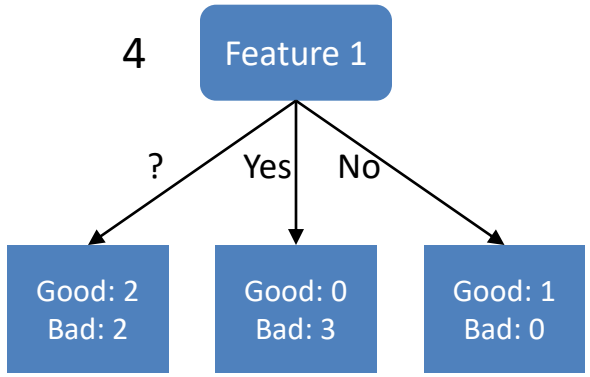
Decision Tree

Prof. Seungchul Lee

Industrial AI Lab.

Feature Test

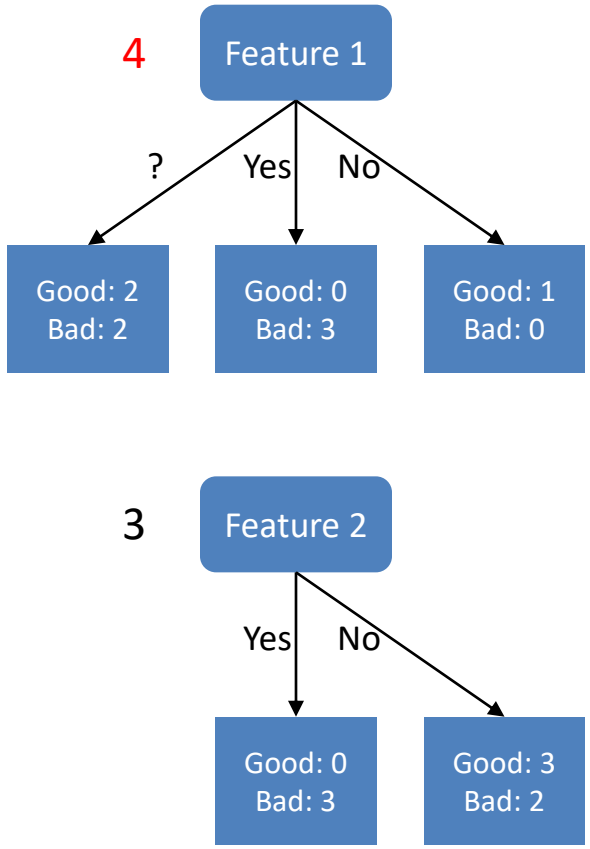
- Feature test in the first level



Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad

Feature Test

- Homogeneous set in the first level

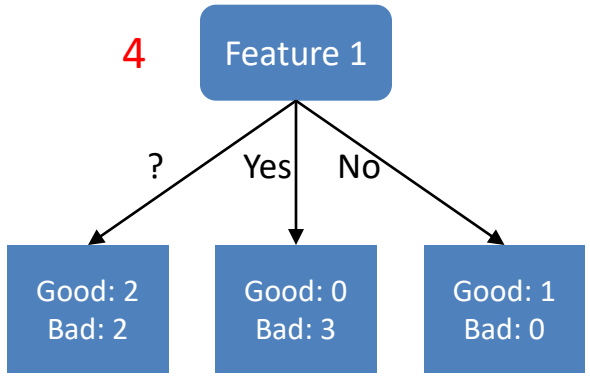


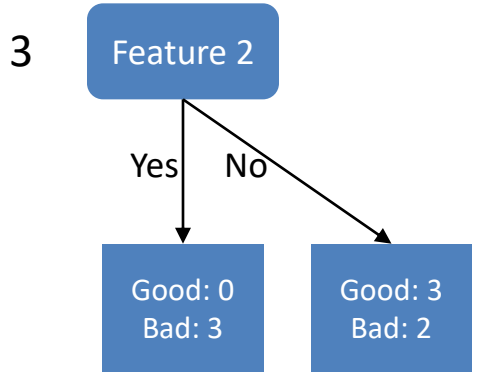
Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad

