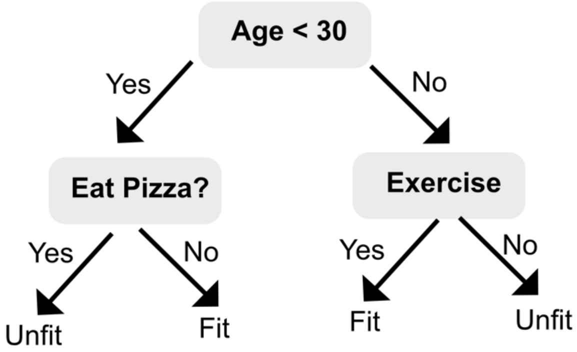
# HW1

**ID3 algorithm: Generating a decision tree**

Consider the following classic decision learning problem. Given data about a discrete 'target' variable that depends on some other discrete 'attribute' variables, we would like to create a hierarchical tree where the values of attribute variables lead us to a target result.



Like in the above figure, fitness is the target variable. Age, Eat pizza and exercise are attributes which are arranged in the right order to make optimal decisions about a person's fitness.

The precursor to building a decision tree is data with different values of the attributes and the corresponding target. Check the datafile provided with this homework (id3_data.csv). The ID3 algorithm splits the target set (like "work at office" in the datafile) based on the attribute that results in maximum information gain at every step. One eventually gets a sequence of splits along each branch until the entropy of the target subsets is 0.

The algorithm can be summarized as:

1. Find entropy H(S) of the target column S.
2. Divide the target column into subsets based on one of the attributes. Eg: The target column (work from home) can be divided into two subsets based on the attribute mode of transport since it takes only two possible values.
3. Calculate entropy H(S|A) of the divided target column conditioned on the attribute. Then the information gain is: IG = H(S) – H(S|A).

   We could use the following simplification: H(S|A) = $\sum_t p(t)H(t)$ where t represents the different subsets.
   p(t) is the proportion of elements in subset t and H(t) is the corresponding entropy for the subset of the target.

4. Similarly, we calculate the information gains based on the other attributes. The attribute with the highest IG is used for the first partition.

5. This process is repeated recursively on the constituent partitions until H(S) is 0 along each branch.

Use the ID3 algorithm on the given data (id3_data.csv) to find the sequence of attributes resulting in the highest information gain. It is sufficient to traverse along only one branch but make sure to pick the attribute with the highest gain at each step. You may choose how to present your results, but be sure to include the following:

- sequence of optimal attributes
- values of those attributes along chosen branch
- H(S|A) at each step given the optimal attribute choice
- H(S) of the target column after each step

The corresponding code should be shared in an appendix.