

MSBD5002: Data Mining and Knowledge Discovery (MicroMasters)

Exercise 5
Classification

Q1 Decision Tree

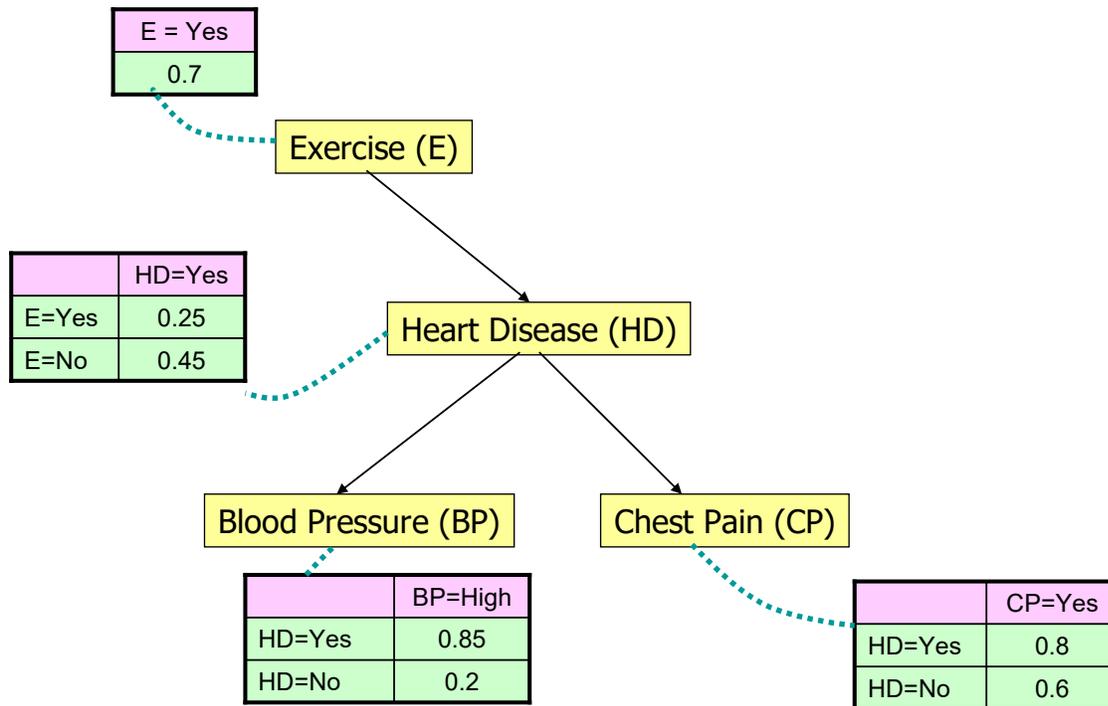
Race	Income	Child	Insurance
black	high	no	yes
white	high	yes	yes
white	low	yes	yes
white	low	yes	yes
black	low	no	no
black	low	no	no
black	low	no	no
white	low	no	no

Consider the above table.

- (a) In C4.5 decision tree, what is the value of $\text{Gain}(\text{Child}, T)$ where T is a set of all tuples in the above table?
- (b) In CART decision tree, what is the value of $\text{Gain}(\text{Child}, T)$ where T is a set of all tuples in the above table?

Q2 Bayes Classifier

We have the following Bayesian Belief Network.



Suppose that there is a new person. We know that

- (1) he does exercises
- (2) he has high blood pressure
- (3) he has chest pain

We would like to know whether he is likely to have Heart Disease.

Exercise	Blood Pressure	Chest Pain	Heart Disease
Yes	High	Yes	?

Please use Bayesian classifier with the use of Bayesian Belief Network to predict whether he is likely to have Heart Disease.