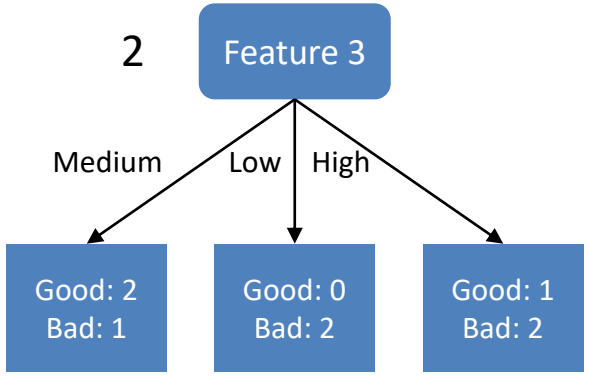
Feature Test

- Homogeneous set in the first level

Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad



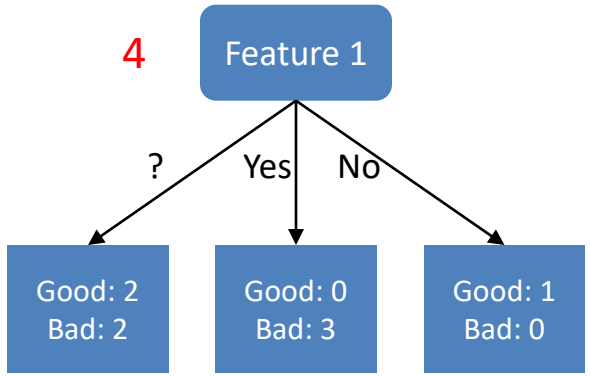


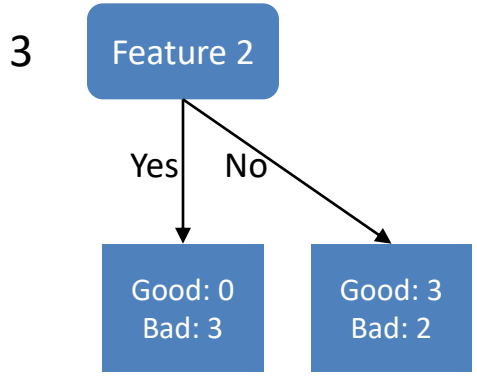


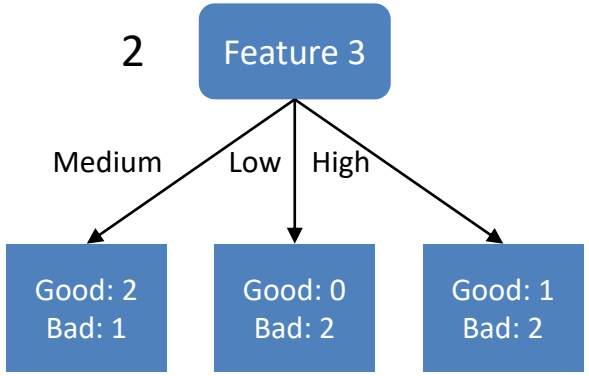
Feature Test

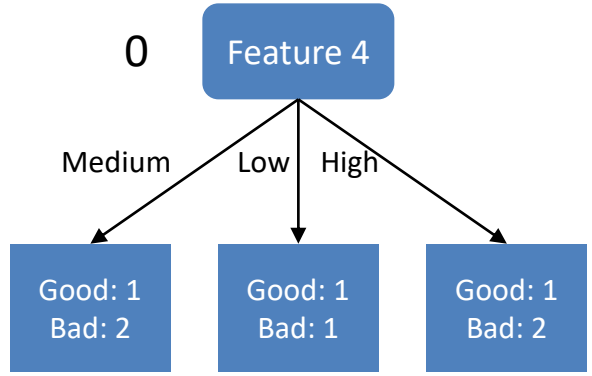
- Homogeneous set in the first level

Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad



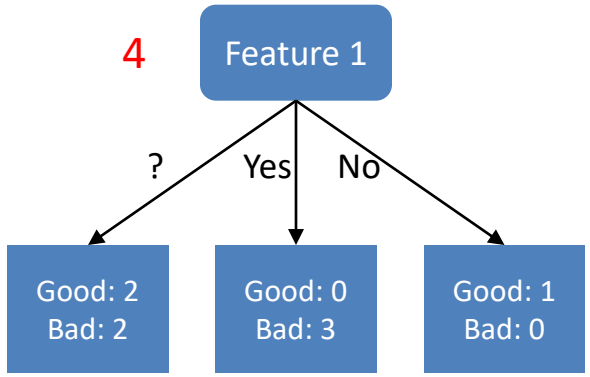






Feature Test

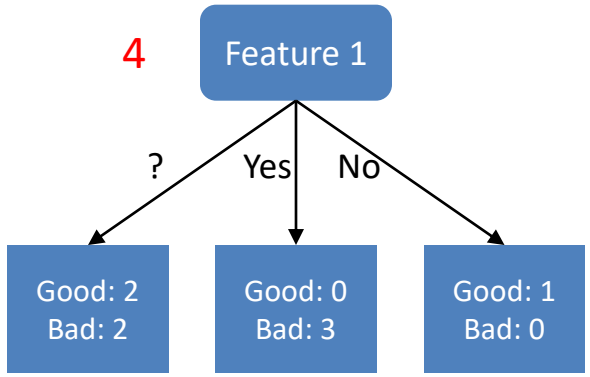
- Homogeneous set in the first level



Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad

Feature Test

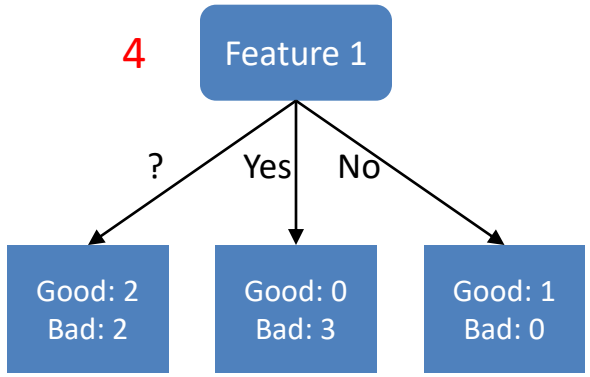
- Homogeneous set in the second level

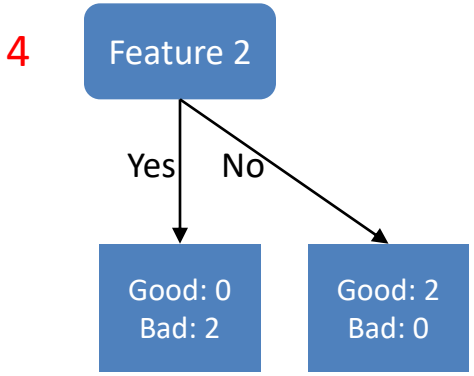


Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad

Feature Test

- Homogeneous set in the second level



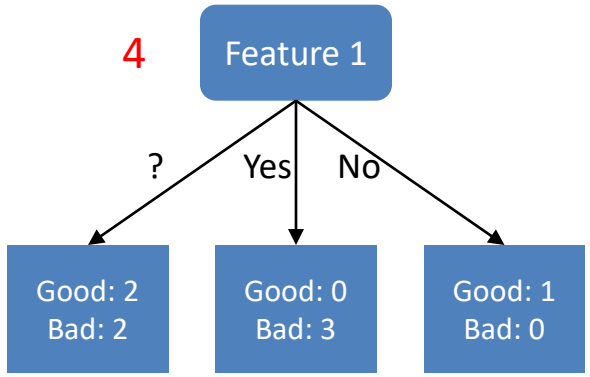


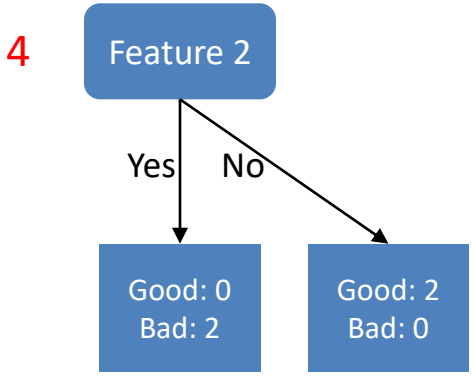
Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad

