Problem1

filter:

Filter methods select the attributes independent to training algorithm. Generally filter methods applied to dataset before training, and there is no interaction between attributes selection methods and training algorithm.

wrapper:

Wrapper methods select the attributes according to each algorithm. There are interaction between attribute selection methods and training algorithm. Wrapper select attributes according to the performance of training algorithm.

Bagging:

Bagging repeatedly resample from dataset, then train classifier on each sample set. The final prediction is by doing unweight vote of all trained classifiers.

Boosting

Boosting keep train classifier on weighted dataset. And boosting will increase the misclassified instance’s weight to make the training algorithm more focus on those instances.

Problem2

Let $Y = \{H, T\}$ be the observation of the third coin.

Let $X = \{HHHT\ HTHH\ TTTH\}$ be the observation of the sequence.

Let $\Theta = \{\pi, \mu_A, \mu_B\}$

$P(Y|\Theta) = \pi$ if $Y=H$ or $(1-\pi)$ if $Y=T$

$P(X|Y, \Theta) = \mu_A^h * (1 - \mu_A)^t$ if $Y=H$ or $\mu_B^h * (1 - \mu_B)^t$ if $Y=T$

Then, we have

$P(X=HHHT, Y=H|\Theta) = \pi * \mu_A^3 * (1 - \mu_A)^1$

$P(X=HHHT, Y=T|\Theta) = (1 - \pi) * \mu_B^3 * (1 - \mu_B)^1$

$P(X=HTHH, Y=H|\Theta) = \pi * \mu_A^3 * (1 - \mu_A)^1$

$P(X=HTHH, Y=T|\Theta) = (1 - \pi) * \mu_B^3 * (1 - \mu_B)^1$

$P(X=TTTH, Y=H|\Theta) = \pi * \mu_A^1 * (1 - \mu_A)^3$

$P(X=TTTH, Y=T|\Theta) = (1 - \pi) * \mu_B^1 * (1 - \mu_B)^3$

$$P(Y=H|X=HHHT) = \frac{P(X=HHHT, Y=H|\Theta)}{P(X=HHHT, Y=H|\Theta) + P(X=HHHT, Y=T|\Theta)}$$

$$P(Y=T|X=HHHT) = \frac{P(X=HHHT, Y=T|\Theta)}{P(X=HHHT, Y=H|\Theta) + P(X=HHHT, Y=T|\Theta)}$$

$$P(Y=H|X=HTHH) = \frac{P(X=HTHH, Y=H|\Theta)}{P(X=HTHH, Y=H|\Theta) + P(X=HTHH, Y=T|\Theta)}$$

$$P(Y=T|X=HTHH) = \frac{P(X=HTHH, Y=T|\Theta)}{P(X=HTHH, Y=H|\Theta) + P(X=HTHH, Y=T|\Theta)}$$

$$P(Y=H|X=TTTH) = \frac{P(X=TTTH, Y=H|\Theta)}{P(X=TTTH, Y=H|\Theta) + P(X=TTTH, Y=T|\Theta)}$$

$$P(Y=T|X=TTTH) = \frac{P(X=TTTH, Y=T|\Theta)}{P(X=TTTH, Y=H|\Theta) + P(X=TTTH, Y=T|\Theta)}$$

According to the possibilities, we can update Θ as:

$$\pi = \frac{P(Y=H|X=HHHT) + P(Y=H|X=HTHH) + P(Y=H|X=TTTH)}{3}$$

$$\mu_A = \frac{3 * P(Y=H|X=HHHT) + 3 * P(Y=H|X=HTHH) + 1 * P(Y=H|X=TTTH)}{4 * P(Y=H|X=HHHT) + 4 * P(Y=H|X=HTHH) + 4 * P(Y=H|X=TTTH)}$$

$$\mu_B = \frac{3 * P(Y=T|X=HHHT) + 3 * P(Y=T|X=HTHH) + 1 * P(Y=T|X=TTTH)}{4 * P(Y=T|X=HHHT) + 4 * P(Y=T|X=HTHH) + 4 * P(Y=T|X=TTTH)}$$

The initial value of Θ is:

$\pi=0.5$, $\mu_A = 0.6$ $\mu_B = 0.4$

Value of π, μ_A, μ_B in each iteration is:

iteration	π	μ_A	μ_B
1	0.5641025641025641	0.651517956312477	0.4497466741396895
2	0.5832562761214085	0.5832562761214085	0.5832562761214085
3	0.5832562761214085	0.5832562761214085	0.5832562761214085

Problem 3

Suppose we could denote each item with a letter, then we denote the six item as A to F.

For the first level:

Itemset	Support count
A	5
B	3
C	6
D	5
E	4
F	4

For the second level:

Itemset	Support count
BC	3
CE	3
CF	3
DF	3

For the third level, no item set combination have support count greater or equal than 3.

The maximal frequent sets in first 3 levels are: {A}, {B,C}, {C,E}, {C,F},

{D,F}

If item set {B,C} was picked, the support of {B,C} is 0.3 and it will always greater or equals than 0.3 because support threshold is 3 and the total transaction is 10.

If subset {B} was chosen, then confidence of {B} => {C} is 3/3 = 1. If {C} was chosen, then confidence of {C} => {B} is 3/6 = 0.5. That means item C appears in transaction that contain B more often that item B appears in transaction that contain C.

