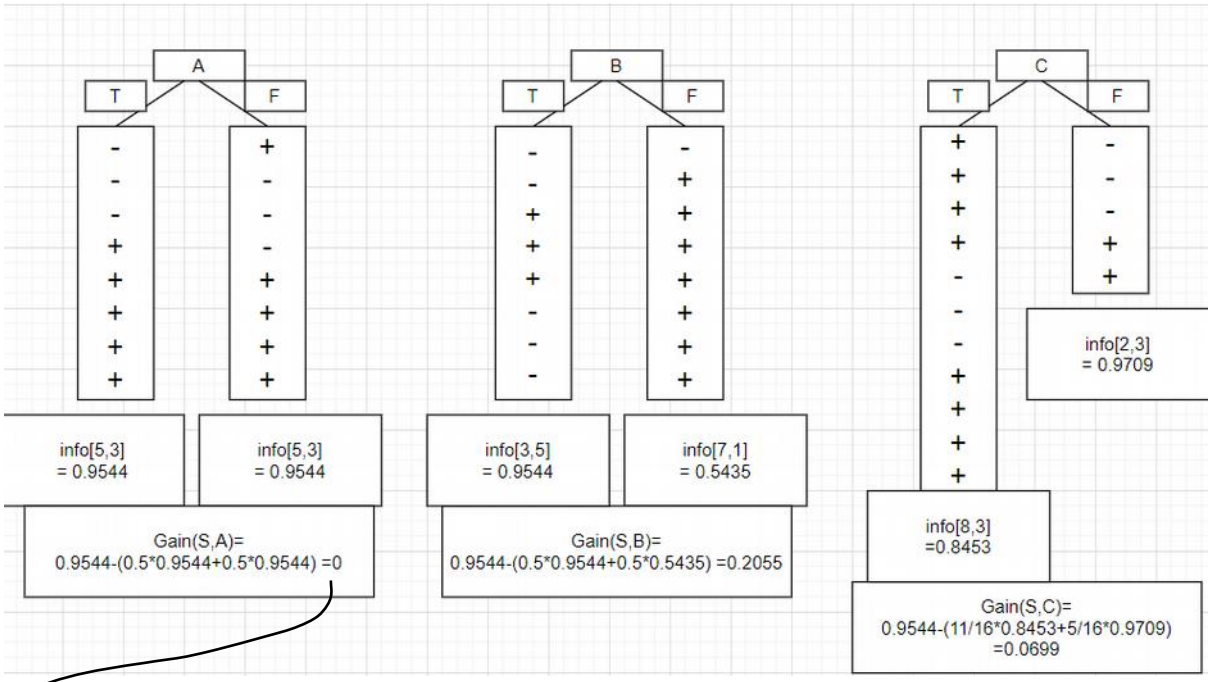
Homework 5

JAESANGPARK

5.1

Entropy[10/16 , 6/16] = 0.9544



Since the information gain of B is the largest, so the decision tree would be built with B.