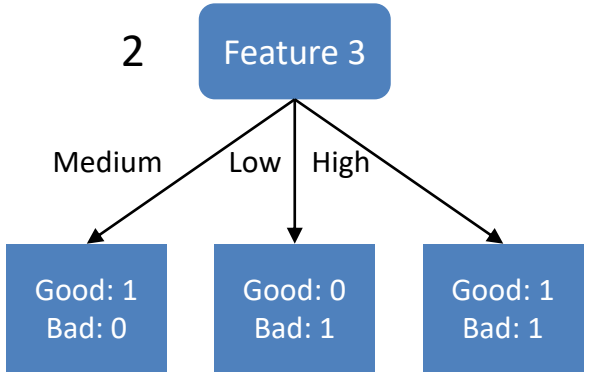
Feature Test

- Homogeneous set in the second level

Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad



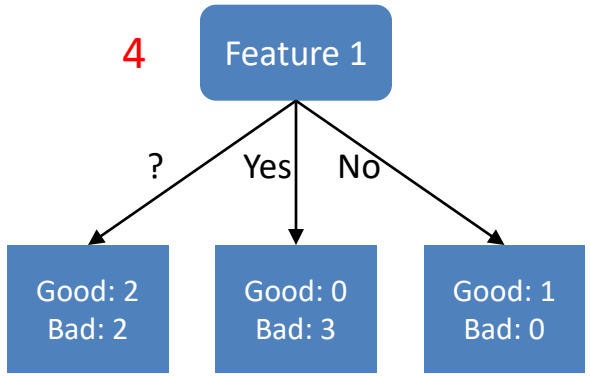


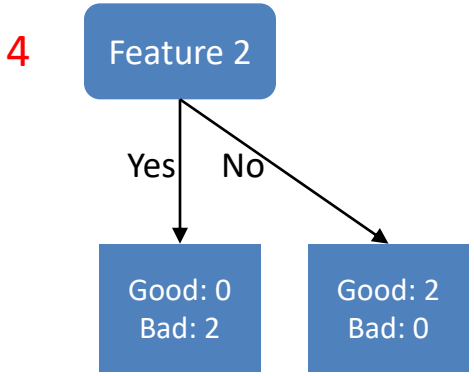


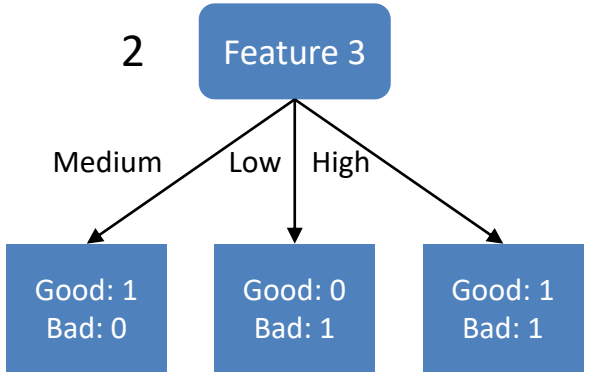
Feature Test

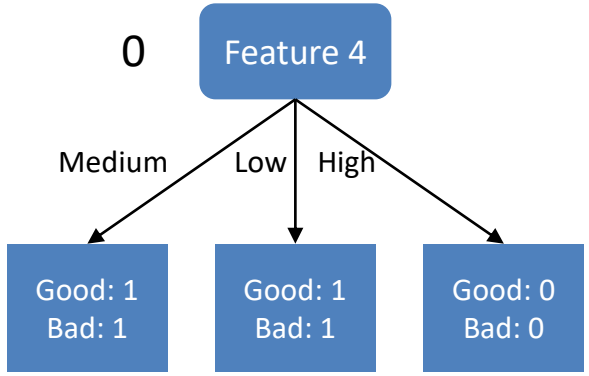
- Homogeneous set in the second level

Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad



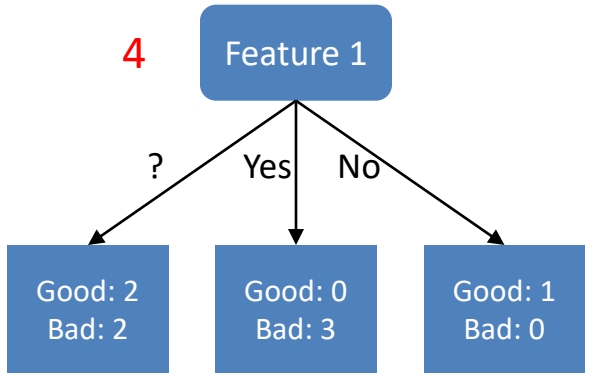


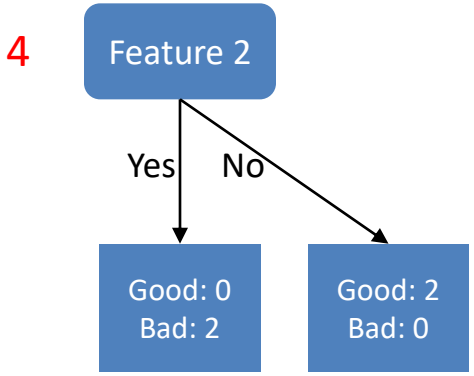




Feature Test

- Homogeneous set in the second level

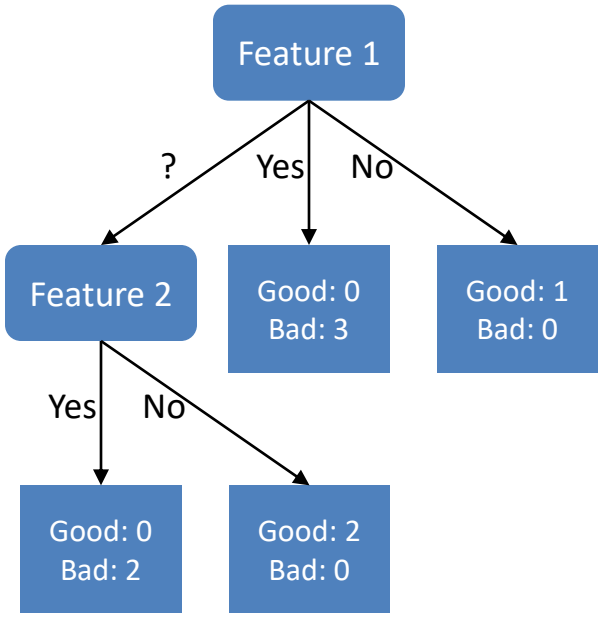




Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad

Feature Test

- Decision tree



Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad

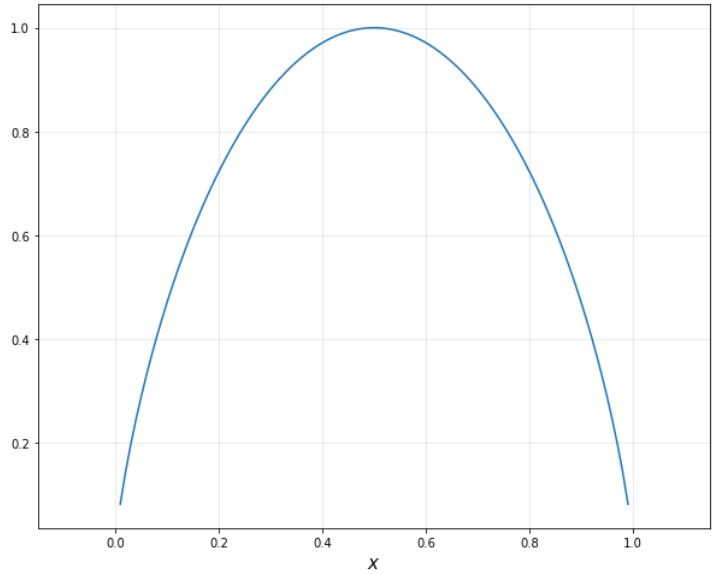
- Issue: a large data set will be most likely to produce zero homogeneous set