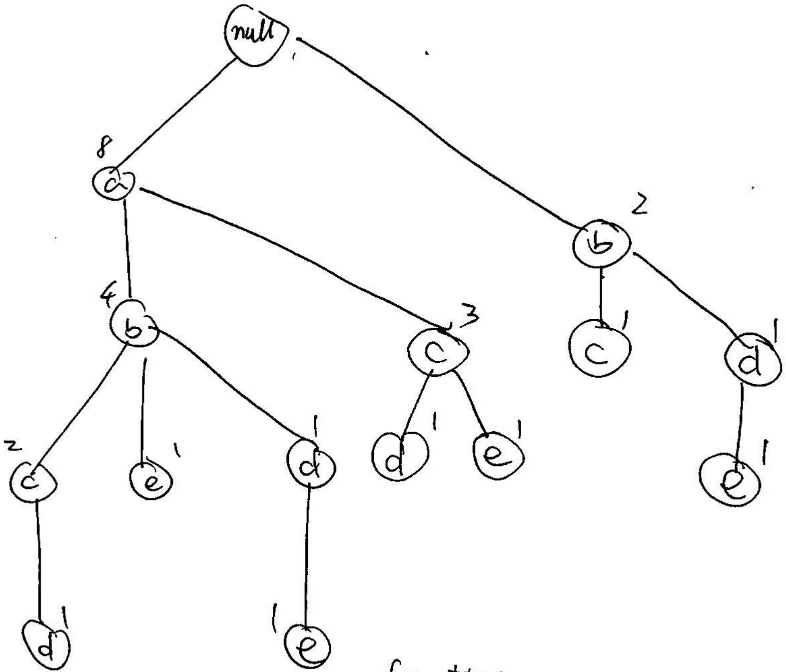
Problem 4

Support count of each item is:

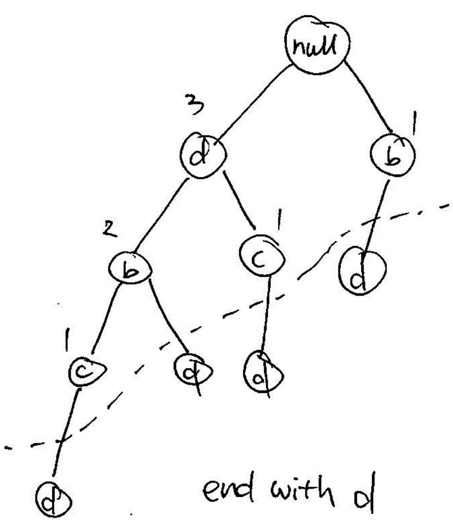
Item	Support count
A	8
B	6
C	6
D	4
E	4
F	1

And the f-list is ABCDE

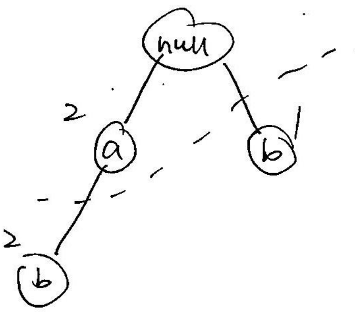
Please refer to the picture below.



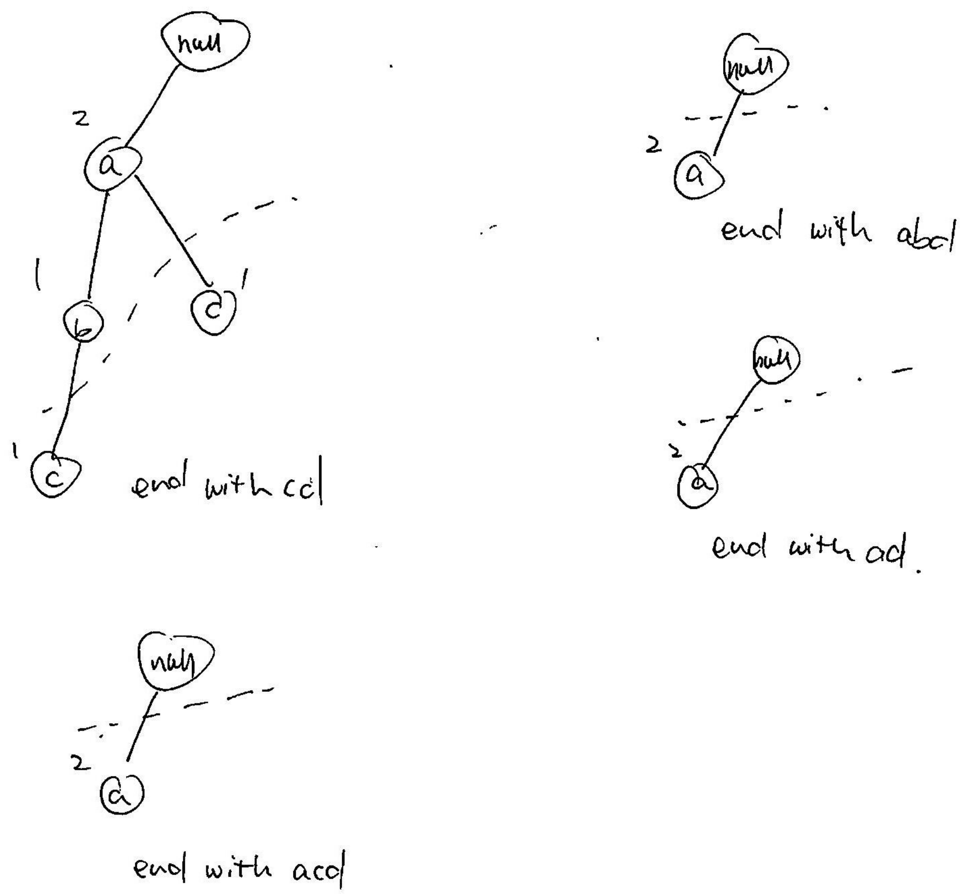
fp-tree



end with d



end with bd



According to these trees, frequent itemset ending with D are: {D}, {C, D}, {A, C, D}, {B, D}, {A, B, D}, {A, D}