# ASSIGNMENT ON DECISION TREES

1. The following table contains training examples that help predict whether a patient is likely to have a heart attack.

|  |  |  |  |
| --- | --- | --- | --- |
| PATIENT ID CHEST PAIN? | MALE? SMOKES? | EXERCISES? | HEART ATTACK? |
| 1. yes | yes no | yes | yes |
| 2. yes | yes yes | no | yes |
| 3. no | no yes | no | yes |
| 4. no | yes no | yes | no |
| 5. yes | no yes | yes | yes |
| 6. no | yes yes | yes | no |

Use information theory to construct a minimal decision tree that predicts whether or not a patient is likely to have a heart attack. SHOW EACH STEP OF THE COMPUTATION.

# Soln:

YES

NO

YES

NO

YES

NO

EXERCISE ?

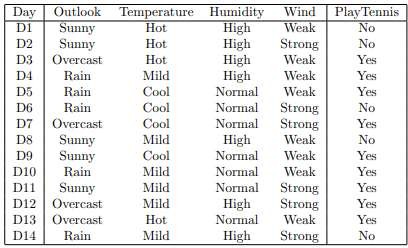
YES

**CHEST PAIN ?**

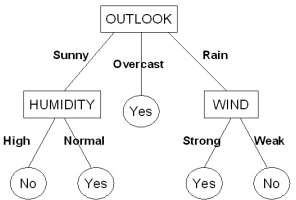
The following decision tree is used to predict whether a person can have a heart attack or not.

* The above decision tree tells us that if a person has a chest pain then he is likely to have heart attack.
* If the person does not have chest pain, then we take another attribute from the dataset into consideration- exercise.
* If the person does not have chest pain but does exercise, then is not likely to have heart attack.
* If the person does not have chest pain but does not exercise, then he is likely to have heart attack.
* We have chosen the attributes from the given dataset in such a way that the depth of the decision tree is as minimal as possible.

1. A) Build a decision tree from the given tennis dataset. You should build a tree to predict Play Tennis, based on the other attributes (but, do not use the Day attribute in your tree.). Show all of your work, calculations, and decisions as you build the tree.



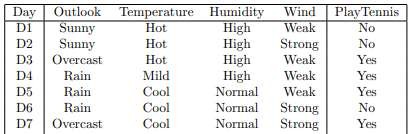
# Soln:



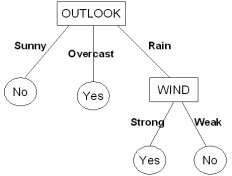
The following decision tree is used to predict whether we can play tennis or not.

* + If the outlook is overcast, then we can play tennis.
  + If outlook is rain, then we consider another attribute from the given dataset- wind.
  + If the outlook is rain and the wind is strong, then we cannot play tennis.
  + If the outlook is rain and the wind is weak, then we can play tennis.
  + If the outlook is sunny, then we consider another attribute from the given dataset- humidity.
  + If the outlook is sunny and humidity is high, we cannot play tennis.
  + If the outlook is sunny and humidity is normal, we can play tennis.
  + The above decision tree is constructed by choosing attributes from the dataset which makes the depth of the decision tree to be as minimal as possible.

1. Now, build a tree using only examples D1–D7.



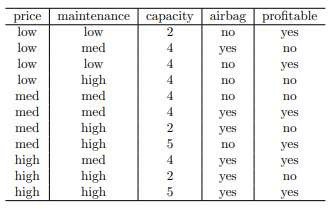
**Soln:**



The following decision tree is used to predict whether we can play tennis or not.

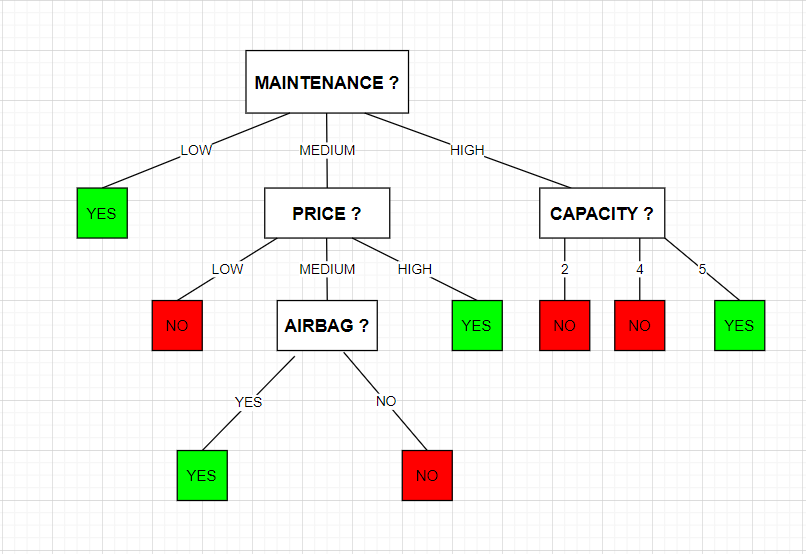
* + If the outlook is overcast, then we can play tennis.
  + If the outlook is sunny, then we cannot play tennis.
  + If outlook is rain, then we consider another attribute from the given dataset- wind.
  + If the outlook is rain and the wind is strong, then we cannot play tennis.
  + If the outlook is rain and the wind is weak, then we can play tennis.
  + The above decision tree is constructed by choosing attributes from the dataset which makes the depth of the decision tree to be as minimal as possible.

1. Consider the following data set:



Considering ’profitable’ as the binary values attribute, we are trying to predict construct a decision tree.

**Soln:**



The following decision tree is used to predict whether it is profitable or not.

* If the maintenance is low, then it is profitable.
* If maintenance is medium, then we consider another attribute from the dataset- price.
* If the maintenance is medium and the price is high, then it is profitable.
* If the maintenance is medium and price is low, then it is not profitable.
* If the maintenance is medium and the price is also medium, we consider another attribute called airbag.
* If the maintenance is medium, price is medium and if airbag is yes, then its profitable.
* If the maintenance is medium, price is medium and if airbag is no, then it’s not profitable.
* If the maintenance is high then we consider another attribute from the dataset- capacity.
* If maintenance is high and capacity is 2, it’s not profitable.
* If maintenance is high and capacity is 4, it’s not profitable.
* If maintenance is high and capacity is 5, it’s profitable.
* The above decision tree is constructed by choosing attributes from the dataset which makes the depth of the decision tree to be as minimal as possible.