Disorder

- Disorder of single set

$$D = -x \log_2 x - (1-x) \log_2 (1-x)$$

$$D = -\frac{G}{T} \log_2 \frac{G}{T} - \frac{B}{T} \log_2 \frac{B}{T}$$

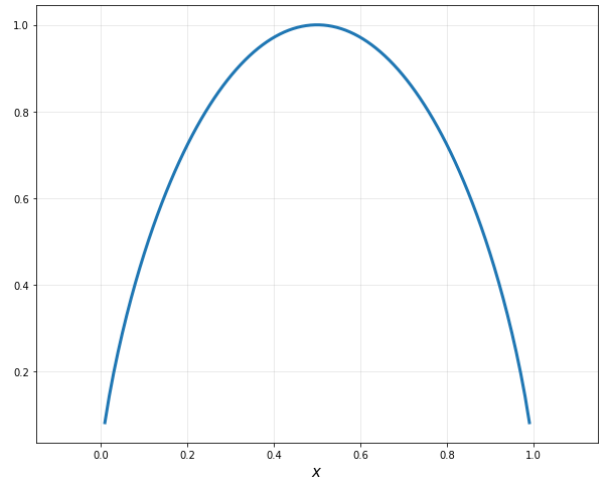


Quality of Test

- Disorder of single set

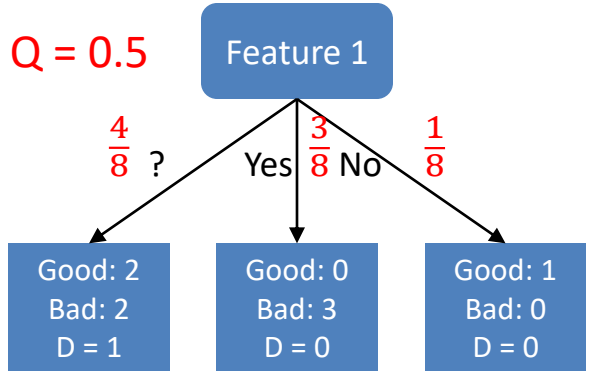
$$D = -x \log_2 x - (1-x) \log_2 (1-x)$$

$$D = -\frac{G}{T} \log_2 \frac{G}{T} - \frac{B}{T} \log_2 \frac{B}{T}$$



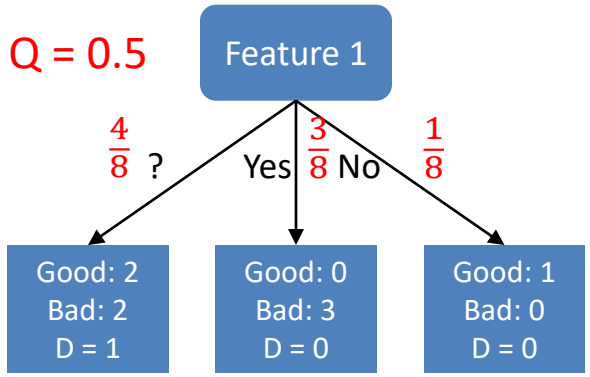
- Quality of test

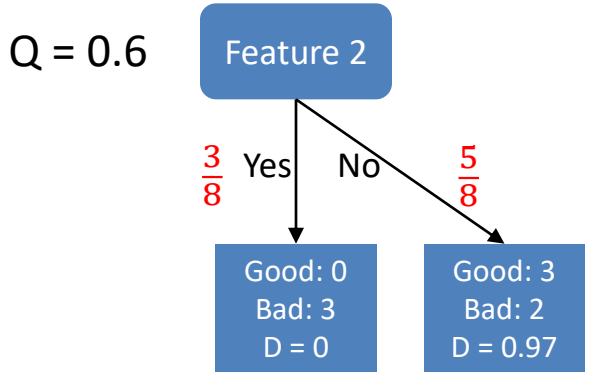
$$Q(\text{test}) = \sum D(\text{set}) \times \frac{\#\text{ of samples in set}}{\#\text{ of samples in all sets}}$$



