

**Introduction to Classification: single feature models***(find and work with a partner)*

Consider the following dataset, modified from the Cleveland Heart Disease dataset here: <https://archive.ics.uci.edu/ml/datasets/heart+Disease>. This dataset has  $p = 2$  features and  $n = 16$  examples. Our goal is to predict the class label as negative ( $-1$ , no heart disease) vs. positive ( $+1$ , heart disease). We will consider this dataset as the *training data*.

```
@relation cleveland-14-heart-disease
@attribute 'slope' { up, flat, down}
@attribute 'thal' { fixed_defect, normal, reversable_defect}
@attribute 'class' { negative, positive}
@data
down, fixed_defect,      negative
down, normal,           negative
up, normal,             negative
up, normal,             negative
up, normal,             negative
flat, fixed_defect,      negative
flat, normal,           negative
up, reversable_defect,   negative
flat, normal,           positive
flat, reversable_defect, positive
down, normal,           positive
flat, reversable_defect, positive
down, reversable_defect, positive
flat, fixed_defect,      positive
down, reversable_defect, positive
flat, normal,           positive
```

1. Draw a decision tree for the feature `slope` (which relates to heart rate during exercise). Each edge corresponds to one feature value. At each leaf, compute the *probability* of a positive classification.
2. Do the same thing for the feature `thal` (which relates to the blood disorder *Thalassemia*).
3. If our threshold is  $1/2$ , which feature values would classify a new test example as positive? What if the threshold is  $1/12$ ?