

Homework set on Tree Distribution

1. **Tree Distribution:** In the following items you should write whether the statement is true or false, and explain your answer.

(a) **True/False:** In our Tree Distribution lecture, we concluded that the criteria for an optimal tree is

$$\max_{\text{All Trees}} \sum_{i=1}^n I(x_i, x_{j(i)}),$$

where x_i is the i_{th} feature, $x_{j(i)}$ is the parent of the i_{th} feature, and I is the mutual information between the two features. This criteria is equivalent to the *Maximum-Likelihood* criteria.

(b) **True/False:** There can be more than two 'sons' per node in a Tree Distribution.

2. **Tree Distribution:** You wish to generate a model to predict if a mushroom is poisonous or not. You have some empirical data:

Example	Is heavy	Is smelly	Is spotted	Is smooth	Is poisonous
A	0	0	0	0	0
B	0	0	1	0	0
C	1	1	0	1	0
D	1	0	0	1	1
E	0	1	1	0	1
F	0	0	1	1	1
G	0	0	0	1	1
H	1	1	0	0	1

(a) Calculate the empirical mutual information between all couples of features (including *Is poisonous*).

(b) Build tree distribution for the data according to the maximum-likelihood criteria. You have a constraint that the node of 'Is poisonous' must be the main root (head) of the tree.

(c) Use the tree you built to determine by the maximum-likelihood criteria whether U,V,W are poisonous or not. If it happens to be that there is a tie, you define it as poisonous.

Example	Is heavy	Is smelly	Is spotted	Is smooth	Is poisonous
U	0	1	1	1	?
V	0	1	0	1	?
W	1	1	0	0	?