Quality

- Quality of test in the first level



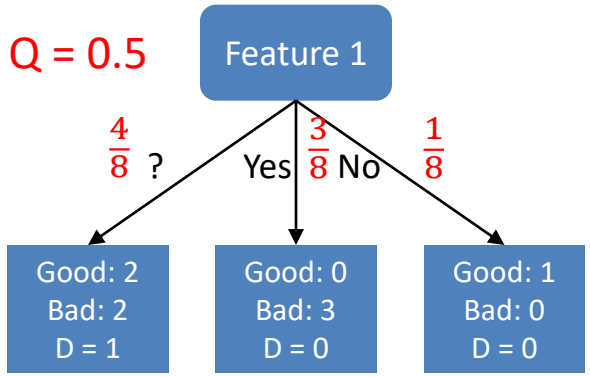


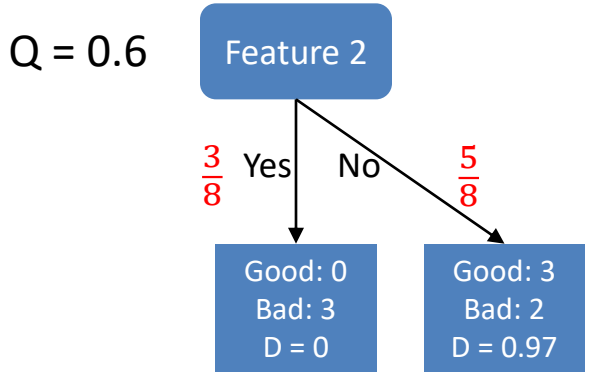
Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad

