(h_i\big|d) && (1)\\ &= \frac{1}{P(d)}\sum_i P(lime\big|h_i)P(d\big|h_i)P(h_i) && (2)\\ &= \frac{0.25\times0.2\times0.578125+0.5\times0.4\times0.03125+0.75\times0.2\times0.484375}{0.2\times0.578125+0.4\times0.03125+0.2\times0.454375} && (3)\\ &= \frac{0.10781175}{0.219} && (4)\\ &= 0.49 && (5)\\ P(cherry\big|d) &= 1-0.49 = 0.51 && (6) \end{aligned}$$

The 6th candy is more possibly to be of cherry taste.

2.2 MAP

MAP make prediction based on most probable hypothesis:

$$h_{MAP} = \text{argmax}_{h_i}$$

$$P(h_i\big|d) = P(X\big|d) \simeq P(X\big|h_{MAP})$$

With $d =< lime, cherry, cherry, lime, lime >$:

Prediction after:

1. lime: $h_{MAP} = h_3, Pr(lime|h_{MAP}) = 0.5$
2. cherry: $h_{MAP} = h_3, Pr(lime|h_{MAP}) = 0.5$

3. cherry: $h_{MAP} = h_3, Pr(lime|h_{MAP}) = 0.5$

4. lime: $h_{MAP} = h_3, Pr(lime|h_{MAP}) = 0.5$

5. lime: $h_{MAP} = h_3, Pr(lime|h_{MAP}) = 0.5$

With hypothesis 3, $Pr(lime\big|h_3) = 0.5 = Pr(cherry\big|h_3)$

The 6th candy has the same possibility to be of cherry or lime taste.

2.3 ML

ML make prediction based on most probable conditional hypothesis:

$$h_{ML} = \text{argmax}_h P(d\big|h)$$

With $d =< lime, cherry, cherry, lime, lime >$:

Prediction after:

1. lime: $h_{ML} = h_5, Pr(lime|h_{ML}) = 1$

2. cherry: $h_{ML} = h_3, Pr(lime|h_{ML}) = 0.5$

3. cherry: $h_{ML} = h_2, Pr(lime|h_{ML}) = 0.25$

4. lime: $h_{ML} = h_3, Pr(lime|h_{ML}) = 0.5$

5. lime: $h_{ML} = h_4, Pr(lime|h_{ML}) = 0.75$

With hypothesis 4, $Pr(lime\big|h_4) = 0.75 > Pr(cherry\big|h_3) = 0.25$

The 6th candy is more possibly to be of lime taste.

3 Q3

According to Naive Bayes Classification, these two rules are followed:

$$P\left(y \mid x_1, \ldots, x_n\right) \propto P(y) \prod_{i=1}^n P\left(x_i \mid y\right)$$

$$\hat{y}=\arg \max_y P(y) \prod_{i=1}^n P\left(x_i \mid y\right)$$

Then, build the truth table without term A = 1, B = 1, C = 1, D = 1.

A	B	C	D	Value
0	0	0	0	0
0	0	0	1	0
0	0	1	0	0
0	0	1	1	0
0	1	0	0	0
0	1	0	1	1
0	1	1	0	1
0	1	1	1	0
1	0	0	0	0
1	0	0	1	1
1	0	1	0	1
1	0	1	1	0
1	1	0	0	0
1	1	0	1	0
1	1	1	0	0

So with the input $A=1, B=1, C=1, D=1$:

$$\hat{y}=\arg \max_y P(y) \prod_{i=1}^n P\left(x_i \mid y\right) \tag{7}$$

$$=\arg \max_y \{ \tag{8}$$

$$P(1) \times P(x_1=1|1) \times P(x_2=1|1) \times P(x_3=1|1) \times P(x_4=1|1), \tag{9}$$

$$P(0) \times P(x_1=1|0) \times P(x_2=1|0) \times P(x_3=1|0) \times P(x_4=1|0)\} \tag{10}$$

$$=\arg \max_y \{\frac{4}{15} \times \frac{1}{2}^4, \frac{11}{15} \times \frac{5}{11}^4\} \tag{11}$$

$$=\arg \max_y \{0.0167, 0.0313\} \tag{12}$$

$$=\mathbf{0} \tag{13}$$

4 Q4

Since XOR operation is NOT linearly separable, a straight forward neural network without hidden layer will definitely not learn well.
However, AND operation, NOT operation, OR operation is linearly separable, it's possible to build an XOR Neural Network by combining those operations using a hidden layer.