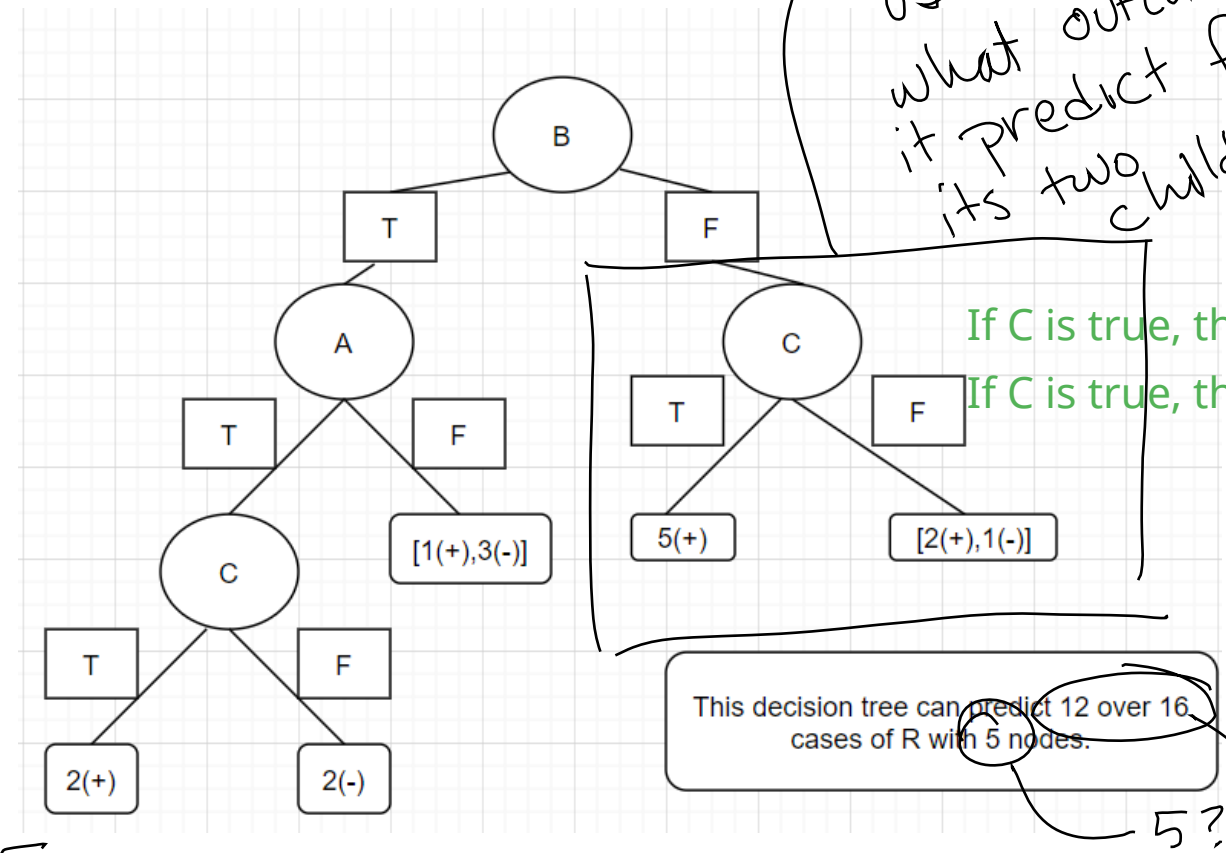
yes

yes - do you see why?

I think there is no change about the distribution.

The distribution of (+,-) is still conserved.

5.2



is this decision node contributing useful information - what outcome would it predict for its two children?

If C is true, then B is tr

If C is true, then 2 over

5?

I think there are 4 no

The tree with only one node in previous problem can't predict any case of R.

Since there is no data about (A,B,C) = (T,T,T) , (F,T,F) and (F,F,F), we can't make adequate decision tree.

I think I need more general training set to make the performance better.

it does better than that!

Sure it can. It's just wrong sometimes. How often is it wrong? what about if the tree has no decision nodes?

Without decision nodes, how can we predict the result?

maybe, maybe not. the tree is capable of making predictions in all of those cases too — they just might be wrong ..

5.3

a. unigram

If the probability of the occurrence of each word is independent, then the probability of the sentence is the product of the probability of each word.

That means,

$p(w=the)*p(w=council)*p(w=has)*p(w=approved)*p(w=wind)*p(w=farms)$

=0.05 * 0.000075 * 0.0035 * 0.00015 * 0.00002* 0.000015

= 0.000000000000000000000590625

b. bigram

If the probability of the occurrence of each word is dependent on previous word, then the probability of the sentence is :

$$P(w_1, w_2, \ldots . w_n) = P(w_1) * \prod_{i=2}^{n} P(w_i \vee w_{i-1})$$

$p(w_1=the)*p(w_2=council|w_1=the¿*p¿$

$p(w_4=approved|w_3=has)*p(w_5=wind|w_4=approved)*p(w_6=farms \vee w_5=wind)$

0.05*0.0004*0.00000 * 0.0006*0.0000*0.00000

= 0

c. pos-driven model

With this model, I have to consider the hidden feature of each word, which means is that noun? verb?.

And like the bigram, I need to consider the order of each word.

$P(D \rightarrow the)*P(D \rightarrow N)*P(N \rightarrow council)*P(N \rightarrow V)*P(V \rightarrow has)*P(V \rightarrow V)$

¿$P(V \rightarrow approved)*P(V \rightarrow N)*P(N \rightarrow wind)*P(N \rightarrow N)*P(N \rightarrow farms)$

0.6* 0.65*0.0002*0.15*0.03*0.15*0.001*0.15*0.000025*0.3*0.00004

=0.000000000000000000000236925

P(v->v)*P(v->wind)*P(v->v)*P(v->farms)

P(v->v)*P(v->wind)*P(v->n)*P(n->farms)

P(v>n)*P(n>wind)*P(n>v)*P(v>farms)

5.4

a. unigram

My utterance is “the the the the the the.”

It is definitely not that good English.

And the probability is $p(w=the)^6$ =0.000000015625 >>>0.00000000000000000000590625

b. bigram

My utterance is “the approved the approved the approved.”

Since the $p(w_i=the \vee w_{i-1}=approved)$ is the largest, I would use that pattern.

And the $p(w=the)$ is the best.

So the probability of my utterance is

$$p(w=the)*\left(p(w_i=approved|w_{i-1}=the)\right)^3*\left(p(w_i=the|w_{i-1}=approved)\right)^2$$

which is absolutely larger than 0.

yes, it's even high compared to many sentences without the zero problem

c. pos-driven model

My utterance is “has has has has has has.”

Since the $p(V \to V)*p(V \to has)$ is the largest followed by $p(D \to D)*p(D \to the)$,

I choose this combination.

The probability of my utterance is $p(V \to V)^5*p(V \to has)^6$.

which is 0.0000000000000553584375.

It is 23,365,384 times higher than the previous “The council has approved wind farms.”

that's a good one