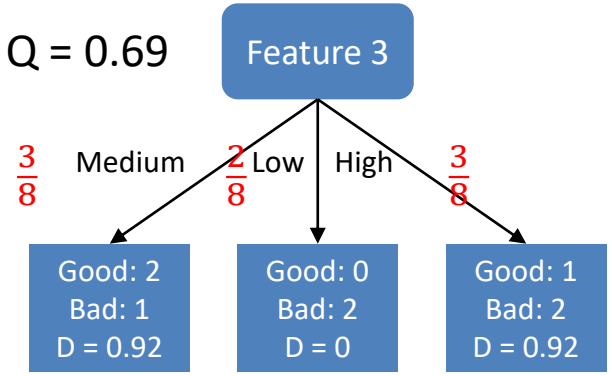
Quality

- Quality of test in the first level

Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad



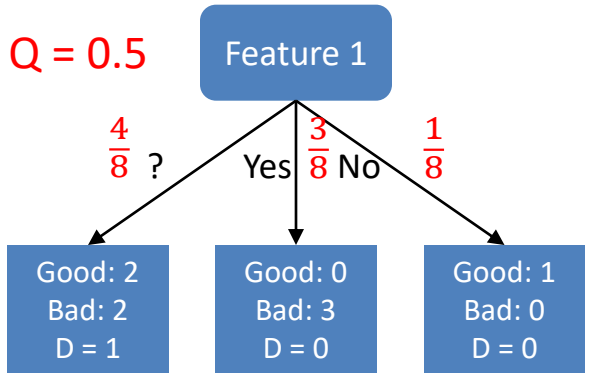


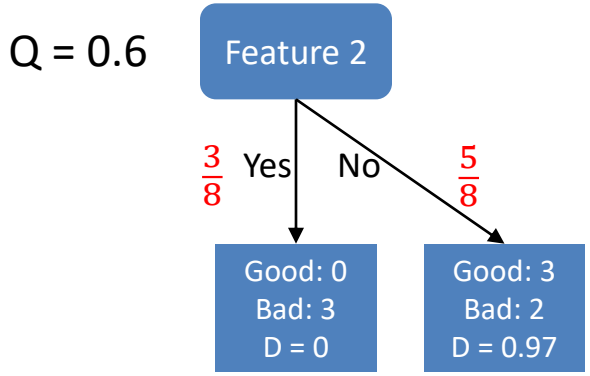


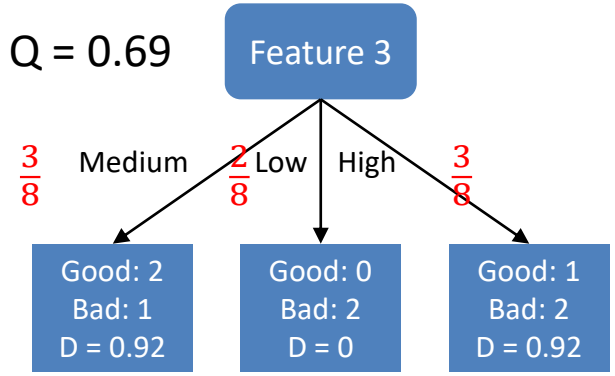
Quality

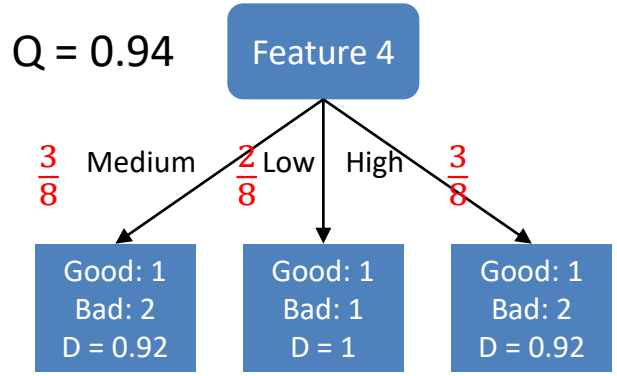
- Quality of test in the first level

Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad



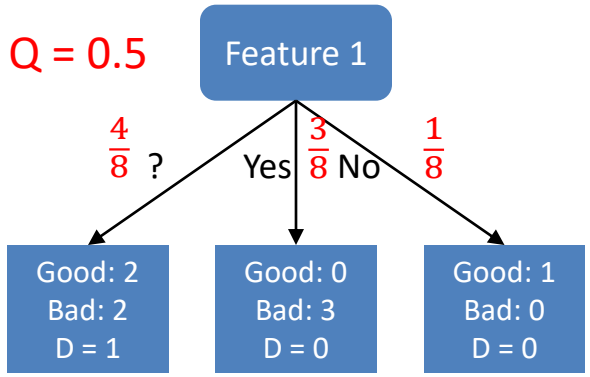






Quality