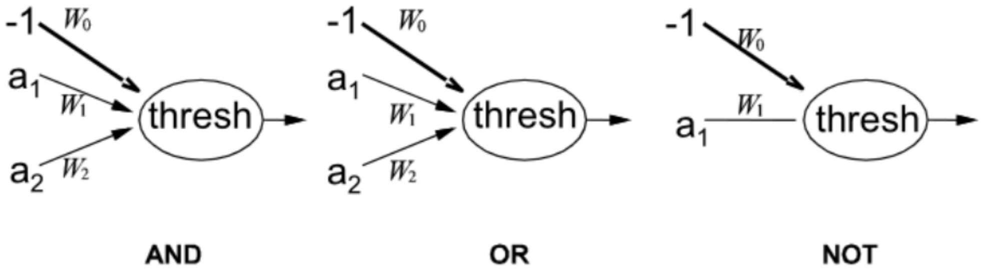

Figure 2: References

For convenience, for sigmoid function: $g(z) = \dfrac{1}{1+e^{-z}}$, I define a threshold of 10, which means for z over 10 or less than -10, the value certainly approach 1 or 0.

So, define **NOT** layer, **AND** layer and **OR** layer:

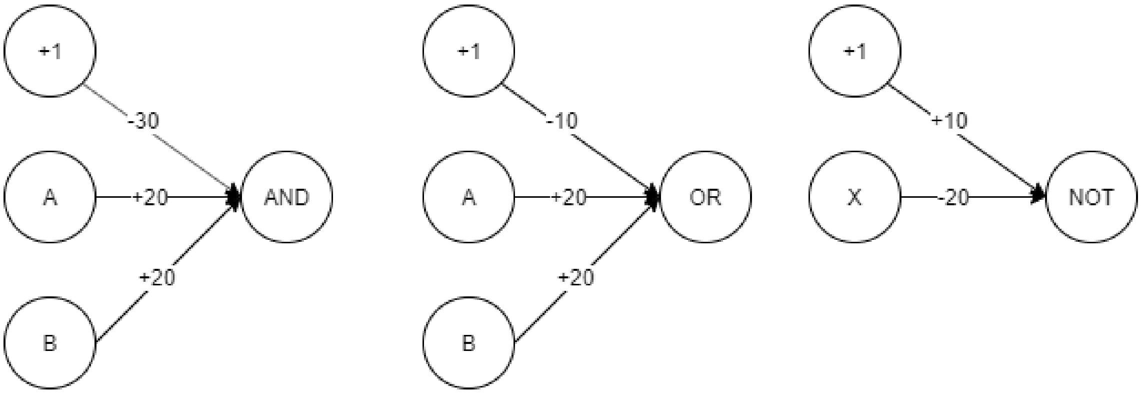

Figure 3: Implementations

And then, build the XOR NN as below:

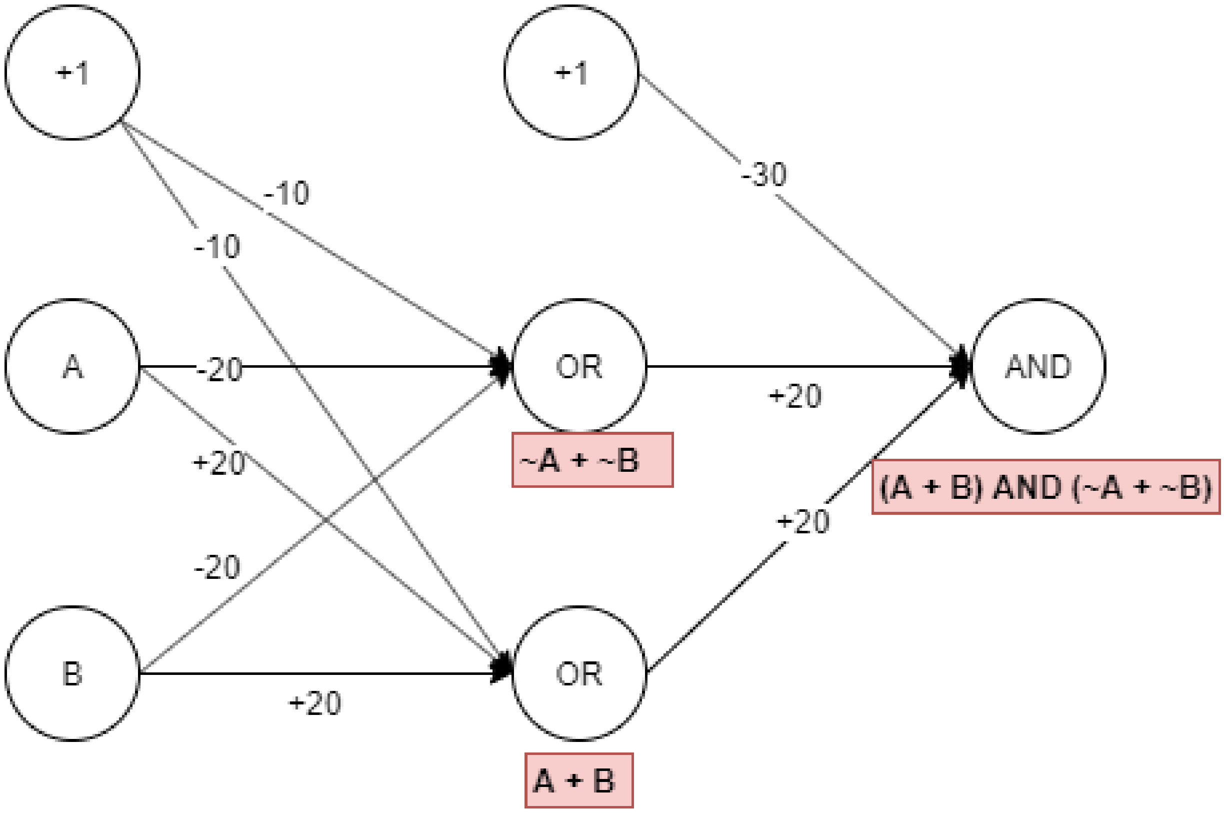

Figure 4: XOR

For a certain input, this model will compute A XOR B.

5 Q5

$$\Delta_{o_i} = g'\left(in_{o_i}\right)\left(y_{o_i} - a_{o_i}\right) = a_{o_i}\left(1 - a_{o_i}\right)\left(y_{o_i} - a_{o_i}\right) \tag{14}$$

$$\Delta_{o_1} = 0.75136507 \times (10.75136507) \times (0.010.75136507) = 0.138498562 \tag{15}$$

$$\Delta_{o_2} = 0.772928465 \times (10.772928465) \times (0.990.772928465) = 0.0380982366 \tag{16}$$

$$in_{h_1} = w_1 i_1 + w_2 i_2 + b_1 = 0.05 \times 0.15 + 0.10 \times 0.20 + 0.35 = 0.3775 \tag{17}$$

$$in_{h_2} = w_2 i_1 + w_4 i_2 + b_1 = 0.05 \times 0.20 + 0.10 \times 0.30 + 0.35 = 0.39 \tag{18}$$

5.1 Using sigmoid function

$$\frac{\partial\,\mathrm{Loss}_{o_1}}{\partial\,\mathrm{w}_1} = i_1 \cdot \Delta_{h_1} \tag{19}$$

$$= 0.05 \times g'\left(inh_1\right)\left(w_5 \Delta_{o_1}\right) \tag{20}$$

$$= 0.05 \times g(0.3775)(1 - g(0.3775)) \times (0.4 \times \Delta_{o_1}) \tag{21}$$

$$= 0.05 \times \frac{1}{1 + e^{-0.3775}} \times (1 - \frac{1}{1 + e^{-0.3775}}) \times (0.4 \times (-0.138)) \tag{22}$$

$$= 0.05 \times 0.241300709 \times (-0.0552) \tag{23}$$

$$= -0.000665988 \tag{24}$$

5.2 Using tanh function

$$\frac{\partial\,\mathrm{Loss}_{o_2}}{\partial\,\mathrm{w}_4} = i_2 \cdot \Delta_{h_2} \tag{25}$$

$$= 0.05 \times g'\left(inh_2\right)\left(w_8 \Delta_{o_2}\right) \tag{26}$$

$$= 0.05 \times (1 - tanh^2(0.39)) \times (0.55 \times \Delta_{o_2}) \tag{27}$$

$$= 0.05 \times 0.862091581151508 \times (0.55 \times 0.0380982366) \tag{28}$$

$$= 0.000903214648313402 \tag{29}$$