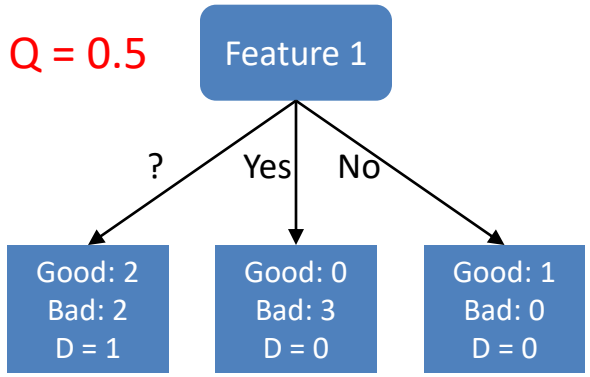
- Quality of test in the first level



Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad

Quality

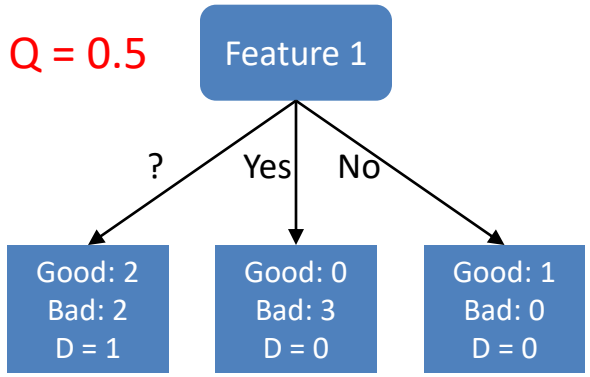
- Quality of test in the second level

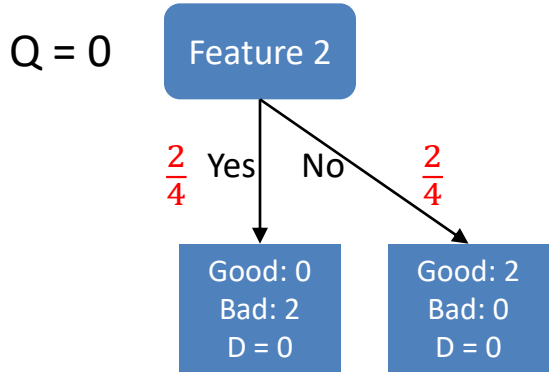


Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad

Quality

- Quality of test in the second level



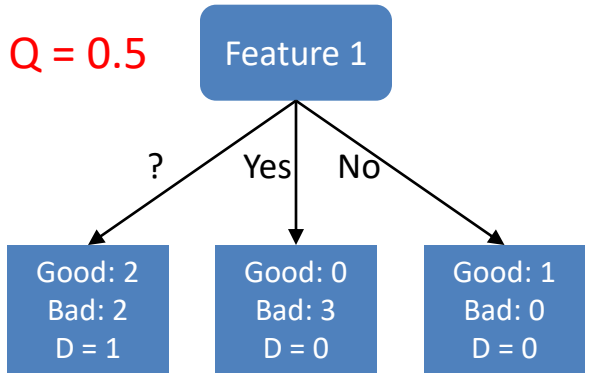


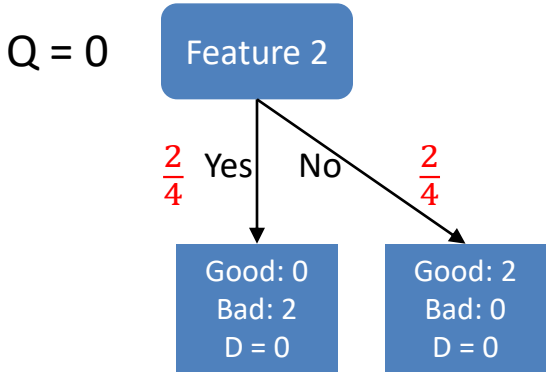
Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad

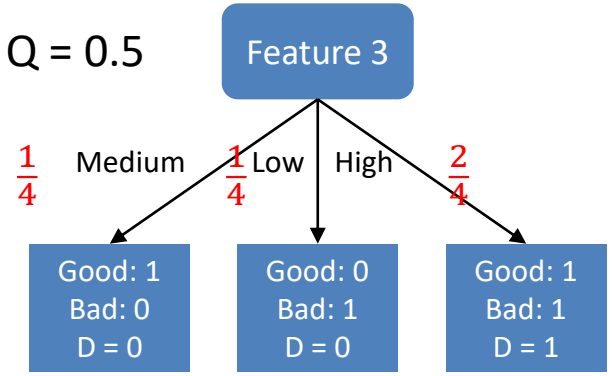
Quality

- Quality of test in the second level

Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad



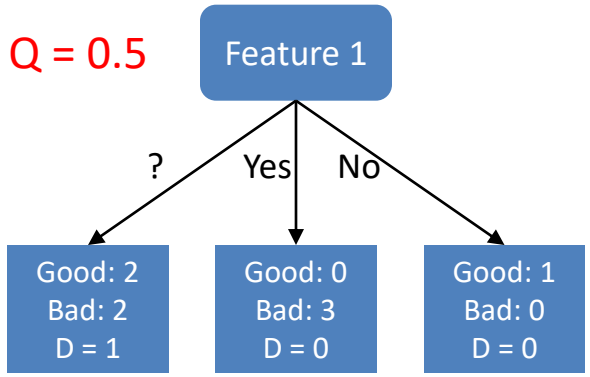


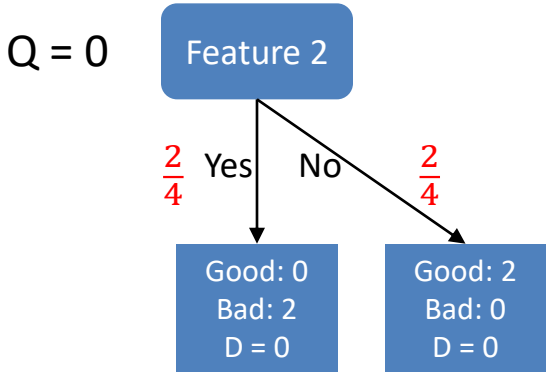


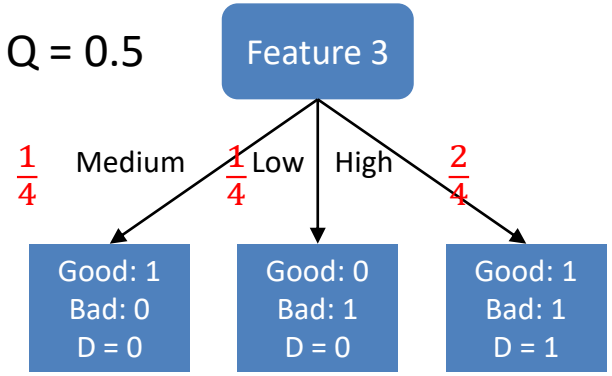
Quality

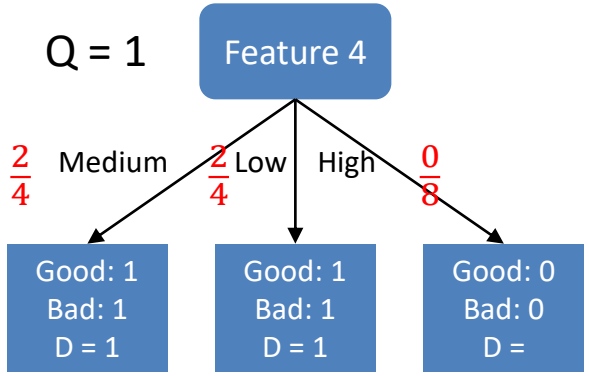
- Quality of test in the second level

Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad



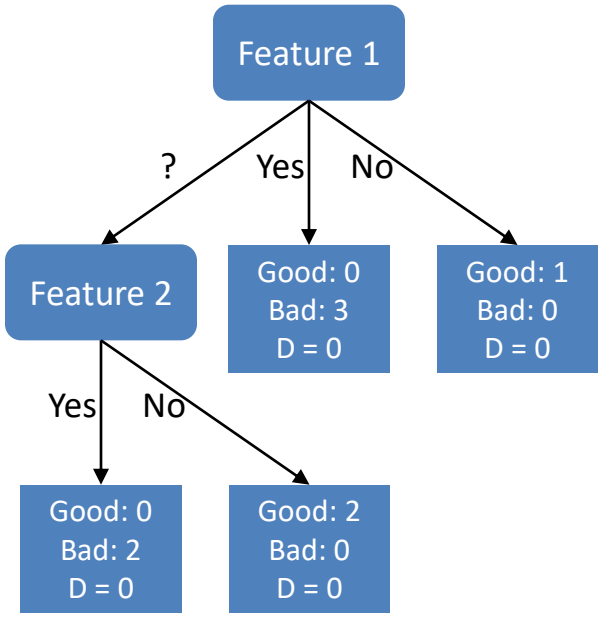






Quality

- Quality of test in the second level



Feature 1	Feature 2	Feature 3	Feature 4	output
?	Yes	Low	Medium	Bad
Yes	Yes	High	Medium	Bad
?	No	High	Medium	Good
No	No	Medium	High	Good
?	No	Medium	Low	Good
Yes	No	Low	High	Bad
Yes	No	Medium	High	Bad
?	Yes	High	Low	Bad